

Parameter Estimation

STAT 330 - Iowa State University

In this lecture students will be introduced to parameter estimation. We look at the definitions of estimators and estimates. Students will learn about different metrics that make “good” estimators, and we introduce the idea of a statistical model.

Inference Overview

Topics:

1. Estimation of parameters
2. Confidence intervals
3. Hypothesis testing
4. Prediction

Estimation

Estimator

Start with $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x)$, where $f_X(x)$ is some distribution with some parameter(s) θ . "population"
 $\theta = \lambda, \quad \underline{\theta} = (\mu, \sigma^2)$

In statistics, θ is unknown, so we need to *estimate* it.

Definition

An estimator is a statistic, $T(X_1, \dots, X_n)$, that is used to learn about an unknown parameter θ .

- The term "estimator" refers to the statistic as a function of random variables X_1, \dots, X_n . Estimator \equiv R.V.
- Estimators usually get "hats".
 $\rightarrow \hat{\theta}$ is an estimator of θ .

Estimate

Definition

An estimate is the observed value of the statistic used to learn about an unknown parameter.

- The term “estimate” refers to the statistic as a function of the observed data x_1, \dots, x_n
- It is a numeric value



Example 1: $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x)$ with some $E(X) = \mu$ (unknown).

You observe values 6, 7, 7, 8, 9, 10

- $\hat{\mu} = \bar{X}$ is an *estimator* of μ
- $\bar{x} = \frac{6+7+7+8+9+10}{6} = 7.83$ is an estimate of μ

Sampling Distribution of the Estimator

- Since the estimator $\hat{\theta}$ is a function of R.V's, it is also considered a R.V.
- Estimators have their own distribution called the sampling distribution of $\hat{\theta}$
 - The mean of the sampling distribution is $E(\hat{\theta})$
 - The standard deviation of the sampling distribution is called the "standard error" $= se(\hat{\theta}) = \sqrt{var(\hat{\theta})}$
- We will make use of the sampling distribution in confidence intervals and hypothesis testing?

$X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x)$

$$\mu = \bar{X} \leftarrow$$

CLT

sampling dist. of \bar{X} is
 $\approx N(\mu, \sigma^2/n)$



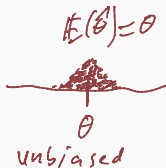
Properties of Estimators

Properties of Estimators

How to tell if our estimator is “good”?

There are some properties of estimators we can look at:

- unbiasedness
- consistency
- mean squared error



Definition

The **bias** of an estimator $\hat{\theta}$ is $Bias(\hat{\theta}) = E(\hat{\theta} - \theta)$.

An estimator $\hat{\theta}$ is **unbiased** if $Bias(\hat{\theta}) = E(\hat{\theta} - \theta) = 0 \Rightarrow E(\hat{\theta}) = \theta$

Definition: Consistent

An estimator $\hat{\theta}$ is a **consistent** estimator of θ if

$$\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| > \epsilon) = 0$$

Unbiased and Consistent Estimators

Earlier in the notes, we said that we should use

- \bar{X} as an estimator for $E(X) = \mu$
 - S^2 as an estimator for $Var(X) = \sigma^2$
- Handwritten note: $\frac{1}{n-1}$ version*

Theorem

\bar{X} and S^2 are both *unbiased* and *consistent* estimators for parameters μ and σ^2 respectively.

Proof: Let $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x)$ with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$
Unbiasedness:

$$E(\bar{X}) = E\left(\frac{1}{n} \sum X_i\right) = \frac{1}{n} E\left(\sum X_i\right) = \frac{1}{n} \sum E(X_i) = \frac{1}{n} \cdot n \cdot \mu = \mu$$

Consistency:

$$P(|\bar{X} - \mu| > \epsilon) \leq \frac{Var(\bar{X})}{\epsilon^2} = \frac{\sigma^2}{n \cdot \epsilon^2} \rightarrow 0 \text{ as } n \rightarrow \infty$$

Handwritten note: σ^2 and $n \cdot \epsilon^2$ are circled with arrows pointing to the denominator.

Mean Squared Error

A popular metric for comparing different estimators is the mean squared error (MSE).

Definition: Mean Squared Error (MSE)

The *mean squared error (MSE)* of an estimator is

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

- It can be shown that $MSE(\hat{\theta}) = Bias^2(\hat{\theta}) + Var(\hat{\theta})$
- This is usually easier to calculate
- Ideally, we want estimator to have small MSE (with small bias and small variance).

Example

Example 2: $X_1, \dots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$. Want estimators for μ and σ^2 .

Consider two estimators for μ :

1. $\hat{\mu}_1 = X_1$
2. $\hat{\mu}_2 = \bar{X}$

Both estimators have sampling distribution that are normal dist.

