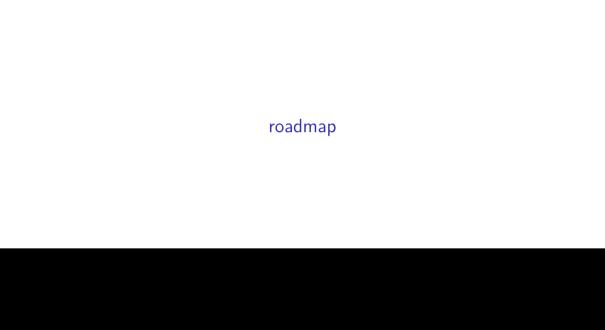
#### STA286 Lecture 23

**Neil Montgomery** 

Last edited: 2017-06-27 15:07



#### probability/statistics

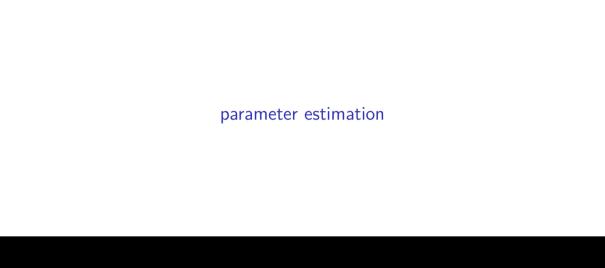
I said probability was "done", but it really wasn't.

In statistics we use a sample to make statements about what is unknown about an underlying distribution.

Then we did more probability - but not for the purpose of modeling actual random processes in the wild.

The purpose of the additional probability was to determine some properties of functions of samples—with a focus on the properties of  $\overline{X}$ 

Now we are actually going to do statistics.



We will treat population parameters as constants (as opposed to Bayesian statistics).

The goal is to use a statistic  $\hat{\theta}$ , i.e. a function of a sample, to estimate the value of a parameter  $\theta$ , which could be a vector. This statistic is called a "point estimator".

We will treat population parameters as constants (as opposed to Bayesian statistics).

The goal is to use a statistic  $\hat{\theta}$ , i.e. a function of a sample, to estimate the value of a parameter  $\theta$ , which could be a vector. This statistic is called a "point estimator".

e.g. Sample is from  $N(\mu,3)$ . We want to estimate  $\mu$ .

We will treat population parameters as constants (as opposed to Bayesian statistics).

The goal is to use a statistic  $\hat{\theta}$ , i.e. a function of a sample, to estimate the value of a parameter  $\theta$ , which could be a vector. This statistic is called a "point estimator".

e.g. Sample is from  $N(\mu,3)$ . We want to estimate  $\mu$ .

e.g. Sample is from Bernoulli(p). We want to estimate p.

We will treat population parameters as constants (as opposed to Bayesian statistics).

The goal is to use a statistic  $\hat{\theta}$ , i.e. a function of a sample, to estimate the value of a parameter  $\theta$ , which could be a vector. This statistic is called a "point estimator".

- e.g. Sample is from  $N(\mu,3)$ . We want to estimate  $\mu$ .
- e.g. Sample is from Bernoulli(p). We want to estimate p.
- e.g. Sample is from  $N(\mu, \sigma)$ . We want to estimate  $(\mu, \sigma)$ .

We will treat population parameters as constants (as opposed to Bayesian statistics).

The goal is to use a statistic  $\hat{\theta}$ , i.e. a function of a sample, to estimate the value of a parameter  $\theta$ , which could be a vector. This statistic is called a "point estimator".

- e.g. Sample is from  $N(\mu,3)$ . We want to estimate  $\mu$ .
- e.g. Sample is from Bernoulli(p). We want to estimate p.
- e.g. Sample is from  $N(\mu, \sigma)$ . We want to estimate  $(\mu, \sigma)$ .

Open questions about point estimators:

what desirable properties should they have?

We will treat population parameters as constants (as opposed to Bayesian statistics).

