14.310x Lecture 11

Probability---the sample mean

The sample mean is the arithmetic average of the n random variables from a random sample of size n. We denote it

$$\bar{X}_n = (1/n)(X_1 + \dots + X_n)$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i$$

We also call the arithmetic average of the *realizations* of those n random variables the sample mean.

Probability---the sample mean

The sample mean is the arithmetic average of the n random variables (or realizations) from a random sample of size n. We denote it

$$\bar{X}_n = (1/n)(X_1 + \ldots + X_n)$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i$$

Why would such a function be useful?

Probability---the sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

The X's are random variables, so $\bar{X}_n$ is also a random variable (since it is a function of random variables).

So let's figure out how it's distributed.

Probability---the sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

The X's are random variables, so $\bar{X}_n$ is also a random variable (since it is a function of random variables).

So let's figure out how it's distributed.

(Note that, if we knew how the X's were distributed, we might be able to use something like the n-version of the convolution formula. For now, let's just use properties of moments to figure out the moments of $\bar{X}_n$ as functions of the moments of the X's.)

Probability---the sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Expectation of the sample mean:

$$E(\bar{X}_n) = E((1/n)(\Sigma X_i)) = 1/n\, \Sigma E(X_i) = 1/n\, \Sigma \mu = \mu$$

Probability---the sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Expectation of the sample mean:

$$E(\bar{X}_n) = E((1/n)(\Sigma X_i)) = 1/n\, \Sigma E(X_i) = 1/n\, \Sigma \mu = \mu$$

by definition

by properties of E

just changing notation

Probability---the sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Expectation of the sample mean:

$$E(\bar{X}_n) = E((1/n)(\Sigma X_i)) = 1/n\, \Sigma E(X_i) = 1/n\, \Sigma \mu = \mu$$

Variance of the sample mean:

$$Var(\bar{X}_n) = Var(1/n(\Sigma X_i)) = 1/n^2\, \Sigma Var(X_i)$$

$$= 1/n^2\, \Sigma \sigma^2 = \sigma^2/n$$

Probability---the sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Expectation of the sample mean:

$$E(\bar{X}_n) = E((1/n)(\sum X_i)) = 1/n \sum E(X_i) = 1/n \sum \mu = \mu$$

by properties of Var

Variance of the sample mean:

$$Var(\bar{X}_n) = Var(1/n(\sum X_i)) = 1/n^2 \sum Var(X_i)$$

$$= 1/n^2 \sum \sigma^2 = \sigma^2/n$$

Probability---the sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Note that we used independence in the variance calculation but did not need it in the expectation calculation.

Expectation of the sample mean:

$$E(\bar{X}_n) = E((1/n)(\Sigma X_i)) = 1/n\, \Sigma E(X_i) = 1/n\, \Sigma \mu = \mu$$

Variance of the sample mean:

$$\mathrm{Var}(\bar{X}_n) = \mathrm{Var}(1/n(\Sigma X_i)) = 1/n^2\, \Sigma \mathrm{Var}(X_i)$$
$$= 1/n^2\, \Sigma \sigma^2 = \sigma^2/n$$

Probability---the sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

What do these calculations tell us?
Distribution of sample mean is centered around the mean, more concentrated than original distribution, becoming more concentrated as n gets large.

Expectation of the sample mean:

$$E(\bar{X}_n) = E((1/n)(\Sigma X_i)) = 1/n\, \Sigma E(X_i) = 1/n\, \Sigma \mu = \mu$$

Variance of the sample mean:

$$\mathrm{Var}(\bar{X}_n) = \mathrm{Var}(1/n(\Sigma X_i)) = 1/n^2\, \Sigma \mathrm{Var}(X_i)$$
$$= 1/n^2\, \Sigma \sigma^2 = \sigma^2/n$$

Probability---the Central Limit Theorem

That brings us to one of the most important and useful results in all of probability theory, which really serves as the basis for statistics, the Central Limit Theorem.

