YSU ASDS, Statistics, Fall 2019

Lecture 15

Michael Poghosyan

09 Oct 2019

Contents

Last Lecture ReCap

- Aystegh Karogh e Linel Dzer Govazd@

Last Lecture ReCap

- Aystegh Karogh e Linel Dzer Govazd@ ◡̈

Point Estimates

General Setting for the Point Estimation

The general setting for the Point Estimation is the following:

General Setting for the Point Estimation

The general setting for the Point Estimation is the following:

- We have a Dataset: $x : x_1, x_2, ..., x_n$;

General Setting for the Point Estimation

The general setting for the Point Estimation is the following:

- We have a Dataset: $x : x_1, x_2, ..., x_n$;
- We model the process that generated x by the Parametric Family of Distributions $\{\mathcal{F}_\theta, \theta \in \Theta\}$;

General Setting for the Point Estimation

The general setting for the Point Estimation is the following:

- We have a Dataset: $x : x_1, x_2, ..., x_n$;
- We model the process that generated x by the Parametric Family of Distributions $\{\mathcal{F}_\theta, \theta \in \Theta\}$;
- We assume the Dataset above is one possible realization of the Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta \qquad \text{for some} \quad \theta \in \Theta;$$

General Setting for the Point Estimation

The general setting for the Point Estimation is the following:

- We have a Dataset: $x : x_1, x_2, ..., x_n$;
- We model the process that generated x by the Parametric Family of Distributions $\{\mathcal{F}_\theta, \theta \in \Theta\}$;
- We assume the Dataset above is one possible realization of the Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta \qquad \text{for some} \quad \theta \in \Theta;$$

Problem: Estimate the unknown Parameter θ.

Motivating Example $\smile$̈

Motivating Example ☺

Example: I have generated the following Data from a Normal Distribution:

```
##        [,1]   [,2]  [,3]   [,4]   [,5]  [,6]
## [1,]  5.568 -3.370 -1.18 -0.937 -0.355 -4.78
## [2,] -0.094 -6.912 -8.74 -1.740  1.915  2.84
## [3,] -4.845  1.969 -8.14 -1.615  3.497 -3.17
## [4,] -4.714  2.517 -3.83 -9.615 -3.439 -8.25
## [5,] -1.472  0.392 -9.26  4.639 -4.623  1.73
```

Motivating Example $\ddot\smile$

Example: I have generated the following Data from a Normal Distribution:

```
##         [,1]   [,2]  [,3]   [,4]   [,5]  [,6]
## [1,]  5.568 -3.370 -1.18 -0.937 -0.355 -4.78
## [2,] -0.094 -6.912 -8.74 -1.740  1.915  2.84
## [3,] -4.845  1.969 -8.14 -1.615  3.497 -3.17
## [4,] -4.714  2.517 -3.83 -9.615 -3.439 -8.25
## [5,] -1.472  0.392 -9.26  4.639 -4.623  1.73
```

Question: Find/Estimate the Parameter values I was using.

Motivating Example ☺

Example: I have generated the following Data from a Normal Distribution:

```
##         [,1]   [,2]  [,3]   [,4]   [,5]  [,6]
## [1,]  5.568 -3.370 -1.18 -0.937 -0.355 -4.78
## [2,] -0.094 -6.912 -8.74 -1.740  1.915  2.84
## [3,] -4.845  1.969 -8.14 -1.615  3.497 -3.17
## [4,] -4.714  2.517 -3.83 -9.615 -3.439 -8.25
## [5,] -1.472  0.392 -9.26  4.639 -4.623  1.73
```

Question: Find/Estimate the Parameter values I was using.

Moral: Statistics is like a Detective Story: you need to find the Unknown (murderer?) using some small amount of Observations, Data you have ☺

Statistics, Estimator and Estimate

Let us recall what is our Problem: assume we have a Dataset $x_1, ..., x_n$. We assume that this is a realization of a Random Sample $X_1, ..., X_n$, coming from one of the Distributions from some Parametric Family:

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta, \qquad \theta \in \Theta.$$

And our problem is to estimate the value of our unknown Parameter θ.

Statistics, Estimator and Estimate

Let us recall what is our Problem: assume we have a Dataset $x_1, ..., x_n$. We assume that this is a realization of a Random Sample $X_1, ..., X_n$, coming from one of the Distributions from some Parametric Family:

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta, \qquad \theta \in \Theta.$$

And our problem is to estimate the value of our unknown Parameter θ.

To estimate θ, we will use only $X_1, ..., X_n$ (or, $x_1, x_2, ..., x_n$), since we do not have any other thing.

Statistics, Estimator and Estimate

Let us recall what is our Problem: assume we have a Dataset $x_1, ..., x_n$. We assume that this is a realization of a Random Sample $X_1, ..., X_n$, coming from one of the Distributions from some Parametric Family:

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta, \qquad \theta \in \Theta.$$

And our problem is to estimate the value of our unknown Parameter θ.

