AUA CS108, Statistics, Fall 2020

Lecture 31

Michael Poghosyan

06 nov 2020

Contents

B-V Decomposition, Again

Recall again the B-V D:

$$MSE(\hat{\theta}, \theta) = \left(Bias(\hat{\theta}, \theta)\right)^2 + Var_\theta(\hat{\theta}).$$

B-V Decomposition, Again

Recall again the B-V D:

$$MSE(\hat{\theta}, \theta) = \left(Bias(\hat{\theta}, \theta)\right)^2 + Var_\theta(\hat{\theta}).$$

We have talked before that it would be nice to be able to minimize $MSE(\hat{\theta}, \theta)$, but, unfortunately, this is impossible, in general.

B-V Decomposition, Again

Recall again the B-V D:

$$MSE(\hat{\theta}, \theta) = \Big(Bias(\hat{\theta}, \theta)\Big)^2 + Var_\theta(\hat{\theta}).$$

We have talked before that it would be nice to be able to minimize $MSE(\hat{\theta}, \theta)$, but, unfortunately, this is impossible, in general. So Statisticians consider the following restricted Problem:

Find an Unbiased Estimator with the Minimal Variance.

B-V Decomposition, Again

Recall again the B-V D:

$$MSE(\hat{\theta}, \theta) = \left(Bias(\hat{\theta}, \theta)\right)^2 + Var_\theta(\hat{\theta}).$$

We have talked before that it would be nice to be able to minimize $MSE(\hat{\theta}, \theta)$, but, unfortunately, this is impossible, in general. So Statisticians consider the following restricted Problem:

Find an Unbiased Estimator with the Minimal Variance.

Well, in general, there will be a lot of Unbiased Estimators for the same Parameter.

B-V Decomposition, Again

Recall again the B-V D:

$$MSE(\hat{\theta}, \theta) = \left(Bias(\hat{\theta}, \theta)\right)^2 + Var_\theta(\hat{\theta}).$$

We have talked before that it would be nice to be able to minimize $MSE(\hat{\theta}, \theta)$, but, unfortunately, this is impossible, in general. So Statisticians consider the following restricted Problem:

Find an Unbiased Estimator with the Minimal Variance.

Well, in general, there will be a lot of Unbiased Estimators for the same Parameter.Say, if $\hat{\theta}_0$ and $\hat{\theta}_1$ are Unbiased Estimators of θ, then for any $\alpha \in [0, 1]$, the Estimator

$$\hat{\theta}_\alpha = \alpha \cdot \hat{\theta}_1 + (1 - \alpha) \cdot \hat{\theta}_0$$

will be an Unbiased Estimator too.

MVUE

So the idea is to restrict our attention to only Unbiased Estimators. In that case, since $Bias(\hat{\theta}, \theta) = 0$,

$$MSE(\hat{\theta}, \theta) = Var_{\theta}(\hat{\theta}).$$

MVUE

So the idea is to restrict our attention to only Unbiased Estimators. In that case, since $Bias(\hat{\theta}, \theta) = 0$,

$$MSE(\hat{\theta}, \theta) = Var_\theta(\hat{\theta}).$$

Then we need to find the Estimator with the smallest Variance, i.e., with the highest precision.

MVUE

So the idea is to restrict our attention to only Unbiased Estimators. In that case, since $Bias(\hat{\theta}, \theta) = 0$,

$$MSE(\hat{\theta}, \theta) = Var_{\theta}(\hat{\theta}).$$

Then we need to find the Estimator with the smallest Variance, i.e., with the highest precision.

And we give the following

Definition: Estimator $\hat{\theta}$ is called the **MVUE (Minimum Variance Unbiased Estimator)** for θ, if

MVUE

So the idea is to restrict our attention to only Unbiased Estimators. In that case, since $Bias(\hat{\theta}, \theta) = 0$,

$$MSE(\hat{\theta}, \theta) = Var_{\theta}(\hat{\theta}).$$

Then we need to find the Estimator with the smallest Variance, i.e., with the highest precision.

And we give the following

Definition: Estimator $\hat{\theta}$ is called the **MVUE (Minimum Variance Unbiased Estimator)** for θ, if

- $\hat{\theta}$ is Unbiased Estimator for θ;

MVUE

So the idea is to restrict our attention to only Unbiased Estimators. In that case, since $Bias(\hat{\theta}, \theta) = 0$,

$$MSE(\hat{\theta}, \theta) = Var_\theta(\hat{\theta}).$$

Then we need to find the Estimator with the smallest Variance, i.e., with the highest precision.

