Lecture 20

Parameter Estimation

STAT 330 - Iowa State University

Inference Overview

Topics:

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

Estimation

Estimator

Start with $X_1, \ldots, X_n \overset{iid}{\sim} f_X(x)$, where $f_X(x)$ is *some* distribution with *some* parameter(s) θ

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

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

- The term "estimator" refers to the statistic as a function of random variables $X_1, \ldots, X_n$
- 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 $x_1, \ldots, x_n$
- It is a numeric value

Example 1: $X_1, \ldots, X_n \overset{iid}{\sim} f_X(x)$ 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

(random variable)

- 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 $\hat{\theta}$
  - $\rightarrow$ The *mean* of the sampling distribution is $E(\hat{\theta})$
  - $\rightarrow$ 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

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

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

//

Definition

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

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

Definition: Consistent

An estimator $\hat{\theta}$ is a *consistent* estimator of θ if (for any $\varepsilon > 0$)

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

(as $n\uparrow$, there is high probability that $\hat{\theta}$ is close to θ.)

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 $Var(X) = \sigma^2$

Theorem

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

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

for $\bar{X}$

Unbiasedness:

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

Consistency:

$$\underbrace{P(|\bar{X} - \mu| > \epsilon) \leq \frac{Var(\bar{X})}{\epsilon^2}}_{\text{Chebychev's Inequality}} = \frac{\sigma^2}{n \cdot \epsilon^2} \to 0 \text{ as } n \to \infty$$

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

Example

← independent & identically distributed.

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

Consider two estimators for μ:

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

} which one is better?

$X_1 \sim N(\mu, \sigma^2)$

$\bar{X} \sim N(\mu, \sigma^2/n)$ by CLT

Both estimators have sampling distribution that are normal dist.

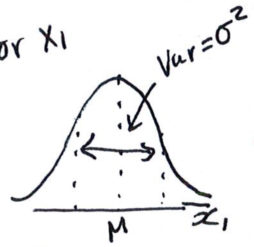
Both estimators are *unbiased*

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

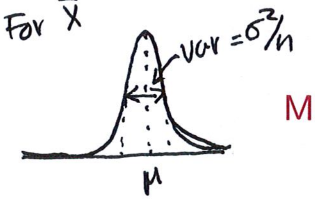
} both estimators have bias $= 0$

Example Cont.

For X_1



For $\bar{X}$



Compare the MSE of both estimators

Recall $MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = \underbrace{Bias^2(\hat{\theta})}_{\text{already have this}} + \underbrace{Var(\hat{\theta})}_{?}$

Variance of the estimators:

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

Mean squared errors (MSE) of the estimators:

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

$MSE(\bar{X}) < MSE(X_1) \rightarrow \bar{X}$ 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

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

when R.Vs are independent joint = product of marginal

Statistical Model Cont.

Let $X_1, \ldots, X_n \overset{iid}{\sim} f_X(x)$. ← indep. & identically distributed.

The joint distribution of our sample is

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

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.

$$X_1, \ldots, X_n \overset{iid}{\sim} f_X(x)$$

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.

  $\rightarrow X_i$'s are discrete random variables

$X_1 \ldots X_n \overset{iid}{\sim} Pois(\lambda)$

- A reasonable model is then the Poisson distribution

For each x_i, $f_X(x_i) = \dfrac{e^{-\lambda}\lambda^{x_i}}{x_i!}$

- The *joint distribution* is

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

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:

Estimator of λ : $\hat{\lambda} = \bar{X}$

Estimate of λ : $\hat{\lambda} = \bar{x} = 0.8$

Now we can assume a model $X \sim Pois(\lambda = 0.8)$

What is the probability of scoring more than 2 points?

$$P(X > 2) = 1 - P(X \leq 2)$$
$$= 0.047$$