To estimate θ, we will use only $X_1, ..., X_n$ (or, $x_1, x_2, ..., x_n$), since we do not have any other thing. Now,

Definition: Any (measurable) function of the Random Sample $X_1, X_2, ..., X_n$ is called a **Statistics**. So Statistics is a r.v. of the form

$$g(X_1, X_2, ..., X_n).$$

Statistics, Estimator and Estimate

Let us recall what is our Problem: assume we have a Dataset $x_1, ..., x_n$. We assume that this is a realization of a Random Sample $X_1, ..., X_n$, coming from one of the Distributions from some Parametric Family:

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta, \qquad \theta \in \Theta.$$

And our problem is to estimate the value of our unknown Parameter θ.

To estimate θ, we will use only $X_1, ..., X_n$ (or, $x_1, x_2, ..., x_n$), since we do not have any other thing. Now,

Definition: Any (measurable) function of the Random Sample $X_1, X_2, ..., X_n$ is called a **Statistics**. So Statistics is a r.v. of the form

$$g(X_1, X_2, ..., X_n).$$

This is our third meaning of the term *Statistics*.

Example

Example: For example, the followings are Statistics:

$$X_1, \qquad \frac{X_1 + X_n}{3}, \qquad \sin(X_1 \cdot X_2 + X_3 + ... + X_n).$$

Example

Example: For example, the followings are Statistics:

$$X_1, \qquad \frac{X_1 + X_n}{3}, \qquad \sin(X_1 \cdot X_2 + X_3 + ... + X_n).$$

Definition: The Distribution of the Statistics $g(X_1, X_2, ..., X_n)$ is called a **Sampling Distribution**.

Example

Example: For example, the followings are Statistics:

$$X_1, \qquad \frac{X_1 + X_n}{3}, \qquad \sin(X_1 \cdot X_2 + X_3 + ... + X_n).$$

Definition: The Distribution of the Statistics $g(X_1, X_2, ..., X_n)$ is called a **Sampling Distribution**.

Example: The Sampling Distribution of the Statistics

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}$$

is

Example

Example: For example, the followings are Statistics:

$$X_1, \qquad \frac{X_1 + X_n}{3}, \qquad \sin(X_1 \cdot X_2 + X_3 + ... + X_n).$$

Definition: The Distribution of the Statistics $g(X_1, X_2, ..., X_n)$ is called a **Sampling Distribution**.

Example: The Sampling Distribution of the Statistics

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}$$

is almost Normal, for large n, by the CLT.

Statistics, Estimator and Estimate

Assume we want to estimate the value of the parameter $\theta \in \Theta$, and we will use the Statistics $g(X_1, ..., X_n)$ for that.

Statistics, Estimator and Estimate

Assume we want to estimate the value of the parameter $\theta \in \Theta$, and we will use the Statistics $g(X_1, ..., X_n)$ for that.

Definition: If

- $g : \mathbb{R}^n \to \Theta$;
- g doesn't depend on the unknown θ;

then the Statistics $g(X_1, X_2, ..., X_n)$ is called an **Estimator** for θ, and it is usually denoted by

$$\hat{\theta} = \hat{\theta}_n = g(X_1, X_2, ..., X_n).$$

Statistics, Estimator and Estimate

Assume we want to estimate the value of the parameter $\theta \in \Theta$, and we will use the Statistics $g(X_1, ..., X_n)$ for that.

Definition: If

- $g : \mathbb{R}^n \to \Theta$;
- g doesn't depend on the unknown θ;

then the Statistics $g(X_1, X_2, ..., X_n)$ is called an **Estimator** for θ, and it is usually denoted by

$$\hat{\theta} = \hat{\theta}_n = g(X_1, X_2, ..., X_n).$$

The value of the Estimator at our observations, $g(x_1, x_2, ..., x_n)$, is called an **Estimate** for θ, and it is again (unfortunately) denoted by $\hat{\theta} = \hat{\theta}_n$.

Example

Example: Say, we want to estimate the parameter λ in the

$$\{Exp(\lambda): \ \lambda > 0\}$$

model, using the Random Sample

$$X_1, X_2, ..., X_n \overset{IID}{\sim} Exp(\lambda).$$

Example

Example: Say, we want to estimate the parameter λ in the

$$\{Exp(\lambda): \ \lambda > 0\}$$

model, using the Random Sample

$$X_1, X_2, ..., X_n \overset{IID}{\sim} Exp(\lambda).$$

Then the followings are **Esimators**:

- $\hat{\lambda} = \dfrac{X_1}{3}$;

Example

Example: Say, we want to estimate the parameter λ in the

$$\{Exp(\lambda) : \ \lambda > 0\}$$

model, using the Random Sample

$$X_1, X_2, ..., X_n \overset{IID}{\sim} Exp(\lambda).$$

Then the followings are **Esimators**:

- $\hat{\lambda} = \dfrac{X_1}{3}$;
- $\hat{\lambda} = \dfrac{X_1 + X_2 + ... + X_n}{n}$;