Probability---the Central Limit Theorem

Let $X_1, \ldots, X_n$ form a random sample of size n from a distribution with finite mean and variance. Then for any fixed number x,

$$\lim_{n \to \infty} P\left[\frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \le x \right] = \Phi(x)$$

Probability---the Central Limit Theorem

Let $X_1, \ldots, X_n$ form a random sample of size n from a distribution with finite mean and variance. Then for any fixed number x,

$$\lim_{n \to \infty} P\left[\frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \leq x \right] = \Phi(x)$$

This is the sample mean, with its mean subtracted off and divided by the square root of its variance.

Probability---the Central Limit Theorem

Let $X_1, \ldots, X_n$ form a random sample of size n from a distribution with finite mean and variance. Then for any fixed number x,

$$\lim_{n \to \infty} P\left[\frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \le x \right] = \Phi(x)$$

This is the sample mean, with its mean subtracted off and divided by the square root of its variance.
So it is now standardized---it has mean zero and variance one.

Probability---the Central Limit Theorem

Let $X_1, \ldots, X_n$ form a random sample of size n from a distribution with finite mean and variance. Then for any fixed number x,

$$\lim_{n \to \infty} P\left[\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma} \leq x \right] = \Phi(x)$$

This is special notation for the CDF of a standard normal random variable.

Probability---the Central Limit Theorem

Let $X_1, \ldots, X_n$ form a random sample of size n from a distribution with finite mean and variance. Then for any fixed number x,

$$\lim_{n \to \infty} P\left[\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leq x \right] = \Phi(x)$$

So, basically, we take a standardized version of the sample mean from any old distribution, let the sample size go to infinity, and note that essentially the definition of the CDF of that thing is the standard normal CDF.

Probability---the Central Limit Theorem

Let $X_1, \ldots, X_n$ form a random sample of size n from a distribution with finite mean and variance. Then for any fixed number x,

$$\lim_{n \to \infty} P\left[\frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \leq x \right] = \Phi(x)$$

Practically speaking, if you have a sample mean from a reasonably large random sample from any distribution, it will have an approximate $N(\mu, \sigma^2/n)$ distribution.

Probability---the Central Limit Theorem

It's remarkable. You can have any kind of crazy, almost degenerate, distribution. As long as the random sample taken from that distribution is reasonably large, the sample mean will be approximately normal.

Probability---the Central Limit Theorem

It's remarkable. You can have any kind of crazy, almost degenerate, distribution. As long as the random sample taken from that distribution is reasonably large, the sample mean will be approximately normal.

It's not just remarkable; it's useful. We don't need to know what distribution we're sampling from to know a lot about the behavior of the sample mean from that distribution.

Probability---the Central Limit Theorem

It's remarkable. You can have any kind of crazy, almost degenerate, distribution. As long as the random sample taken from that distribution is reasonably large, the sample mean will be approximately normal.

It's not just remarkable; it's useful. We don't need to know what distribution we're sampling from to know a lot about the behavior of the sample mean from that distribution.

We will rely on this theorem (implicitly) for the rest of the semester. This also gives you some notion of why the normal distribution is so important.

Statistics

For the first time this semester, the title on my slide does not begin with "Probability."

Statistics

So what is statistics? It's the study of estimation and inference. We'll get to inference a little bit later. For now, we'll focus on estimation.

Statistics

So what is statistics? It's the study of estimation and inference. We'll get to inference a little bit later. For now, we'll focus on estimation.

We've seen examples of estimators (sample mean, "guess" of the size of the bufflehead population), but a more general discussion would be helpful.

Statistics

So what is statistics? It's the study of estimation and inference. We'll get to inference a little bit later. For now, we'll focus on estimation.

We've seen examples of estimators (sample mean, "guess" of the size of the bufflehead population), but a more general discussion would be helpful.

An estimator is a function of the random variables in a random sample. The specific function is chosen to have properties useful for giving us information about the distribution of those random variables.

