Lecture 2.1: Intuition Behind Sampling Distributions

2014/01/29

Behavior of Statistics

Say we draw a random sample of size $n = 100$ from any distribution that has true population mean $\mu = 5$.

Behavior of Statistics

Say we draw a random sample of size $n = 100$ from any distribution that has true population mean $\mu = 5$.

Two Important Conceptual Questions:

Behavior of Statistics

Say we draw a random sample of size $n = 100$ from any distribution that has true population mean $\mu = 5$.

Two Important Conceptual Questions:

1. If we compute the sample mean $\overline{x}$ of these points, are we going to always get exactly 5?

Behavior of Statistics

Say we draw a random sample of size $n = 100$ from any distribution that has true population mean $\mu = 5$.

Two Important Conceptual Questions:

1. If we compute the sample mean $\overline{x}$ of these points, are we going to always get exactly 5?
2. Say we do this once and the sample mean is $\overline{x} = 5.025$. If we repeat this procedure
   (i.e. generate a new sample of 100 points and compute $\overline{x}$)
   are we going to get $\overline{x} = 5.025$ exactly?

Behavior of Statistics

Do this for the 1st time

We get, say, $\overline{x} = 4.831$

Behavior of Statistics

Do this for the 1st time	We get, say, $\overline{x} = 4.831$
Do this for the 2nd time	We get, say, $\overline{x} = 5.104$

Behavior of Statistics

Do this for the 1st time	We get, say, $\overline{x} = 4.831$
Do this for the 2nd time	We get, say, $\overline{x} = 5.104$
Do this for the 3rd time	We get, say, $\overline{x} = 4.965$

Behavior of Statistics

Do this for the 1st time	We get, say, $\overline{x} = 4.831$
Do this for the 2nd time	We get, say, $\overline{x} = 5.104$
Do this for the 3rd time	We get, say, $\overline{x} = 4.965$
...	

Behavior of Statistics

Do this for the 1st time	We get, say, $\overline{x} = 4.831$
Do this for the 2nd time	We get, say, $\overline{x} = 5.104$
Do this for the 3rd time	We get, say, $\overline{x} = 4.965$
...	
Do this for the 1000th time	We get, say, $\overline{x} = 4.957$

Behavior of Statistics

Do this for the 1st time	We get, say, $\overline{x} = 4.831$
Do this for the 2nd time	We get, say, $\overline{x} = 5.104$
Do this for the 3rd time	We get, say, $\overline{x} = 4.965$
...	
Do this for the 1000th time	We get, say, $\overline{x} = 4.957$

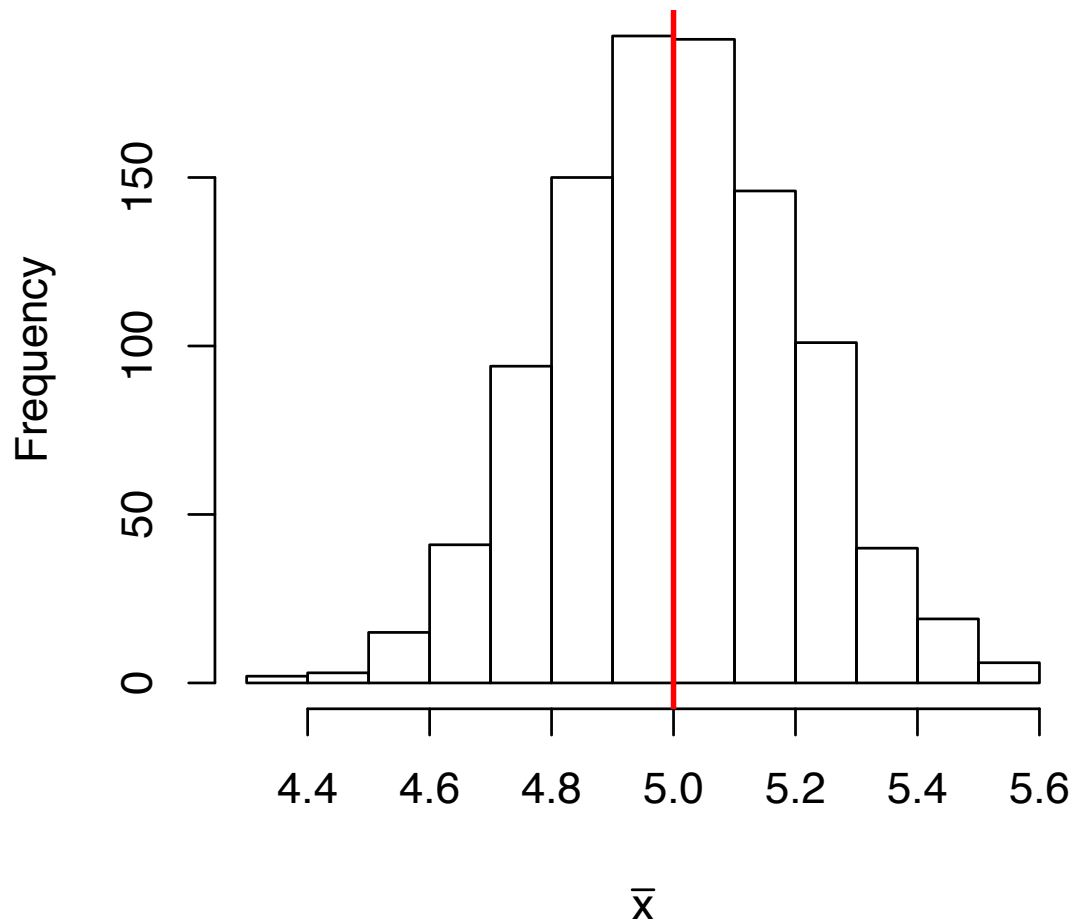
The sampling distribution describes the random behavior of these estimates.

