AUA CS108, Statistics, Fall 2020

Lecture 30

Michael Poghosyan

04 nov 2020

Contents

Asymptotic Unbiasedness

Some years ago Unbiasedness was a very important, desirable property from any good Estimator.

Asymptotic Unbiasedness

Some years ago Unbiasedness was a very important, desirable property from any good Estimator. Nowadays, it is enough to have *Asymptotic Unbiasedness*:

Asymptotic Unbiasedness

Some years ago Unbiasedness was a very important, desirable property from any good Estimator. Nowadays, it is enough to have *Asymptotic Unbiasedness*:

Definition: Estimator $\hat{\theta}_n$ is called **Asymptotically Unbiased** for θ, if

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

Asymptotic Unbiasedness

Some years ago Unbiasedness was a very important, desirable property from any good Estimator. Nowadays, it is enough to have *Asymptotic Unbiasedness*:

Definition: Estimator $\hat{\theta}_n$ is called **Asymptotically Unbiased** for θ, if

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

Idea: If the Sample size is very large, then the behaviour of our Asymptotic Unbiased Estimator is close to an Unbiased one, $Bias(\hat{\theta}_n, \theta) \approx 0$

Asymptotic Unbiasedness

Some years ago Unbiasedness was a very important, desirable property from any good Estimator. Nowadays, it is enough to have *Asymptotic Unbiasedness*:

Definition: Estimator $\hat{\theta}_n$ is called **Asymptotically Unbiased** for θ, if

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

Idea: If the Sample size is very large, then the behaviour of our Asymptotic Unbiased Estimator is close to an Unbiased one, $Bias(\hat{\theta}_n, \theta) \approx 0$

Example: Say, for the Mean μ of the Population,

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

is a Biased, but Asymptotically Unbiased Estimator. OTB, please!

Bias-Variance Decomposition

This is an important result:

Theorem(Bias-Variance Decomposition of the MSE): If $\hat{\theta}$ is an Estimator for θ, then

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

Bias-Variance Decomposition

This is an important result:

Theorem(Bias-Variance Decomposition of the MSE): If $\hat{\theta}$ is an Estimator for θ, then

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

Proof: OTB

Bias-Variance Decomposition

Again, let's recall our BVD:

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

Bias-Variance Decomposition

Again, let's recall our BVD:

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

We interpret the RHS:

- Bias is the **Accuracy** of our Estimator $\hat{\theta}$

Bias-Variance Decomposition

Again, let's recall our BVD:

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

We interpret the RHS:

- Bias is the **Accuracy** of our Estimator $\hat{\theta}$
- Variance is the **Precision** of our Estimator $\hat{\theta}$

Bias-Variance Decomposition

Again, let's recall our BVD:

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

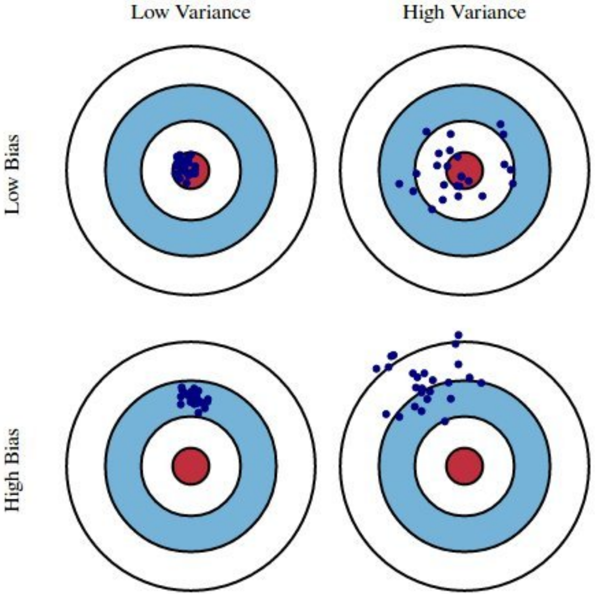
We interpret the RHS:

- Bias is the **Accuracy** of our Estimator $\hat{\theta}$
- Variance is the **Precision** of our Estimator $\hat{\theta}$

Nice Graphical Interpretation: Link, see also the next slide.

Bias-Variance Decomposition/Tradeoff

Bias-Variance Decomposition/Tradeoff



Standard Error and Estimated Standard Error

Definition: The Standard Deviation of the Estimator is called the **Standard Error** of the Estimator $\hat{\theta}$ and is denoted by

$$SE(\hat{\theta}) = SD(\hat{\theta}) = \sqrt{Var_{\theta}(\hat{\theta})}.$$

Standard Error and Estimated Standard Error

Definition: The Standard Deviation of the Estimator is called the **Standard Error** of the Estimator $\hat{\theta}$ and is denoted by

$$SE(\hat{\theta}) = SD(\hat{\theta}) = \sqrt{Var_{\theta}(\hat{\theta})}.$$

Usually, the Standard Error will depend on the unknown value of the Parameter θ. If we use the Estimator $\hat{\theta}$, then the **Estimated Standard Error** of $\hat{\theta}$, $\widehat{SE}(\hat{\theta})$ is the Standard Error, where after calculation we plug $\hat{\theta}$ instead of θ.