Statistics

First, a definition:

A <u>parameter</u> is a constant indexing a family of distributions.
Examples of parameters are μ and σ^2 from the normal distribution, λ from the exponential distribution, a and b from the uniform distribution, n and p from the binomial distribution. We will often use θ as a general notation for a parameter.

Statistics

First, a definition:

A <u>parameter</u> is a constant indexing a family of distributions. Examples of parameters are μ and σ^2 from the normal distribution, λ from the exponential distribution, a and b from the uniform distribution, n and p from the binomial distribution. We will often use θ as a general notation for a parameter.

We typically want to determine the values of parameters that govern an observed stochastic process or phenomenon--- <u>estimating unknown parameters</u>. We will often use $\hat{\theta}$ as general notation for an estimator.

Statistics

And an important distinction:

We need to think of random variables in two ways, as the mathematical construct I introduced several weeks ago, i.e., a function from the sample space to the real numbers, and as a stochastic object that can "take on" different realizations with different probabilities. We use the notation X to stand for the random variable, as always, and use x to stand for the realization (or possible realizations).

Before, I defined a random sample as an i.i.d. collection of random variables. We can also call the *realizations* of those random variables a random sample. Or just data.

Statistics

In other words, we know (or assume) that a set of random variables, a random sample, is distributed i.i.d. normal, or i.i.d. uniform, or i.i.d. exponential. Estimation is trying to determine the specific μ and σ^2, or a and b, or λ.

Statistics

For instance, we might choose a function whose result, when applied to the random sample, is a random variable tightly distributed around the mean of the distribution of those random variables.

Then we plug the realizations of the random sample, or data, into the function to obtain a number, a realization of the function.

The function of the random sample is the <u>estimator</u>. The number, or realization of the function of the random sample, is the <u>estimate</u>. (We use $\hat{\theta}$ to stand for both.)

Statistics---example

Suppose $X \sim U[0,\theta]$

$$f_X(x) = \begin{cases} 1/\theta & 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}$$

Want to estimate θ. What could you do?

Statistics---example

Suppose $X \sim U[0,\theta]$

$$f_X(x) = \begin{cases} 1/\theta & 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}$$

Two reasonable procedures come to mind:

Gather a random sample, compute the sample mean, and multiply by 2. Use that as $\hat{\theta}$.

Gather a random sample, compute the max (n^{th} order statistic) of the random sample. Use that as $\hat{\theta}$.

Statistics---example

Suppose $X \sim U[0,\theta]$

$$f_X(x) = \begin{cases} 1/\theta & 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}$$