The goal is to use a statistic  $\hat{\theta}$ , i.e. a function of a sample, to estimate the value of a parameter  $\theta$ , which could be a vector. This statistic is called a "point estimator".

- e.g. Sample is from  $N(\mu,3)$ . We want to estimate  $\mu$ .
- e.g. Sample is from Bernoulli(p). We want to estimate p.
- e.g. Sample is from  $N(\mu, \sigma)$ . We want to estimate  $(\mu, \sigma)$ .

Open questions about point estimators:

- what desirable properties should they have?
  - ▶ how do I know which one to use? (To be addressed later.)

We would like  $\hat{\theta}$  to be *unbiased* ("for  $\theta$ "):

$$E(\hat{ heta})$$
 :

We would like  $\hat{\theta}$  to be *unbiased* ("for  $\theta$ "):

$$E(\hat{\theta}) = \theta$$

For example,  $\overline{X}$  is always unbiased for the mean  $\mu$  of any population, because we showed earlier that:

$$E(\overline{X}) = \mu$$

We would like  $\hat{\theta}$  to be *unbiased* ("for  $\theta$ "):

$$E(\hat{\theta}) = \theta$$

For example,  $\overline{X}$  is always unbiased for the mean  $\mu$  of any population, because we showed earlier that:

$$E(\overline{X}) = \mu$$

There can be lots of unbiased estimators. How to choose the best one? Take the one with the smallest variance.

We would like  $\hat{\theta}$  to be *unbiased* ("for  $\theta$ "):

$$E(\hat{\theta}) = \theta$$

For example,  $\overline{X}$  is always unbiased for the mean  $\mu$  of any population, because we showed earlier that:

$$E(\overline{X}) = \mu$$

There can be lots of unbiased estimators. How to choose the best one? Take the one with the smallest variance.

For example, if you have a sample  $X_1, \ldots, X_n$  from a  $N(\mu, \sigma)$  distribution, then  $\overline{X}$  is unbiased. But so is just taking  $X_1$ , say, because:

We would like  $\hat{\theta}$  to be *unbiased* ("for  $\theta$ "):

$$E(\hat{\theta}) = \theta$$

For example,  $\overline{X}$  is always unbiased for the mean  $\mu$  of any population, because we showed earlier that:

$$E(\overline{X}) = \mu$$

There can be lots of unbiased estimators. How to choose the best one? Take the one with the smallest variance.

For example, if you have a sample  $X_1, \ldots, X_n$  from a  $N(\mu, \sigma)$  distribution, then  $\overline{X}$  is unbiased. But so is just taking  $X_1$ , say, because:

That was a silly example.

A less silly exampe is that is possible to show that the sample median  $\tilde{X}$  is also unbiased for  $\mu$  when the sample is from a normal population.

That was a silly example.

A less silly example is that is possible to show that the sample median  $\tilde{X}$  is also unbiased for  $\mu$  when the sample is from a normal population.

It turns out (FIXED - the problem was the 4 should have been 2)  $Var(\tilde{X}) \approx \frac{\pi \sigma^2}{2n} \approx 1.57 Var(\overline{X})$ , in the  $N(\mu, \sigma)$  case.

That was a silly example.

A less silly exampe is that is possible to show that the sample median  $\tilde{X}$  is also unbiased for  $\mu$  when the sample is from a normal population.

It turns out (FIXED - the problem was the 4 should have been 2)

 $\mathsf{Var}ig( ilde{X}ig) pprox rac{\pi\sigma^2}{2n} pprox 1.57 \mathsf{Var}ig(\overline{X}ig)$ , in the  $\mathit{N}(\mu,\sigma)$  case.

So  $\overline{X}$  is preferred.

That was a silly example.

A less silly example is that is possible to show that the sample median  $\tilde{X}$  is also unbiased for  $\mu$  when the sample is from a normal population.

It turns out (FIXED - the problem was the 4 should have been 2)

$$\mathsf{Var}ig( ilde{X}ig)pprox rac{\pi\sigma^2}{2n}pprox 1.57 \mathsf{Var}ig(\overline{X}ig)$$
, in the  $\mathit{N}(\mu,\sigma)$  case.