Standard Error and Estimated Standard Error

Definition: The Standard Deviation of the Estimator is called the **Standard Error** of the Estimator $\hat{\theta}$ and is denoted by

$$SE(\hat{\theta}) = SD(\hat{\theta}) = \sqrt{Var_{\theta}(\hat{\theta})}.$$

Usually, the Standard Error will depend on the unknown value of the Parameter θ. If we use the Estimator $\hat{\theta}$, then the **Estimated Standard Error** of $\hat{\theta}$, $\widehat{SE}(\hat{\theta})$ is the Standard Error, where after calculation we plug $\hat{\theta}$ instead of θ.

And statisticians, when reporting the Estimate, usually report also the Estimated Standard Error, as a measure how precise is the result. If the Standard Error is small (and we are using a nice Estimator, say, it is Unbiased), then this is a sign that the result is close to real/actual one.

Example

Example: Assume we are facing an election with Parties A and B, and we want to estimate the percentage of voters for A in advance. So we do a poll, asking 10 persons to give their preferences. Let the result be:

$$A, B, B, B, A, B, B, A, B, B.$$

Problem: Estimate the percentage of voters for the Party A, and give the Estimated Standard Error.

Solution: OTB.

Bias-Variance Decomposition, Corollary

Recall from the last lecture the BVD:

Theorem(Bias-Variance Decomposition of the MSE): If $\hat{\theta}$ is an Estimator for θ, then

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

Bias-Variance Decomposition, Corollary

Recall from the last lecture the BVD:

Theorem(Bias-Variance Decomposition of the MSE): If $\hat{\theta}$ is an Estimator for θ, then

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

Corollary: If $\hat{\theta}_1$ and $\hat{\theta}_2$ are both *Unbiased* Estimators for the unknown Parameter θ, then $\hat{\theta}_1$ is preferable to $\hat{\theta}_2$ if and only if

$$\mathit{Var}(\hat{\theta}_1) \le \mathit{Var}(\hat{\theta}_2), \qquad \text{for all } \theta,$$

with a strict inequality for at least one value of θ.

Bias-Variance Decomposition, Corollary

Recall from the last lecture the BVD:

Theorem(Bias-Variance Decomposition of the MSE): If $\hat{\theta}$ is an Estimator for θ, then

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

Corollary: If $\hat{\theta}_1$ and $\hat{\theta}_2$ are both *Unbiased* Estimators for the unknown Parameter θ, then $\hat{\theta}_1$ is preferable to $\hat{\theta}_2$ if and only if

$$\mathit{Var}(\hat{\theta}_1) \le \mathit{Var}(\hat{\theta}_2), \qquad \text{for all } \theta,$$

with a strict inequality for at least one value of θ.

Idea: If we consider Unbiased Estimators only, then we prefer that one, which has the least variability.

Bias-Variance Decomposition, Corollary

Recall from the last lecture the BVD:

Theorem(Bias-Variance Decomposition of the MSE): If $\hat{\theta}$ is an Estimator for θ, then

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

Corollary: If $\hat{\theta}_1$ and $\hat{\theta}_2$ are both *Unbiased* Estimators for the unknown Parameter θ, then $\hat{\theta}_1$ is preferable to $\hat{\theta}_2$ if and only if

$$\mathit{Var}(\hat{\theta}_1) \leq \mathit{Var}(\hat{\theta}_2), \qquad \text{for all } \theta,$$

with a strict inequality for at least one value of θ.

Idea: If we consider Unbiased Estimators only, then we prefer that one, which has the least variability. Unbiased means that the values of our Estimators are centered around the true value of the Parameter.

Bias-Variance Decomposition, Corollary

Recall from the last lecture the BVD:

Theorem(Bias-Variance Decomposition of the MSE): If $\hat{\theta}$ is an Estimator for θ, then

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

Corollary: If $\hat{\theta}_1$ and $\hat{\theta}_2$ are both *Unbiased* Estimators for the unknown Parameter θ, then $\hat{\theta}_1$ is preferable to $\hat{\theta}_2$ if and only if

$$Var(\hat{\theta}_1) \leq Var(\hat{\theta}_2), \qquad \text{for all } \theta,$$

with a strict inequality for at least one value of θ.

Idea: If we consider Unbiased Estimators only, then we prefer that one, which has the least variability. Unbiased means that the values of our Estimators are centered around the true value of the Parameter. Variance of the Estimator is measuring how concentrated are the values of our Estimator around the true value of the Parameter.

Bias-Variance Decomposition, Corollary

Recall from the last lecture the BVD:

Theorem(Bias-Variance Decomposition of the MSE): If $\hat{\theta}$ is an Estimator for θ, then

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

Corollary: If $\hat{\theta}_1$ and $\hat{\theta}_2$ are both *Unbiased* Estimators for the unknown Parameter θ, then $\hat{\theta}_1$ is preferable to $\hat{\theta}_2$ if and only if

$$\mathit{Var}(\hat{\theta}_1) \leq \mathit{Var}(\hat{\theta}_2), \qquad \text{for all } \theta,$$

with a strict inequality for at least one value of θ.

Idea: If we consider Unbiased Estimators only, then we prefer that one, which has the least variability. Unbiased means that the values of our Estimators are centered around the true value of the Parameter. Variance of the Estimator is measuring how concentrated are the values of our Estimator around the true value of the Parameter. If variability, Variance is small, it will give better results.