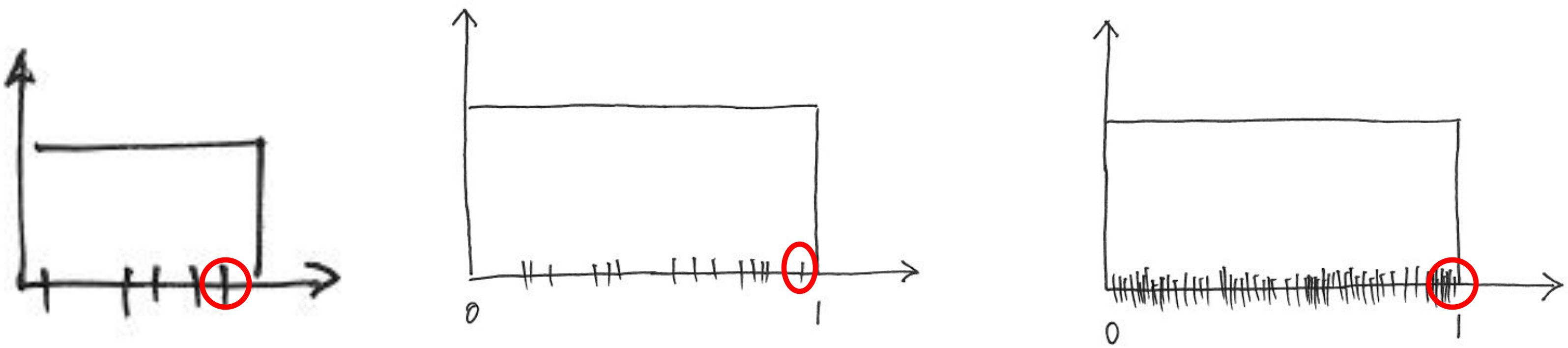
In other words,

$$\hat{\theta}_1 = \max\{X_1, X_2, \ldots X_n\}$$

$$\hat{\theta}_2 = 2 \, \frac{1}{n} \sum_{i=1}^{n} X_i$$

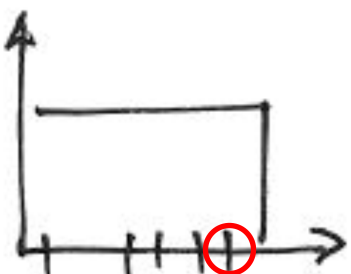
Statistics---example

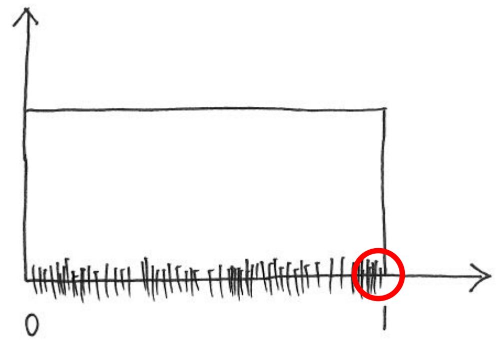
$$\hat{\theta}_1 = \max\{X_1, X_2, \ldots X_n\}$$