Example

Example: Say, we want to estimate the parameter λ in the

$$\{Exp(\lambda): \ \lambda > 0\}$$

model, using the Random Sample

$$X_1, X_2, ..., X_n \stackrel{IID}{\sim} Exp(\lambda).$$

Then the followings are **Esimators**:

- $\hat{\lambda} = \dfrac{X_1}{3}$;
- $\hat{\lambda} = \dfrac{X_1 + X_2 + ... + X_n}{n}$;
- $\hat{\lambda} = (X_1 + X_2 + ... + X_n)^2$;

Example

Example: Say, we want to estimate the parameter λ in the

$$\{Exp(\lambda): \ \lambda > 0\}$$

model, using the Random Sample

$$X_1, X_2, ..., X_n \overset{IID}{\sim} Exp(\lambda).$$

Then the followings are **Esimators**:

- $\hat{\lambda} = \dfrac{X_1}{3}$;
- $\hat{\lambda} = \dfrac{X_1 + X_2 + ... + X_n}{n}$;
- $\hat{\lambda} = (X_1 + X_2 + ... + X_n)^2$;

And the following is not an estimator:

- $\hat{\lambda} = \dfrac{\lambda}{X_1 + X_n}$, since it depends on λ - the unkown parameter value.

Estimators and Estimates

Note: We require our Estimator to be independent of the Parameter θ, since we want to be able to calculate the value of the Estimator on the observation, i.e., we want to be able to calculate the Estimate, we want to make our Estimator *observable*. Since θ is unknown, then we cannot use it in the Estimator - we will not be able to calculate the Estimate.

Estimators and Estimates

Note: We require our Estimator to be independent of the Parameter θ, since we want to be able to calculate the value of the Estimator on the observation, i.e., we want to be able to calculate the Estimate, we want to make our Estimator *observable*. Since θ is unknown, then we cannot use it in the Estimator - we will not be able to calculate the Estimate.

Note: Again,

- **Estimator** is a function of our Random Sample, it is a r.v.. It doesn't depend on the unknown Parameter. Plugging the values of our observation into the Estimator we will get the Estimate

Estimators and Estimates

Note: We require our Estimator to be independent of the Parameter θ, since we want to be able to calculate the value of the Estimator on the observation, i.e., we want to be able to calculate the Estimate, we want to make our Estimator *observable*. Since θ is unknown, then we cannot use it in the Estimator - we will not be able to calculate the Estimate.

Note: Again,

- **Estimator** is a function of our Random Sample, it is a r.v.. It doesn't depend on the unknown Parameter. Plugging the values of our observation into the Estimator we will get the Estimate
- **Estimate** is a number, it is the result of plugging the observation into the Estimator.

Example

Example: Assume we want to estimate the probability of a birth (birthrate) of a girl child in Armenia. To that end, say, we ask a maternity hospital to provide the birth data for a day. And say, our data is:

$$b, g, g, b, b, g, b$$

Example

Example: Assume we want to estimate the probability of a birth (birthrate) of a girl child in Armenia. To that end, say, we ask a maternity hospital to provide the birth data for a day. And say, our data is:

$$b, g, g, b, b, g, b$$

First we need to build a Model. To that end, we encode our data by

$$x: \; 0, 1, 1, 0, 0, 1, 0,$$

where $b = 0$ and $g = 1$:

Example

Example: Assume we want to estimate the probability of a birth (birthrate) of a girl child in Armenia. To that end, say, we ask a maternity hospital to provide the birth data for a day. And say, our data is:

$$b, g, g, b, b, g, b$$

First we need to build a Model. To that end, we encode our data by

$$x: \;\; 0, 1, 1, 0, 0, 1, 0,$$

where $b = 0$ and $g = 1$: this is to be able to use one of our standard Distributions.

Example

Example: Assume we want to estimate the probability of a birth (birthrate) of a girl child in Armenia. To that end, say, we ask a maternity hospital to provide the birth data for a day. And say, our data is:

$$b, g, g, b, b, g, b$$

First we need to build a Model. To that end, we encode our data by

$$x: \ 0, 1, 1, 0, 0, 1, 0,$$

where $b = 0$ and $g = 1$: this is to be able to use one of our standard Distributions. Next, from a Dataset we pass, for a generalization, to a Random Sample

$$X_1, X_2, X_3, X_4, X_5, X_6, X_7,$$

where X_k is the gender of the k-th child *before the observation was made* ($X_k = 1$ if the child will be a girl, and 0 otherwise).

Example, cont'd

Then we will have

$$X_1, X_2, X_3, X_4, X_5, X_6, X_7 \sim Bernoulli(p),$$

where p is the Probability of having a girl child.

Example, cont'd

Then we will have

$$X_1, X_2, X_3, X_4, X_5, X_6, X_7 \sim \mathit{Bernoulli}(p),$$

where p is the Probability of having a girl child. So our Model will be

$$\{\mathit{Bernoulli}(p) : p \in [0, 1]\},$$

and our task will be to estimate p based on our Sample x.

