

# STA286 Lecture 23

Neil Montgomery

Last edited: 2017-03-17 12:49

roadmap

## 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  $\bar{X}$

Now we are actually going to do statistics.

parameter estimation

## point estimation

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”.

## point estimation

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  $\mu$ .

## point estimation

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  $\text{Bernoulli}(p)$ . We want to estimate  $p$ .

## point estimation

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  $\text{Bernoulli}(p)$ . We want to estimate  $p$ .

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



## point estimation

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  $\text{Bernoulli}(p)$ . 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?

## point estimation

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  $\text{Bernoulli}(p)$ . 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.)

## bias and variance

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

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

## bias and variance

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

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

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

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

## bias and variance

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

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

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

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

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

## bias and variance

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

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

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

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

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  $X_1, \dots, X_n$  from a  $N(\mu, \sigma)$  distribution, then  $\bar{X}$  is unbiased. But so is just taking  $X_1$ , say, because:

$$E(X_1) = \mu$$

## bias and variance

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

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

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

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

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  $X_1, \dots, X_n$  from a  $N(\mu, \sigma)$  distribution, then  $\bar{X}$  is unbiased. But so is just taking  $X_1$ , say, because:

$$E(X_1) = \mu$$

But  $\text{Var}(\bar{X}) = \sigma^2/n$  which is smaller than  $\text{Var}(X_1) = \sigma^2$ .

## bias and variance

That was a silly example.

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



## bias and variance

That was a silly example.

A less silly example is that it 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)

$\text{Var}(\tilde{X}) \approx \frac{\pi\sigma^2}{2n} \approx 1.57\text{Var}(\bar{X})$ , in the  $N(\mu, \sigma)$  case.

## bias and variance

That was a silly example.

A less silly example is that it 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)

$\text{Var}(\tilde{X}) \approx \frac{\pi\sigma^2}{2n} \approx 1.57\text{Var}(\bar{X})$ , in the  $N(\mu, \sigma)$  case.

So  $\bar{X}$  is preferred.

## bias and variance

That was a silly example.

A less silly example is that it 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)

$\text{Var}(\tilde{X}) \approx \frac{\pi\sigma^2}{2n} \approx 1.57\text{Var}(\bar{X})$ , in the  $N(\mu, \sigma)$  case.

So  $\bar{X}$  is preferred.

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

## bias and variance

That was a silly example.

A less silly example is that it 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)

$\text{Var}(\tilde{X}) \approx \frac{\pi\sigma^2}{2n} \approx 1.57\text{Var}(\bar{X})$ , in the  $N(\mu, \sigma)$  case.

So  $\bar{X}$  is preferred.

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

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

## another unbiased estimator

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

Parameter to estimate:  $\sigma^2$ .

## another unbiased estimator

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

Parameter to estimate:  $\sigma^2$ .

Since:

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

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

## another unbiased estimator

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

Parameter to estimate:  $\sigma^2$ .

Since:

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

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

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

## another unbiased estimator

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

Parameter to estimate:  $\sigma^2$ .

Since:

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

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

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

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



## another unbiased estimator

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

Parameter to estimate:  $\sigma^2$ .

Since:

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

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

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

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

This explains the embarrassing  $n-1$  in the denominator of  $S^2$ .

$Exp(\lambda)$

How to estimate the rate parameter  $\lambda$ ?