

# Lecture 20

## Parameter Estimation

---

STAT 330 - Iowa State University

## Topics:

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

# Estimation

---

Start with  $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x)$ , where  $f_X(x)$  is *some* distribution with *some* parameter(s)  $\theta$

In statistics,  $\theta$  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  $\theta$ .

- The term “estimator” refers to the statistic as a function of random variables  $X_1, \dots, X_n$
- Estimators usually get “hats”.  
→  $\hat{\theta}$  is an estimator of  $\theta$ .

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

# 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

# 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  $Bias(\hat{\theta}) = E(\hat{\theta} - \theta) = 0$ .

## Definition: Consistent

An estimator  $\hat{\theta}$  is a *consistent* estimator of  $\theta$  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  $\text{Var}(X) = \sigma^2$

## Theorem

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

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

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:

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

# 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  $MSE(\hat{\theta}) = Bias^2(\hat{\theta}) + Var(\hat{\theta})$
- 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  $\mu$  and  $\sigma^2$ .

Consider two estimators for  $\mu$ :

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

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

Variance of the estimators:

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

Mean squared errors (MSE) of the estimators:

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

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

# 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

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

## Statistical Model Cont.

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

The joint distribution of our sample is

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

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  $i$ .

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

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  $i$ .  
→  $X_i$ '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  $\lambda$ .
- Observed values: 0, 0, 1, 0, 1, 2, 2, 0, 1, 1
- My *estimate* of  $\lambda$ :