Example, cont'd

Then we will have

$$X_1, X_2, X_3, X_4, X_5, X_6, X_7 \sim Bernoulli(p),$$

where p is the Probability of having a girl child. So our Model will be

$$\{Bernoulli(p) : p \in [0, 1]\},$$

and our task will be to estimate p based on our Sample x.

To estimate p, let us take the following **Estimator**:

$$\hat{p} = \frac{X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7}{7}.$$

Example, cont'd

Then we will have

$$X_1, X_2, X_3, X_4, X_5, X_6, X_7 \sim Bernoulli(p),$$

where p is the Probability of having a girl child. So our Model will be

$$\{Bernoulli(p) : p \in [0, 1]\},$$

and our task will be to estimate p based on our Sample x.

To estimate p, let us take the following **Estimator**:

$$\hat{p} = \frac{X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7}{7}.$$

This is a r.v. . The **Estimate** for p, using our Observation, will be

$$\hat{p} = \frac{0 + 1 + 1 + 0 + 0 + 1 + 0}{7} = \frac{3}{7}.$$

Point Estimation Setup

In the Inferential Statistics part, for the Point Estimation of the Parameter(s), we usually do the following steps, after describing the problem (and, maybe after obtaining Data):

Point Estimation Setup

In the Inferential Statistics part, for the Point Estimation of the Parameter(s), we usually do the following steps, after describing the problem (and, maybe after obtaining Data):

- First we specify the Model and unknown Parameter(s);

Point Estimation Setup

In the Inferential Statistics part, for the Point Estimation of the Parameter(s), we usually do the following steps, after describing the problem (and, maybe after obtaining Data):

- First we specify the Model and unknown Parameter(s);
- Then we find Estimators for our Parameter(s);

Point Estimation Setup

In the Inferential Statistics part, for the Point Estimation of the Parameter(s), we usually do the following steps, after describing the problem (and, maybe after obtaining Data):

- First we specify the Model and unknown Parameter(s);
- Then we find Estimators for our Parameter(s);
- Prove that the Estimators found are good ones;

Point Estimation Setup

In the Inferential Statistics part, for the Point Estimation of the Parameter(s), we usually do the following steps, after describing the problem (and, maybe after obtaining Data):

- First we specify the Model and unknown Parameter(s);
- Then we find Estimators for our Parameter(s);
- Prove that the Estimators found are good ones;
- Using Estimators and Data, calculate the Estimate for the Parameter(s).

Point Estimation Setup

In the Inferential Statistics part, for the Point Estimation of the Parameter(s), we usually do the following steps, after describing the problem (and, maybe after obtaining Data):

- First we specify the Model and unknown Parameter(s);
- Then we find Estimators for our Parameter(s);
- Prove that the Estimators found are good ones;
- Using Estimators and Data, calculate the Estimate for the Parameter(s).

In the next few lectures, we will consider what it means that an Estimator is a good one.

Point Estimation Setup

In the Inferential Statistics part, for the Point Estimation of the Parameter(s), we usually do the following steps, after describing the problem (and, maybe after obtaining Data):

- First we specify the Model and unknown Parameter(s);
- Then we find Estimators for our Parameter(s);
- Prove that the Estimators found are good ones;
- Using Estimators and Data, calculate the Estimate for the Parameter(s).

In the next few lectures, we will consider what it means that an Estimator is a good one. Later, we will consider some general methods to find good Estimators.

General Setting for the Point Estimation, Updated

In more Mathematical terms, the general setting for the Point Estimation is the following:

General Setting for the Point Estimation, Updated

In more Mathematical terms, the general setting for the Point Estimation is the following:

- We have a Dataset: $x : x_1, x_2, ..., x_n$;

General Setting for the Point Estimation, Updated

In more Mathematical terms, the general setting for the Point Estimation is the following:

- We have a Dataset: $x : x_1, x_2, ..., x_n$;
- We model the process that generated x by the Parametric Family of Distributions $\{\mathcal{F}_\theta, \theta \in \Theta\}$;

General Setting for the Point Estimation, Updated

In more Mathematical terms, the general setting for the Point Estimation is the following:

- We have a Dataset: $x : x_1, x_2, ..., x_n$;
- We model the process that generated x by the Parametric Family of Distributions $\{\mathcal{F}_\theta, \theta \in \Theta\}$;
- We assume the Dataset above is one possible realization of the Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta \qquad \text{for some} \quad \theta \in \Theta;$$

General Setting for the Point Estimation, Updated

In more Mathematical terms, the general setting for the Point Estimation is the following:

- We have a Dataset: $x : x_1, x_2, ..., x_n$;
- We model the process that generated x by the Parametric Family of Distributions $\{\mathcal{F}_\theta, \theta \in \Theta\}$;
- We assume the Dataset above is one possible realization of the Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta \qquad \text{for some} \quad \theta \in \Theta;$$

Problem: Estimate the unknown Parameter θ.