And we give the following

Definition: Estimator $\hat{\theta}$ is called the **MVUE (Minimum Variance Unbiased Estimator)** for θ, if

- $\hat{\theta}$ is Unbiased Estimator for θ;
- $\hat{\theta}$ has the smallest Variance among all *Unbiased* Estimators of θ, i.e., for any Unbiased Estimator $\tilde{\theta}$,

$$Var_\theta(\hat{\theta}) \leq Var_\theta(\tilde{\theta}), \qquad \forall \theta \in \Theta.$$

Consistency

Another important and desirable property of an Estimator is the Consistency: this is about its behaviour when we increase the number of Datapoints:

Consistency

Another important and desirable property of an Estimator is the Consistency: this is about its behaviour when we increase the number of Datapoints:

Definition: A point estimator $\hat{\theta}_n$ of the parameter θ is called

- **consistent**, if $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$ for any $\theta \in \Theta$;

Consistency

Another important and desirable property of an Estimator is the Consistency: this is about its behaviour when we increase the number of Datapoints:

Definition: A point estimator $\hat{\theta}_n$ of the parameter θ is called

- **consistent**, if $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$ for any $\theta \in \Theta$;
- **strongly consistent**, if $\hat{\theta}_n \xrightarrow{a.s.} \theta$ for any $\theta \in \Theta$;

Consistency

Another important and desirable property of an Estimator is the Consistency: this is about its behaviour when we increase the number of Datapoints:

Definition: A point estimator $\hat{\theta}_n$ of the parameter θ is called

- **consistent**, if $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$ for any $\theta \in \Theta$;
- **strongly consistent**, if $\hat{\theta}_n \xrightarrow{a.s.} \theta$ for any $\theta \in \Theta$;
- **weakly or Mean Square consistent**, if $\hat{\theta}_n \xrightarrow{q.m.} \theta$ for any $\theta \in \Theta$, i.e., if

$$MSE(\hat{\theta}_n, \theta) = \mathbb{E}_\theta((\hat{\theta}_n - \theta)^2) \to 0 \qquad \forall \theta \in \Theta.$$

Example

Example: Consider a Random Sample

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

where p is our unknown Parameter to be estimated.

Example

Example: Consider a Random Sample

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

where p is our unknown Parameter to be estimated.

We consider the following Estimator for p:

$$\hat{p} = \frac{X_1 + X_2 + ... + X_n}{n+1}.$$

Example

Example: Consider a Random Sample

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

where p is our unknown Parameter to be estimated.

We consider the following Estimator for p:

$$\hat{p} = \frac{X_1 + X_2 + ... + X_n}{n+1}.$$

Then:

- $\hat{p}$ is a Biased Estimator for p;
- $\hat{p}$ is Consistent Estimator for p.

Proof: OTB

Some Properties

Proposition:

- If the estimator is strongly consistent, then it is consistent;

Some Properties

Proposition:

- If the estimator is strongly consistent, then it is consistent;
- If the estimator is weakly consistent, then it is consistent;

Some Properties

Proposition:

- If the estimator is strongly consistent, then it is consistent;
- If the estimator is weakly consistent, then it is consistent;
- If $\hat{\theta}_n$ is an *Unbiased Estimator* for θ and

$$Var(\hat{\theta}_n) \to 0, \qquad n \to \infty,$$

  then $\hat{\theta}_n$ is consistent;

Some Properties

Proposition:

- If the estimator is strongly consistent, then it is consistent;
- If the estimator is weakly consistent, then it is consistent;
- If $\hat{\theta}_n$ is an *Unbiased Estimator* for θ and

  $$Var(\hat{\theta}_n) \to 0, \qquad n \to \infty,$$

  then $\hat{\theta}_n$ is consistent;
- If $\hat{\theta}_n$ is an *Asymptotically Unbiased Estimator* for θ and

  $$Var(\hat{\theta}_n) \to 0, \qquad n \to \infty,$$

  then $\hat{\theta}_n$ is consistent.

Example

Example: Assume $X_1, X_2, ..., X_n, ...$ are IID from a Distribution with the Mean μ, Variance σ^2 and finite 4-th order Moment $\mathbb{E}(X_1^4)$.

Example

Example: Assume $X_1, X_2, ..., X_n, ...$ are IID from a Distribution with the Mean μ, Variance σ^2 and finite 4-th order Moment $\mathbb{E}(X_1^4)$. Let us denote this as

$$X_1, X_2, ..., X_n \sim \mathcal{F}_{\mu,\sigma^2}$$

and let our aim be to estimate σ^2 based on $X_1, X_2, ...$

Example

Example: Assume $X_1, X_2, ..., X_n, ...$ are IID from a Distribution with the Mean μ, Variance σ^2 and finite 4-th order Moment $\mathbb{E}(X_1^4)$. Let us denote this as

$$X_1, X_2, ..., X_n \sim \mathcal{F}_{\mu,\sigma^2}$$

and let our aim be to estimate σ^2 based on $X_1, X_2, ...$

Let's consider the following Estimator for σ^2, the Sample Variance with the denominator n:

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

Example

Example: Assume $X_1, X_2, ..., X_n, ...$ are IID from a Distribution with the Mean μ, Variance σ^2 and finite 4-th order Moment $\mathbb{E}(X_1^4)$. Let us denote this as

$$X_1, X_2, ..., X_n \sim \mathcal{F}_{\mu,\sigma^2}$$

and let our aim be to estimate σ^2 based on $X_1, X_2, ...$.