Statistics---example

$$\hat{\theta}_1 = \max\{X_1, X_2, \ldots X_n\}$$

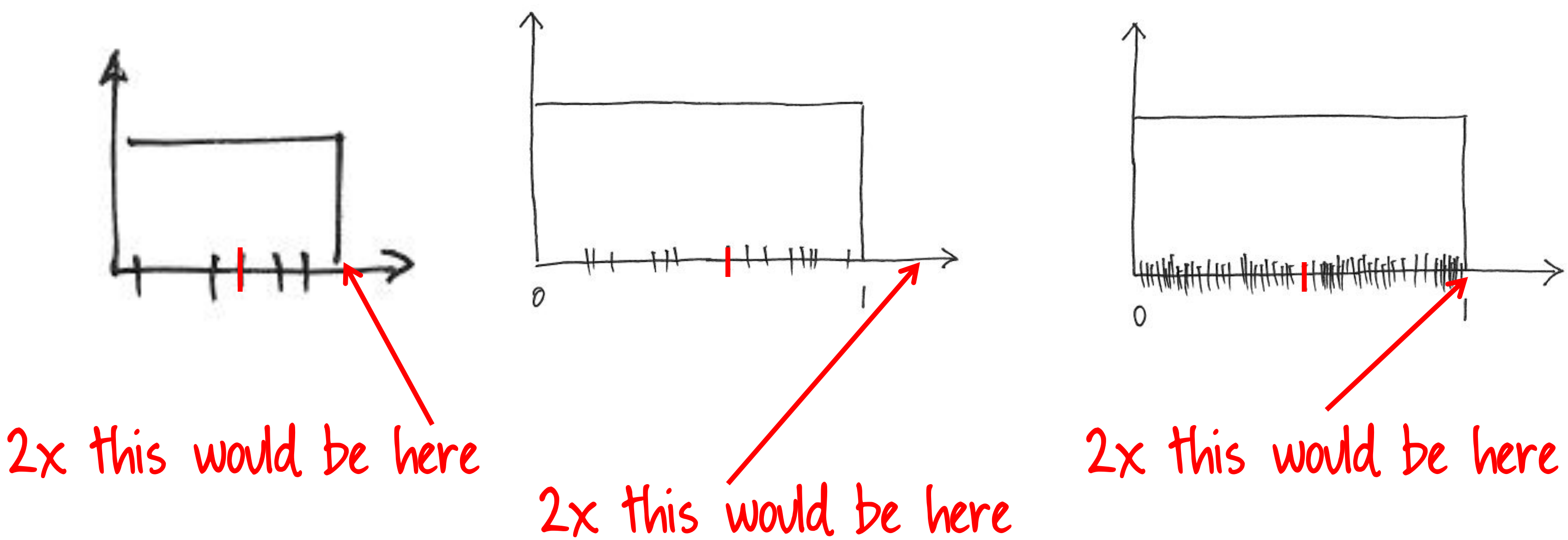




It's getting to be a better estimator as n gets larger.

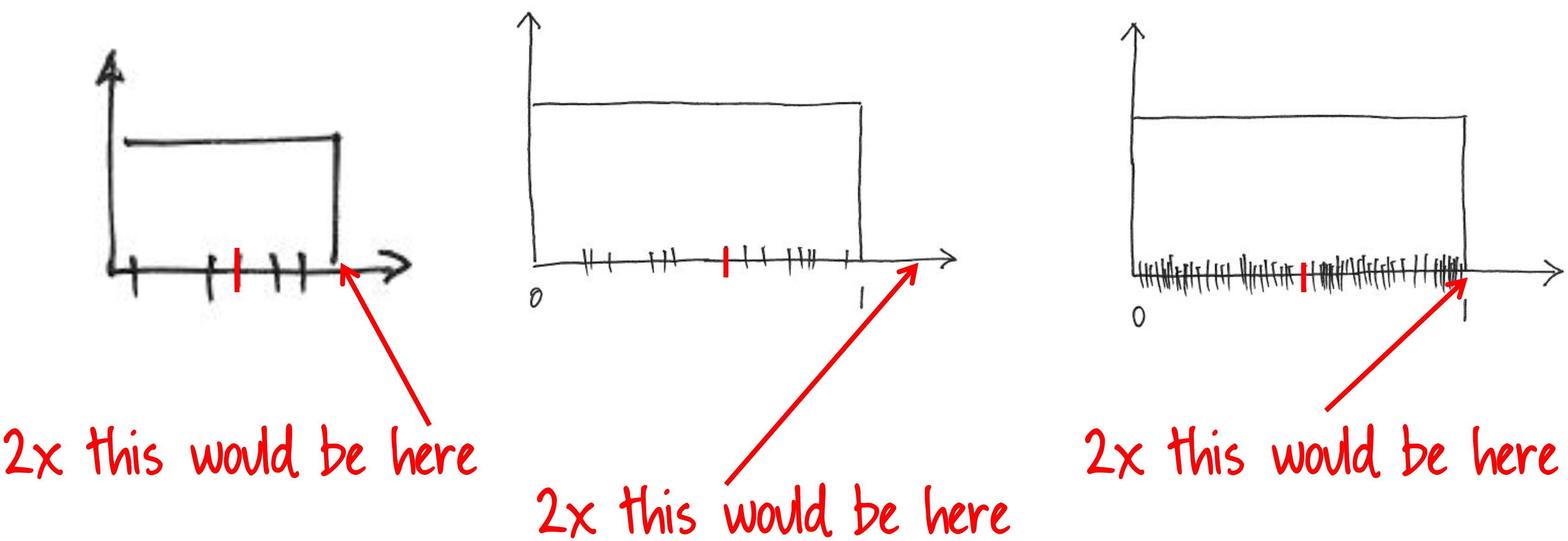
Statistics---example

$$\hat{\theta}_2 = 2\,\frac{1}{n}\sum_{i=1}^{n} x_i$$



Statistics---example

$$\hat{\theta}_2 = 2\frac{1}{n}\sum_{i=1}^{n} X_i$$



Also a better estimator as n gets larger (but bounces around).