General Setting for the Point Estimation, Updated

In more Mathematical terms, the general setting for the Point Estimation is the following:

- We have a Dataset: $x : x_1, x_2, ..., x_n$;
- We model the process that generated x by the Parametric Family of Distributions $\{\mathcal{F}_\theta, \theta \in \Theta\}$;
- We assume the Dataset above is one possible realization of the Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta \qquad \text{for some} \quad \theta \in \Theta;$$

Problem: Estimate the unknown Parameter θ.

To estimate θ,

- We find a good **Estimator** $\hat{\theta}$;

General Setting for the Point Estimation, Updated

In more Mathematical terms, the general setting for the Point Estimation is the following:

- We have a Dataset: $x : x_1, x_2, ..., x_n$;
- We model the process that generated x by the Parametric Family of Distributions $\{\mathcal{F}_\theta, \theta \in \Theta\}$;
- We assume the Dataset above is one possible realization of the Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta \qquad \text{for some} \quad \theta \in \Theta;$$

Problem: Estimate the unknown Parameter θ.

To estimate θ,

- We find a good **Estimator** $\hat{\theta}$;
- We plug the Dataset values into the Estimator to get the **Estimate** for θ

Example

Example: Assume we work with the Bernoulli Model: we have a Random Sample

$$X_1, X_2, ..., X_n \sim Bernoulli(p),$$

and we want to estimate the Parameter p.

Question: Which Estimator to use?

Example

Example: Assume we work with the Bernoulli Model: we have a Random Sample

$$X_1, X_2, ..., X_n \sim Bernoulli(p),$$

and we want to estimate the Parameter p.

Question: Which Estimator to use? Say, is

$$\hat{p} = \overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}$$

good enough to estimate the unknown p?

Example

Example: Assume we work with the Bernoulli Model: we have a Random Sample

$$X_1, X_2, ..., X_n \sim Bernoulli(p),$$

and we want to estimate the Parameter p.

Question: Which Estimator to use? Say, is

$$\hat{p} = \overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}$$

good enough to estimate the unknown p?

Or, maybe,

$$\hat{p} = \frac{X_1 + X_n}{2}$$

is better?

Example

Example: Assume we work with the Bernoulli Model: we have a Random Sample

$$X_1, X_2, ..., X_n \sim Bernoulli(p),$$

and we want to estimate the Parameter p.

Question: Which Estimator to use? Say, is

$$\hat{p} = \overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}$$

good enough to estimate the unknown p?

Or, maybe,

$$\hat{p} = \frac{X_1 + X_n}{2}$$

is better?

Or, maybe

$$\hat{p} = \frac{X_{(1)} + X_{(n)}}{2} \qquad or \qquad \hat{p} = Median(X_1, ..., X_n)?$$

Example

Example: Assume we work with the Gaussian Model: we have a Random Sample

$$X_1, X_2, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2),$$

and we want to estimate the Parameters μ and σ^2.

Question: Which Estimator to use?

Example

Example: Assume we work with the Gaussian Model: we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2),$$

and we want to estimate the Parameters μ and σ^2.

Question: Which Estimator to use? Say, is

$$\hat{\mu} = \overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}$$

good enough to estimate the unknown μ?

Example

Example: Assume we work with the Gaussian Model: we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2),$$

and we want to estimate the Parameters μ and σ^2.

Question: Which Estimator to use? Say, is

$$\hat{\mu} = \overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}$$

good enough to estimate the unknown μ?

And what about estimating σ^2? Can you suggest Estimators?

Example

Example: Assume we work with the Gaussian Model: we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2),$$

and we want to estimate the Parameters μ and σ^2.

Question: Which Estimator to use? Say, is

$$\hat{\mu} = \overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}$$

good enough to estimate the unknown μ?

And what about estimating σ^2? Can you suggest Estimators? Say, which one to choose:

$$\widehat{\sigma^2} = \left(\frac{\sum_{k=1}^n |X_k - \overline{X}_n|}{n}\right)^2 \qquad or \qquad \widehat{\sigma^2} = \frac{\sum_{k=1}^n (X_k - \overline{X}_n)^2}{n} \qquad or$$

$$\widehat{\sigma^2} = \frac{\sum_{k=1}^n (X_k - \overline{X}_n)^2}{n-1} \qquad or \qquad \widehat{\sigma^2} = \text{other Estimator?}$$

Example

Example: Assume we work with the Exponential Model: we have a Random Sample

$$X_1, X_2, ..., X_n \sim Exp(\lambda),$$

and we want to estimate the Parameter λ.

Question: Which Estimator to use?

Example

Example: Assume we work with the Exponential Model: we have a Random Sample

$$X_1, X_2, ..., X_n \sim Exp(\lambda),$$

and we want to estimate the Parameter λ.

Question: Which Estimator to use? Say, is

$$\hat{\lambda} = \overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}$$

good enough to estimate the unknown λ?

Notation

Before starting, few Notations.