Let's consider the following Estimator for σ^2, the Sample Variance with the denominator n:

$$\widehat{\sigma^2} = \frac{\sum\limits_{k=1}^{n}(X_k - \overline{X}_n)^2}{n}$$

Then

- $\widehat{\sigma^2}$ is Biased;

Example

Example: Assume $X_1, X_2, ..., X_n, ...$ are IID from a Distribution with the Mean μ, Variance σ^2 and finite 4-th order Moment $\mathbb{E}(X_1^4)$. Let us denote this as

$$X_1, X_2, ..., X_n \sim \mathcal{F}_{\mu,\sigma^2}$$

and let our aim be to estimate σ^2 based on $X_1, X_2, ...$

Let's consider the following Estimator for σ^2, the Sample Variance with the denominator n:

$$\widehat{\sigma^2} = \frac{\displaystyle\sum_{k=1}^{n}(X_k - \overline{X}_n)^2}{n}$$

Then

- $\widehat{\sigma^2}$ is Biased;
- $\widehat{\sigma^2}$ is Consistent.

Proof: OTB. Use the relation $\widehat{\sigma^2} = \frac{\sum_{k=1}^{n}(X_k)^2}{n} - \left(\frac{\sum_{k=1}^{n} X_k}{n}\right)^2$.

Choosing Good Estimators

We have talked about different properties of Estimators:

Choosing Good Estimators

We have talked about different properties of Estimators:

- Unbiasedness

Choosing Good Estimators

We have talked about different properties of Estimators:

- Unbiasedness
- Consistency

Choosing Good Estimators

We have talked about different properties of Estimators:

- Unbiasedness
- Consistency

Unbiasedness is good if we can do resamplings: to estimate our parameter, we can calculate estimates for many samples and average the obtained values - that average will be close to the real value of the parameter;

Choosing Good Estimators

We have talked about different properties of Estimators:

- Unbiasedness
- Consistency

Unbiasedness is good if we can do resamplings: to estimate our parameter, we can calculate estimates for many samples and average the obtained values - that average will be close to the real value of the parameter;

Consistency is good when we have a large sample: we can just calculate the estimate for that sample, and the obtained value will be close to the real value of the parameter with high probability.

Choosing Good Estimators

We have talked about different properties of Estimators:

- Unbiasedness
- Consistency

Unbiasedness is good if we can do resamplings: to estimate our parameter, we can calculate estimates for many samples and average the obtained values - that average will be close to the real value of the parameter;

Consistency is good when we have a large sample: we can just calculate the estimate for that sample, and the obtained value will be close to the real value of the parameter with high probability.

Of course, it is better to have dva v odnom: an Unbiased and Consistent Estimator $\ddot\smile$

Choosing Good Estimators

We have talked about different properties of Estimators:

- Unbiasedness
- Consistency

Unbiasedness is good if we can do resamplings: to estimate our parameter, we can calculate estimates for many samples and average the obtained values - that average will be close to the real value of the parameter;

Consistency is good when we have a large sample: we can just calculate the estimate for that sample, and the obtained value will be close to the real value of the parameter with high probability.

Of course, it is better to have dva v odnom: an Unbiased and Consistent Estimator $\ddot\smile$

And also, the universal measure for goodness is: *an Estimator is good if it has a small MSE*.

Question

Question: Is sampling 200 times with a sample size 10 the same as sampling once with a size 2000?

Question

Question: Is sampling 200 times with a sample size 10 the same as sampling once with a size 2000? And, in general, isn't it the same to sample many times with a small sample size, as to take a large dataset once?

Question

Question: Is sampling 200 times with a sample size 10 the same as sampling once with a size 2000? And, in general, isn't it the same to sample many times with a small sample size, as to take a large dataset once?

Answer: No, in general. This is because, say,

- we can do a lot of resamplings even when our dataset is not big enough, but one large sample will not be available

Question

Question: Is sampling 200 times with a sample size 10 the same as sampling once with a size 2000? And, in general, isn't it the same to sample many times with a small sample size, as to take a large dataset once?

Answer: No, in general. This is because, say,

- we can do a lot of resamplings even when our dataset is not big enough, but one large sample will not be available
- when taking a large sample, we will take each individual just once. But if we are doing resamplings, we can have the same individual in different samples.