Example: Sampling Distribution of the Sample Mean

Consider the histogram of 1000 instances of $\overline{x}$, where each $\overline{x}$ is computed from a sample of $n = 100$ Normal($\mu = 5, \sigma^2 = 4$) RV's.

Example: Sampling Distribution of the Sample Mean

Consider the histogram of 1000 instances of $\overline{x}$, where each $\overline{x}$ is computed from a sample of $n = 100$ Normal($\mu = 5, \sigma^2 = 4$) RV's.



Standard Error of the Sample Mean $\overline{X}$

The standard deviation associated with a statistic is called the standard error.

Standard Error of the Sample Mean $\overline{X}$

The standard deviation associated with a statistic is called the standard error.

If the X_i's have standard deviation σ, the standard error of the sample mean is

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

Standard Error of the Sample Mean $\overline{X}$

The standard deviation associated with a statistic is called the standard error.

If the X_i's have standard deviation σ, the standard error of the sample mean is

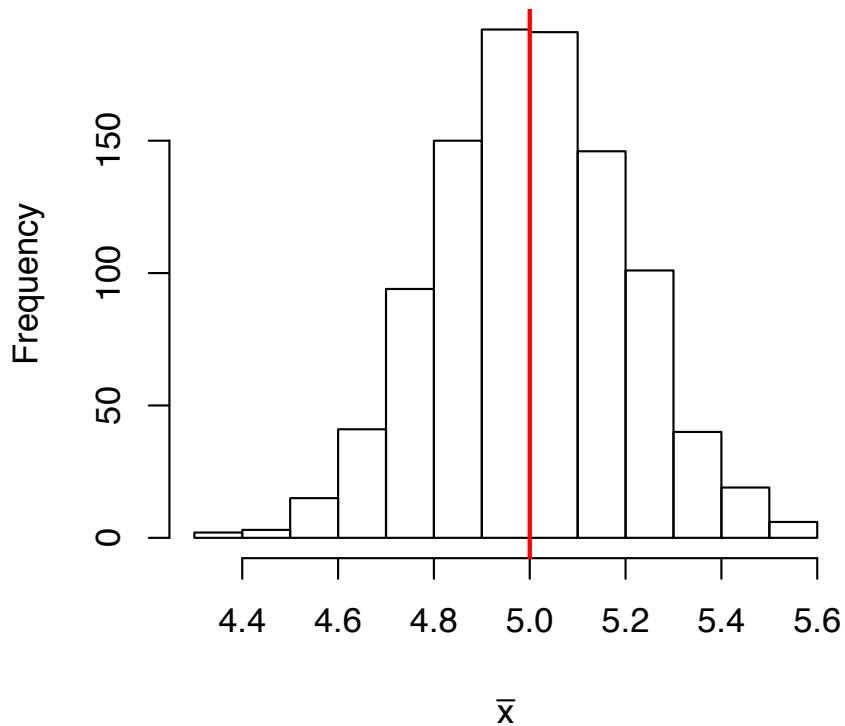
$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

Notice the $\sqrt{n}$ in the denominator: n increases, SE decreases!

Standard Error of the Sample Mean $\overline{X}$

Recall $n = 100$ with $X_i \sim \text{Normal}(\mu = 5, \sigma^2 = 4)$. The standard error is the SD of the sampling distribution:

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{100}} = \frac{2}{10} = 0.2$$



Standard Error of the Sample Mean $\overline{x}$

Now compare the sampling distributions based on

Standard Error of the Sample Mean $\overline{x}$

Now compare the sampling distributions based on

- 1000 instances of $\overline{x}$ where $n = 100$. $SE = \frac{2}{\sqrt{100}} = 0.2$

Standard Error of the Sample Mean $\overline{x}$

Now compare the sampling distributions based on

- 1000 instances of $\overline{x}$ where $n = 100$. $SE = \frac{2}{\sqrt{100}} = 0.2$
- 1000 instances of $\overline{x}$ where $n = 1000$. $SE = \frac{2}{\sqrt{1000}} = 0.0632$

Standard Error of the Sample Mean $\overline{x}$

Now compare the sampling distributions based on

- 1000 instances of $\overline{x}$ where $n = 100$. $SE = \frac{2}{\sqrt{100}} = 0.2$
- 1000 instances of $\overline{x}$ where $n = 1000$. $SE = \frac{2}{\sqrt{1000}} = 0.0632$