Notation

Before starting, few Notations. Let us consider the Model

$$\{\mathcal{F}_\theta : \theta \in \Theta\}.$$

Notation

Before starting, few Notations. Let us consider the Model

$$\{\mathcal{F}_\theta : \theta \in \Theta\}.$$

Assume $F_\theta(x) = F(x|\theta)$ is the CDF of the Distribution $\mathcal{F}_\theta$ and $f_\theta(x) = f(x|\theta)$ is the PD(M)F of $\mathcal{F}_\theta$.

Notation

Before starting, few Notations. Let us consider the Model

$$\{\mathcal{F}_\theta : \theta \in \Theta\}.$$

Assume $F_\theta(x) = F(x|\theta)$ is the CDF of the Distribution $\mathcal{F}_\theta$ and $f_\theta(x) = f(x|\theta)$ is the PD(M)F of $\mathcal{F}_\theta$.

If $X \sim \mathcal{F}_\theta$, then we will write

$$\mathbb{E}_\theta(X)$$

the Expected value of X: since $X \sim \mathcal{F}_\theta$, then the Expected value depends on θ.

Notation

Before starting, few Notations. Let us consider the Model

$$\{\mathcal{F}_\theta : \theta \in \Theta\}.$$

Assume $F_\theta(x) = F(x|\theta)$ is the CDF of the Distribution $\mathcal{F}_\theta$ and $f_\theta(x) = f(x|\theta)$ is the PD(M)F of $\mathcal{F}_\theta$.

If $X \sim \mathcal{F}_\theta$, then we will write

$$\mathbb{E}_\theta(X)$$

the Expected value of X: since $X \sim \mathcal{F}_\theta$, then the Expected value depends on θ. Say, if $X \sim \mathcal{F}_\theta$ is continuous, then

$$\mathbb{E}_\theta(X) = \int_{-\infty}^{+\infty} x \cdot f(x|\theta) dx.$$

Notation

Before starting, few Notations. Let us consider the Model

$$\{\mathcal{F}_\theta : \theta \in \Theta\}.$$

Assume $F_\theta(x) = F(x|\theta)$ is the CDF of the Distribution $\mathcal{F}_\theta$ and $f_\theta(x) = f(x|\theta)$ is the PD(M)F of $\mathcal{F}_\theta$.

If $X \sim \mathcal{F}_\theta$, then we will write

$$\mathbb{E}_\theta(X)$$

the Expected value of X: since $X \sim \mathcal{F}_\theta$, then the Expected value depends on θ. Say, if $X \sim \mathcal{F}_\theta$ is continuous, then

$$\mathbb{E}_\theta(X) = \int_{-\infty}^{+\infty} x \cdot f(x|\theta) dx.$$

And we will use $Var_\theta(X)$ for the Variance of X.

Properties of Estimators

Risk, Mean Squared Error of the Estimator

Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta, \qquad \theta \in \Theta,$$

and we use the Estimator $\hat{\theta}$ to estimate θ.

Risk, Mean Squared Error of the Estimator

Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta, \qquad \theta \in \Theta,$$

and we use the Estimator $\hat{\theta}$ to estimate θ.

Definition: The **Mean Squared Error** or the **Quadratic Risk** of the estimator $\hat{\theta}_n$ of θ is

$$MSE(\hat{\theta}_n, \theta) = Risk(\hat{\theta}_n, \theta) = \mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2].$$

Risk, Mean Squared Error of the Estimator

Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta, \qquad \theta \in \Theta,$$

and we use the Estimator $\hat{\theta}$ to estimate θ.

Definition: The **Mean Squared Error** or the **Quadratic Risk** of the estimator $\hat{\theta}_n$ of θ is

$$MSE(\hat{\theta}_n, \theta) = Risk(\hat{\theta}_n, \theta) = \mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2].$$

Note: *MSE* calculates how close are, in the Quadratic Mean sense, possible values of the Estimator $\hat{\theta}$ to the actual (unknown) value of θ.

Risk, Mean Squared Error of the Estimator

Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta, \qquad \theta \in \Theta,$$

and we use the Estimator $\hat{\theta}$ to estimate θ.

Definition: The **Mean Squared Error** or the **Quadratic Risk** of the estimator $\hat{\theta}_n$ of θ is

$$MSE(\hat{\theta}_n, \theta) = Risk(\hat{\theta}_n, \theta) = \mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2].$$

Note: *MSE* calculates how close are, in the Quadratic Mean sense, possible values of the Estimator $\hat{\theta}$ to the actual (unknown) value of θ. The smaller the value of *MSE*, the better, of course.

Risk, Mean Squared Error of the Estimator

Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta, \qquad \theta \in \Theta,$$

and we use the Estimator $\hat{\theta}$ to estimate θ.

Definition: The **Mean Squared Error** or the **Quadratic Risk** of the estimator $\hat{\theta}_n$ of θ is

$$MSE(\hat{\theta}_n, \theta) = Risk(\hat{\theta}_n, \theta) = \mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2].$$

Note: *MSE* calculates how close are, in the Quadratic Mean sense, possible values of the Estimator $\hat{\theta}$ to the actual (unknown) value of θ. The smaller the value of *MSE*, the better, of course.

