Statistical Data Analysis

Dr. Jana de Wiljes

23. November 2021

Universität Potsdam

Unbiased estimators

Def:

- An estimator is an arbitrary (Borel-measurable) function

$$\widehat{\theta} : \mathcal{X} \to \Theta, \quad x \mapsto \widehat{\theta}(x) \tag{1}$$

- An estimator $\widehat{\theta}$ is called unbiased, if

$$\mathbb{E}_{\theta}[\widehat{\theta}(X)] = \theta \tag{2}$$

  for all $\theta \in \Theta$.

- The bias of an estimator $\widehat{\theta}$ is

$$\text{Bias}_{\theta}(\widehat{\theta}) = \mathbb{E}_{\theta}[\widehat{\theta}(X)] - \theta \tag{3}$$

$= 0$

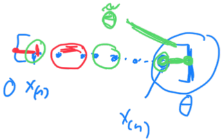
Note: $\text{Bias}_{\theta}(\widehat{\theta})$ is a function in $\widehat{\theta}$

Example: $X_1, \dots, X_n \sim \underline{U[0, \theta]}$ independent and uniformly distributed on $[0, \theta]$, where $\theta > 0$ is the unknown parameter

Further let $X_{(1)} < \dots < X_{(n)}$ be the order statistic of $x_1, \dots, x_n$

1st: Let us consider the Maximum Likelihood estimator

$$\hat{\theta}_1(x_1, \dots, x_n) = x_{(n)} = \max\{x_1, \dots, x_n\}$$



$\Rightarrow \boxed{\hat{\theta}_1 < \theta}$ $\rightsquigarrow$ $\hat{\theta}_1$ has a negative bias

2nd estimator: Goal: trying to improve $\hat{\theta}_1$ by increasing it

$\rightsquigarrow$ would like to increase by $\theta - x_{(n)}$

$\rightsquigarrow$ use that it makes sense to assume that the intervals $(0, x_{(1)})$ and $(x_{(n)}, \theta]$ are approximately the same size

$\rightsquigarrow$ $x_{(1)} \approx \theta - x_{(n)}$ $\rightsquigarrow$ $\hat{\theta}_2(x_1, \dots, x_n) = \boxed{x_{(1)} + x_{(n)}}$

Example

4th estimator: one can also use the Momentestimator. As we know

$\mathbb{E}_\theta[X_i] = \frac{\theta}{2}$ therefore we can use $\overline{X_n}$ to ~~est~~ estimate

$\mathbb{E}_\theta[X_i]$ and set $\hat{\theta}_4(x_1, \dots, x_n) = 2 \cdot \overline{x_n}$

Note: $\hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4$ are unbiased but $\hat{\theta}_1$ is not an unbiased estimator

Questions: which of the unbiased estimates is "the best"?

Example

Mean square error

Def: Let $\Theta = (a, b) \subset \mathbb{R}$ be an interval. The mean square error (MSE) of an estimator $\widehat{\theta} : \mathcal{X} \to \Theta$

$$\mathsf{MSE}_\theta(\widehat{\theta}) = \mathbb{E}_\theta[(\widehat{\theta}(X) - \theta)^2] \tag{4}$$

Mean square error

Lemma: The relationship between the mean square error (MSE) of an estimator $\widehat{\theta} : \mathcal{X} \to \Theta$ and the BIAS is given by

$$\mathrm{MSE}_\theta(\widehat{\theta}) = \mathrm{Var}_\theta \widehat{\theta} + (\underbrace{\mathrm{Bias}_\theta(\widehat{\theta})}_{=0 \text{ for } \hat{\theta} \text{ unbiased}})^2 \tag{5}$$

Proof: $\mathrm{MSE}_\theta(\hat\theta) = \mathbb{E}[(\hat\theta - \theta)^2]$

$= \mathbb{E}[(\hat\theta - \mathbb{E}_\theta[\hat\theta] + \mathbb{E}_\theta[\hat\theta] - \theta)^2]$

$= \mathbb{E}[(\hat\theta - \mathbb{E}_\theta[\hat\theta])^2] + 2\,\mathbb{E}_\theta[(\hat\theta - \mathbb{E}_\theta[\hat\theta])(\mathbb{E}_\theta[\hat\theta] - \theta)]$

$+ \mathbb{E}[(\mathbb{E}_\theta[\hat\theta] - \theta)^2]$

$= \mathrm{Var}_\theta(\hat\theta) + 2(\mathbb{E}_\theta[\hat\theta] - \theta) \cdot \mathbb{E}_\theta[\hat\theta - \mathbb{E}_\theta[\hat\theta]] + \mathrm{Bias}_\theta(\hat\theta)^2$

$\mathbb{E}[\hat\theta - \mathbb{E}_\theta[\hat\theta]] = \mathbb{E}_\theta[\hat\theta] - \mathbb{E}_\theta[\hat\theta] = 0$

$= \mathrm{Var}_\theta(\hat\theta) + \mathrm{Bias}_\theta(\hat\theta)^2$ $\square$

Proof

Proof

Consistently better

Def: Let $\widehat{\Theta}_1$ and $\widehat{\Theta}_2$ be two estimators. The estimator θ_1 is called consistently better than θ_2 if,

$$\boxed{MSE_\theta(\widehat{\theta}_1)} \leq MSE_\theta(\widehat{\theta}_2) \quad \boxed{\forall\theta \in \Theta} \tag{6}$$

Minimum-variance unbiased estimator

Def: An unbiased estimator $\widehat{\theta}$ is called minimum-variance unbiased estimator if all unbiased estimators $\tilde{\theta}$ the following inequality holds

$$\boxed{\mathsf{Var}_\theta \widehat{\theta} \leq \mathsf{Var}_\theta \tilde{\theta}} \qquad \text{MVUE} \tag{7}$$

for all $\theta \in \Theta$.