So  $\overline{X}$  is preferred.

**Fact:** when the population is normal,  $\overline{X}$  is the unbiased estimator with the smallest variance.

That was a silly example.

A less silly exampe is that is possible to show that the sample median  $\tilde{X}$  is also unbiased for  $\mu$  when the sample is from a normal population.

It turns out (FIXED - the problem was the 4 should have been 2)  $Var(\tilde{X}) \approx \frac{\pi\sigma^2}{2n} \approx 1.57 Var(\overline{X})$ , in the  $N(\mu, \sigma)$  case.

So  $\overline{X}$  is preferred.

**Fact:** when the population is normal,  $\overline{X}$  is the unbiased estimator with the smallest variance.

Another desirable property (that  $\overline{X}$  has, for example) is *consistency*, which means the variance tends to 0 as  $n \to \infty$ 

Population:  $N(\mu, \sigma)$ . Sample:  $X_1, \ldots, X_n$ .

Paramater to estimate:  $\sigma^2$ .

Population:  $N(\mu, \sigma)$ . Sample:  $X_1, \ldots, X_n$ .

Paramater to estimate:  $\sigma^2$ .

Since:

$$\frac{n-1}{\sigma^2}S^2 \sim \chi_{n-1}^2$$

and the expected value of a  $\mathsf{Gamma}(\alpha,\lambda)$  is  $\frac{\alpha}{\lambda}$ , we get:

Population:  $N(\mu, \sigma)$ . Sample:  $X_1, \ldots, X_n$ .

Paramater to estimate:  $\sigma^2$ .

 $\frac{n-1}{\sigma^2}S^2 \sim \chi_{n-1}^2$ 

and the expected value of a 
$$\operatorname{Gamma}(\alpha,\lambda)$$
 is  $\frac{\alpha}{\lambda}$ , we get:

of a Gamma
$$(\alpha,\lambda)$$
 is  $\frac{\alpha}{\lambda}$ , we get:

 $E\left(\frac{n-1}{\sigma^2}S^2\right) = n-1$ 

$$-(n-1-2)$$

$$E\left(\frac{1}{\sigma^2}S^2\right) = n-1$$

$$\langle \sigma^2 \rangle$$

Population:  $N(\mu, \sigma)$ . Sample:  $X_1, \ldots, X_n$ .

and therefore  $E(S^2) = \sigma^2$ .

Paramater to estimate:  $\sigma^2$ .

Since: 
$$n-1$$

and the expected value of a Gamma $(\alpha, \lambda)$  is  $\frac{\alpha}{\lambda}$ , we get:

and the expected value of a Gamma
$$(\alpha, \lambda)$$
 is  $\frac{\alpha}{\lambda}$ , we get:

and the expected value of a Gamma
$$(lpha,\lambda)$$
 is  $rac{\pi}{\lambda}$ , we get:  $E\Big(rac{n-1}{\sigma^2}S^2\Big)=n-1$ 

 $\frac{n-1}{\sigma^2}S^2 \sim \chi_{n-1}^2$ 

and therefore  $E(S^2) = \sigma^2$ .

Since:

Population: 
$$N(\mu, \sigma)$$
. Sample:  $X_1$ ,  
Paramater to estimate:  $\sigma^2$ .

Population:  $N(\mu, \sigma)$ . Sample:  $X_1, \ldots, X_n$ .

Sample: 
$$X_1$$
,

ole: 
$$X_1$$
,

 $\frac{n-1}{\sigma^2}S^2 \sim \chi_{n-1}^2$ 

 $E\left(\frac{n-1}{\sigma^2}S^2\right) = n-1$ 

Sample: 
$$X_1, \dots$$

and the expected value of a Gamma $(\alpha, \lambda)$  is  $\frac{\alpha}{\lambda}$ , we get:

 $Exp(\lambda)$ 

How to estimate the rate parameter  $\lambda$ ?