Definition: We say that the estimator $\hat{\theta}_n^1$ of θ is **preferable** to $\hat{\theta}_n^2$, another estimator of θ, if

$$MSE(\hat{\theta}_n^1, \theta) \leq MSE(\hat{\theta}_n^2, \theta), \qquad \forall \theta \in \Theta,$$

and there exists a θ s.t. $MSE(\hat{\theta}_n^1, \theta) < MSE(\hat{\theta}_n^2, \theta)$.

Example

Example: Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim Pois(\lambda),$$

and we want to Estimate the unknown λ.

Example

Example: Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim Pois(\lambda),$$

and we want to Estimate the unknown λ. We want to use one of the following Estimators:

$$\hat{\lambda}_1 = X_1 \qquad or \qquad \hat{\lambda}_2 = \frac{X_1 + X_n}{2}.$$

Example

Example: Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim Pois(\lambda),$$

and we want to Estimate the unknown λ. We want to use one of the following Estimators:

$$\hat{\lambda}_1 = X_1 \qquad or \qquad \hat{\lambda}_2 = \frac{X_1 + X_n}{2}.$$

Question: Calculate the Risks of $\hat{\lambda}_1$ and $\hat{\lambda}_2$. Compare these two Estimators: which of $\hat{\lambda}_1$ and $\hat{\lambda}_2$ is preferable?

Example

Example: Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim Pois(\lambda),$$

and we want to Estimate the unknown λ. We want to use one of the following Estimators:

$$\hat{\lambda}_1 = X_1 \qquad or \qquad \hat{\lambda}_2 = \frac{X_1 + X_n}{2}.$$

Question: Calculate the Risks of $\hat{\lambda}_1$ and $\hat{\lambda}_2$. Compare these two Estimators: which of $\hat{\lambda}_1$ and $\hat{\lambda}_2$ is preferable?

Solution: OTB: Use the fact that for $X \sim Pois(\lambda)$,

$$\mathbb{E}(X) = \lambda = Var(X).$$

Best MSE Estimators

Now, having the idea of the Error of estimation for an Estimator, we can try to find the best Estimator in the sense of MSE, i.e., we can try to solve

$$\text{Find } \hat{\theta} \text{ that minimizes } MSE(\hat{\theta}, \theta).$$

Best MSE Estimators

Now, having the idea of the Error of estimation for an Estimator, we can try to find the best Estimator in the sense of MSE, i.e., we can try to solve

$$\text{Find } \hat{\theta} \text{ that minimizes } MSE(\hat{\theta}, \theta).$$

Unfortunately, this Minimization problem has no solution in general.

Best MSE Estimators

Now, having the idea of the Error of estimation for an Estimator, we can try to find the best Estimator in the sense of MSE, i.e., we can try to solve

$$\text{Find } \hat{\theta} \text{ that minimizes } MSE(\hat{\theta}, \theta).$$

Unfortunately, this Minimization problem has no solution in general.

So we move forward considering other Properties of Estimators making them useful in the estimation process.

Bias, Biased and Unbiased Estimators

Definition: The **Bias** of Estimator $\hat{\theta}$ of θ is

$$\mathit{Bias}(\hat{\theta}, \theta) = \mathbb{E}_\theta(\hat{\theta} - \theta) = \mathbb{E}_\theta(\hat{\theta}) - \theta.$$

Bias, Biased and Unbiased Estimators

Definition: The **Bias** of Estimator $\hat{\theta}$ of θ is

$$\mathit{Bias}(\hat{\theta}, \theta) = \mathbb{E}_{\theta}(\hat{\theta} - \theta) = \mathbb{E}_{\theta}(\hat{\theta}) - \theta.$$

Definition: We say that our Estimator $\hat{\theta}$ (of the parameter θ) is **Unbiased**, if

$$\mathit{Bias}(\hat{\theta}, \theta) = 0 \qquad \text{for any} \qquad \theta \in \Theta.$$

Bias, Biased and Unbiased Estimators

Definition: The **Bias** of Estimator $\hat{\theta}$ of θ is

$$\mathit{Bias}(\hat{\theta}, \theta) = \mathbb{E}_\theta(\hat{\theta} - \theta) = \mathbb{E}_\theta(\hat{\theta}) - \theta.$$

Definition: We say that our Estimator $\hat{\theta}$ (of the parameter θ) is **Unbiased**, if

$$\mathit{Bias}(\hat{\theta}, \theta) = 0 \qquad \text{for any} \qquad \theta \in \Theta.$$

If $\hat{\theta}$ is not Unbiased, then we say that it is **Biased**.

