AUA CS108, Statistics, Fall 2020

Lecture 27

Michael Poghosyan

28 Oct 2020

Contents

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

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

Then the followings are **Esimators**:

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

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**:

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

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

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

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.

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.

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

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

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

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

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

Question: Which Estimator to use? Say, is

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

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:

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

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

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

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

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.