Both estimators are unbiased

- $E(X_1) = \mu$
 $\rightarrow \text{Bias}(X_1) = E(X_1 - \mu) = E(X_1) - \mu = \mu - \mu = 0$
- $E(\bar{X}) = \mu$
 $\rightarrow \text{Bias}(\bar{X}) = E(\bar{X} - \mu) = E(\bar{X}) - \mu = \mu - \mu = 0$

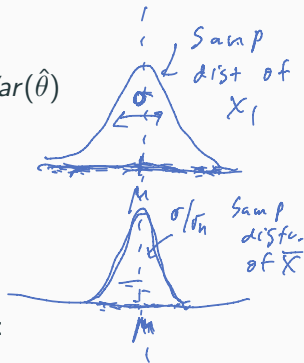
Example Cont.

Compare the MSE of both estimators

$$\text{Recall } \text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = \text{Bias}^2(\hat{\theta}) + \text{Var}(\hat{\theta})$$

Variance of the estimators:

- $\text{Var}(X_1) = \sigma^2$
- $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$



Mean squared errors (MSE) of the estimators:

- $\text{MSE}(X_1) = \text{Bias}^2(X_1) + \text{Var}(X_1) = 0^2 + \sigma^2$
- $\text{MSE}(\bar{X}) = \text{Bias}^2(\bar{X}) + \text{Var}(\bar{X}) = 0^2 + \frac{\sigma^2}{n}$

$\text{MSE}(\bar{X}) < \text{MSE}(X_1) \rightarrow \bar{X}$ is the “better” estimator for μ

Statistical Model

Statistical Models

We want a model for our sample to use for making inference

Definition

A *statistical model* is the joint distribution of our sample.

Recall:

- We've seen the joint distribution for 2 discrete R.V's:

$$P_{X,Y}(x,y) = P(X = x, Y = y)$$

- If X, Y are independent, the the joint distribution can be written as

$$\begin{aligned}\underline{P_{X,Y}(x,y)} &= P(X = x, Y = y) \\ &= P(X = x) \cdot P(Y = y) \\ &= \underline{P_X(x) \cdot P_Y(y)}\end{aligned}$$

Statistical Model Cont.

Let $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x)$.

Identical \Rightarrow marginals are all the same $f_X(x)$
Independent

The joint distribution of our sample is

$$f(x_1, \dots, x_n) = \prod_{i=1}^n \underline{f_X(x_i)}$$

We can use the statistical model and data to obtain a single estimate (point estimate) for the parameter(s) in our model.

- In statistics, this is called “fitting” the model (using “data”)
- In machine learning, this is called “learning” the model (using “training data”)

Example

Example 3: Let $X_i = \#$ of goals scored by the ISU womens's soccer team in game i .

A diagram illustrating the relationship between data and a generating mechanism. On the left, a blue arrow points to the expression X_1, \dots, X_n , which is underlined in blue. To the right of this is a tilde symbol \sim followed by $f_X(x)$. The expression $f_X(x)$ is circled in blue. A blue arrow points from the handwritten text "data (goals)" to the circle around $f_X(x)$. Below "data (goals)" is the handwritten text "generating mechanism".

$$\underline{X_1, \dots, X_n} \sim f_X(x)$$

data (goals)
generating mechanism

We are interested in the probability the team scores more than 2 goals in a game.

How do we approach solving this problem?

1. Come up with a model for the sample.
2. Estimate the parameters of the model
3. Use fitted model to estimate the probability of scoring more than 2 goals.

Example Cont.

- $X_i = \#$ of goals scored by the the soccer team in game i .
→ X_i 's are discrete random variables
- A reasonable model is then the Poisson distribution
- The joint distribution is

$$\begin{aligned} f(x_1, \dots, x_n) &= \prod_{i=1}^n f_X(x_i) \\ &= \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \\ &= \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \end{aligned}$$

Example Cont.

- Since, for Poisson distribution, $E(X) = \lambda$, it makes sense to use the *estimator* \bar{X} for λ .
- Observed values: 0, 0, 1, 0, 1, 2, 2, 0, 1, 1
- My *estimate* of λ :

$$\bar{x} = \frac{0+0+\dots+1}{10} = 8/10 = .8$$

$$\hat{\lambda} = .8$$

$$X_i \sim \text{pois}(.8)$$

$$\begin{aligned} &\Rightarrow P(X > 2) \\ &= 1 - P(X \leq 2) \\ &= \dots \end{aligned}$$

Recap

Students should now know the difference between an estimator (function of random variables) and an estimate (function with the observed values plugged in). They should be able to compute the bias of an estimator and compared mean squared errors. They should be familiar with the idea of the joint distribution of a sample and a statistical model.