Statistics---example

Suppose $X \sim U[0,\theta]$

$$f_X(x) = \begin{cases} 1/\theta & 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}$$

Here's another procedure:

Gather a random sample, compute the sample median (the number above and below which half of the sample falls), and multiply by 2. Use that as $\hat{\theta}$.

Statistics---example

Suppose $X \sim U[0,\theta]$

$$f_X(x) = \begin{cases} 1/\theta & 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}$$

Here's another procedure:

Gather a random sample, compute the sample median (the number above and below which half of the sample falls), and multiply by 2. Use that as $\hat{\theta}$.

Also seems reasonable.

Statistics---example

Suppose $X \sim U[0,\theta]$

$$f_X(x) = \begin{cases} 1/\theta & 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}$$

Here's another procedure:

Gather a random sample, throw the whole thing away, and have R generate a random value for you. Use that as $\hat{\theta}$.

Statistics---example

Suppose $X \sim U[0,\theta]$

$$f_X(x) = \begin{cases} 1/\theta & 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}$$

Here's another procedure:

Gather a random sample, throw the whole thing away, and have R generate a random value for you. Use that as $\hat{\theta}$.

We can guess that this procedure does not have good properties.

Statistics---example

Suppose $X \sim U[0,\theta]$

$$f_X(x) = \begin{cases} 1/\theta & 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}$$

$$\hat{\theta}_1 = \max\{X_1, X_2, \ldots X_n\}$$

$$\hat{\theta}_2 = 2 \, \frac{1}{n} \sum_{i=1}^{n} X_i$$

How did we come up with these functions? How do we know if they're reasonable? How do we choose among them?

Statistics

For the rest of this lecture and some of next, we will talk about two topics, criteria for assessing estimators and frameworks for choosing estimators. These topics will answer those questions I just posed.

Statistics---criteria for assessing estimators

Recall that an estimator is a random variable. So it has a distribution. Our criteria for assessing estimators will be based on characteristics of their distributions.

Statistics---criteria for assessing estimators

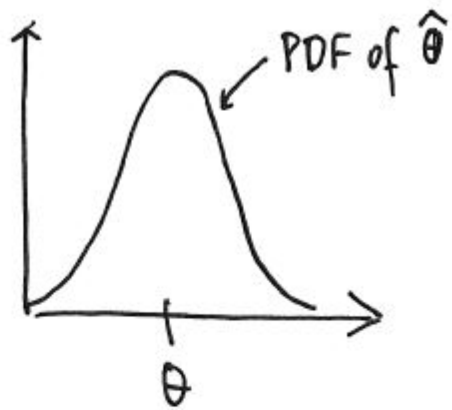
Recall that an estimator is a random variable. So it has a distribution. Our criteria for assessing estimators will be based on characteristics of their distributions.

An estimator is <u>unbiased</u> for θ if $E(\hat{\theta}) = \theta$ for all θ in Θ.

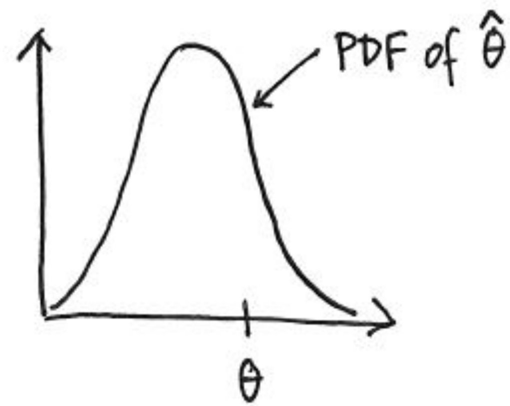
Statistics---criteria for assessing estimators

Recall that an estimator is a random variable. So it has a distribution. Our criteria for assessing estimators will be based on characteristics of their distributions.

An estimator is <u>unbiased</u> for θ if $E(\hat{\theta}) = \theta$ for all θ in Θ.



unbiased



biased