Bias, Biased and Unbiased Estimators

Definition: The **Bias** of Estimator $\hat{\theta}$ of θ is

$$\mathit{Bias}(\hat{\theta}, \theta) = \mathbb{E}_\theta(\hat{\theta} - \theta) = \mathbb{E}_\theta(\hat{\theta}) - \theta.$$

Definition: We say that our Estimator $\hat{\theta}$ (of the parameter θ) is **Unbiased**, if

$$\mathit{Bias}(\hat{\theta}, \theta) = 0 \qquad \text{for any} \qquad \theta \in \Theta.$$

If $\hat{\theta}$ is not Unbiased, then we say that it is **Biased**.

In other words, the Estimator $\hat{\theta}$ is **Unbiased**, if

$$\mathbb{E}_\theta\left(\hat{\theta}\right) = \theta \qquad \text{for any} \qquad \theta \in \Theta.$$

Bias, Biased and Unbiased Estimators

Definition: The **Bias** of Estimator $\hat{\theta}$ of θ is

$$Bias(\hat{\theta}, \theta) = \mathbb{E}_\theta(\hat{\theta} - \theta) = \mathbb{E}_\theta(\hat{\theta}) - \theta.$$

Definition: We say that our Estimator $\hat{\theta}$ (of the parameter θ) is **Unbiased**, if

$$Bias(\hat{\theta}, \theta) = 0 \qquad \text{for any} \qquad \theta \in \Theta.$$

If $\hat{\theta}$ is not Unbiased, then we say that it is **Biased**.

In other words, the Estimator $\hat{\theta}$ is **Unbiased**, if

$$\mathbb{E}_\theta\left(\hat{\theta}\right) = \theta \qquad \text{for any} \qquad \theta \in \Theta.$$

Note: We require the above hold **for any parameter value** since, if, say, it is correct for *some* values of θ, then it can happen that the true value of our unknown θ is exactly that value, for whihc we do not have the equality.

Bias and Unbiasedness

- The idea of Unbiased Estimator is the following: if we will calculate Estimates many-many times using our Unbiased Estimator, and then average the obtained Estimates, we will obtain the (almost exact, exact when many-many$\to \infty$) value of our Parameter.

Bias and Unbiasedness

- The idea of Unbiased Estimator is the following: if we will calculate Estimates many-many times using our Unbiased Estimator, and then average the obtained Estimates, we will obtain the (almost exact, exact when many-many$\to \infty$) value of our Parameter.
- So we can say about an Unbiased Estimator as: **In average, it is Exact**

Bias and Unbiasedness

- The idea of Unbiased Estimator is the following: if we will calculate Estimates many-many times using our Unbiased Estimator, and then average the obtained Estimates, we will obtain the (almost exact, exact when many-many$\to \infty$) value of our Parameter.
- So we can say about an Unbiased Estimator as: **In average, it is Exact**
- Bias can be interpreted, in some sense, as the *accuracy* of the Estimator.

Example

Example: Assume we have a Sample of size n from some Population, and we want to estimate the Mean of that Population.

Example

Example: Assume we have a Sample of size *n* from some Population, and we want to estimate the Mean of that Population. Let $\mathcal{F}_\mu$ be the Distribution of our Population, and μ be the Mean (Expectation) of that Distribution.

Example

Example: Assume we have a Sample of size *n* from some Population, and we want to estimate the Mean of that Population. Let $\mathcal{F}_\mu$ be the Distribution of our Population, and μ be the Mean (Expectation) of that Distribution. So our unknown Parameter here is μ.

Example

Example: Assume we have a Sample of size n from some Population, and we want to estimate the Mean of that Population. Let $\mathcal{F}_\mu$ be the Distribution of our Population, and μ be the Mean (Expectation) of that Distribution. So our unknown Parameter here is μ.

We use size n Random Sample to Model the situation:

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\mu.$$

Example

Example: Assume we have a Sample of size n from some Population, and we want to estimate the Mean of that Population. Let $\mathcal{F}_\mu$ be the Distribution of our Population, and μ be the Mean (Expectation) of that Distribution. So our unknown Parameter here is μ.

We use size n Random Sample to Model the situation:

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\mu.$$

Now, we choose several Estimators to estimate μ:

$$\hat{\mu}_1 = X_1, \qquad \hat{\mu}_2 = \frac{X_1 + X_3}{2}, \qquad \hat{\mu}_3 = \frac{X_1 + X_4}{10},$$

$$\hat{\mu}_4 = \overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}.$$

Example

Example: Assume we have a Sample of size n from some Population, and we want to estimate the Mean of that Population. Let $\mathcal{F}_\mu$ be the Distribution of our Population, and μ be the Mean (Expectation) of that Distribution. So our unknown Parameter here is μ.

We use size n Random Sample to Model the situation:

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\mu.$$

Now, we choose several Estimators to estimate μ:

$$\hat{\mu}_1 = X_1, \qquad \hat{\mu}_2 = \frac{X_1 + X_3}{2}, \qquad \hat{\mu}_3 = \frac{X_1 + X_4}{10},$$

$$\hat{\mu}_4 = \overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}.$$

Problem: Check if each Estimator is Biased or Unbiased.

Solution: OTB