STA286 Lecture 23

Neil Montgomery

Last edited: 2017-03-16 13:06

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 $\overline{X}$

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 θ, 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 θ, which could be a vector. This statistic is called a "point estimator".

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

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 θ, which could be a vector. This statistic is called a "point estimator".

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

e.g. Sample is from 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 θ, which could be a vector. This statistic is called a "point estimator".

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

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

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

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 θ, which could be a vector. This statistic is called a "point estimator".

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

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

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

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 θ, which could be a vector. This statistic is called a "point estimator".

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

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

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

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 θ"):

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

bias and variance

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

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

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

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

bias and variance

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

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

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

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

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 θ"):

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

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

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

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:

$$E(X_1) = \mu$$

bias and variance

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

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

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

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

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:

$$E(X_1) = \mu$$

But $\mathrm{Var}\left(\overline{X}\right) = \sigma^2/n$ which is smaller than $\mathrm{Var}(X_1) = \sigma^2$.

bias and 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 μ when the sample is from a normal population.

bias and 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 μ when the sample is from a normal population.

But $\text{Var}\left(\overline{X}\right) < \text{Var}\left(\tilde{X}\right)$ so $\overline{X}$ is preferred.

bias and 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 μ when the sample is from a normal population.

But $\text{Var}\left(\overline{X}\right) < \text{Var}\left(\tilde{X}\right)$ so $\overline{X}$ is preferred.

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

bias and 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 μ when the sample is from a normal population.

But $\text{Var}\left(\overline{X}\right) < \text{Var}\left(\tilde{X}\right)$ 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 $n \to \infty$.

another unbiased estimator

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

Paramater to estimate: σ^2.

another unbiased estimator

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

Paramater to estimate: σ^2.

Since:

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

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

another unbiased estimator

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

Paramater to estimate: σ^2.

Since:

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

and the expected value of a Gamma(α, λ) 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: $X_1, \ldots, X_n$.

Paramater to estimate: σ^2.

Since:

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

and the expected value of a Gamma(α, λ) 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: $X_1, \ldots, X_n$.

Paramater to estimate: σ^2.

Since:

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

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

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

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

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

$Exp(\lambda)$

How to estimate the rate parameter λ?