Minimum-variance unbiased estimator

Lemma: Let $\widehat{\theta}_1, \widehat{\theta}_2 : \mathcal{X} \to \Theta$ are two minimum-variance unbiased estimator the

$$\widehat{\theta}_1 = \widehat{\theta}_2 \quad \text{almost surely under } \mathbb{P} \text{ for all } \theta \in \Theta \tag{8}$$

for all $\theta \in \Theta$.

Bernoulli Experiment MVUE

Lemma: The estimator $\widehat{\theta}(x_1, \ldots, x_n) = \bar{x}_n$ is the minimum-variance unbiased estimator of θ in n Bernoulli experiments.

~~Proof:~~ • Consider $X_1, \ldots, X_n \sim Ber(\theta)$ i.i.d with parameter $\theta \in [0,1]$

• we observe a realisation $(x_1 \ldots x_n)$ and the goal is to estimate θ

Statistical: 1) sample space is $\mathcal{X} = \{0,1\}^n$

2) $A = 2^{\mathcal{X}}$ is the σ-algebra of the measurable events

3) $$\mathbb{P}_\theta[A] = \sum_{(x_1, \ldots, x_n) \in A} \theta^{x_1 + \ldots x_n} (1-\theta)^{n-(x_1+\ldots+x_n)}$$

$A \subset \mathcal{X}$

Proof: Note: 1) $\bar{X}_n$ is unbiased as we already showed that

$$\mathbb{E}[\bar{X}_n] = \mathbb{E}[X_i] = \theta$$

Still need to show that there is no other better estimator

Proof

Let $\varphi : \mathcal{X} \to [0,1]$ be unbiased estimator. Need to show

$$\operatorname{Var}_\theta \varphi \geq \operatorname{Var}_\theta \overline{X}_n \quad \forall\ \theta \in [0,1]$$

Idea: Sufficiency   Intuitively a good estimator would care about how many success but not about when

$\varphi(0,0,0,1,1) \neq \varphi(1,0,1,0,0) \rightsquigarrow \varphi$ is bad estimator

Concept of sufficiency: is about of estimator to project from high dimensional sample space to a some smaller space that still carries all the crucial information

Therefore we define:

$$A_k = \{ x = (x_1, \dots, x_n) \in \{0,1\}^n : x_1 + \dots + x_n = k \}$$

Note:
- $A_0 \cup \dots \cup A_n = \mathcal{X}$
- $|A_k| = \binom{n}{k}$

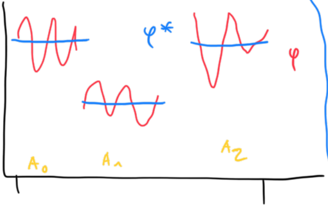
Proof

Further we consider for each estimator $\varphi : \mathcal{X} \to [0,1]$ the so called Rao-Blackwell estimator $\varphi^* : \mathcal{X} \to [0,1]$ with

$$\varphi^*(x) = \frac{1}{\binom{n}{k}} \sum_{y \in A_k} \varphi(y) \quad \text{for } x = (x_1, \dots, x_n) \in A_k$$

Note: • φ^* is constant on each of the set $A_0, \dots, A_n$

• value of φ^* on A_k is the average of φ over A_k



Step 1: Show that φ^* is unbiased (we know φ unbiased)

$$\underline{\mathbb{E}_\theta[\varphi^*]} = \frac{1}{2^n} \sum_{x \in \mathcal{X}} \varphi^*(x)$$

$$= \frac{1}{2^n} \sum_{k=0}^{n} \sum_{x \in A_k} \varphi^*(x) \qquad = \quad x_k \in A_k \qquad \cancel{\binom{n}{k}} \, \varphi^*(x_k)$$

$$= \frac{1}{2^n} \sum_{k=0}^{n} \sum_{y \in A_k} \varphi(y) \qquad \cancel{\frac{1}{\binom{n}{k}}} \quad \sum \varphi($$

$$= \mathbb{E}[\varphi] = \theta$$

$\Rightarrow \varphi^*$ unbiased

□

Proof

Step 2: $\mathrm{Var}_\theta\, \varphi^* \leq \mathrm{Var}_\theta\, \varphi \quad \forall\, \theta \in [0,1]$

use auxillary lemma: $\frac{a_1^2 + \dots + a_N^2}{N} \geq \left(\frac{a_1 + \dots + a_N}{N}\right)^2$

$$① \sum_{y \in A_k} (\varphi(y) - \theta)^2 = \binom{n}{k} \frac{1}{\binom{n}{k}} \sum_{y \in A_k} (\varphi(y) - \theta)^2 \geq \binom{n}{k} \Big(\underbrace{\frac{1}{\binom{n}{k}} \sum_{y \in A_k} \varphi(y)}_{\varphi^*(x)} - \theta \Big)^2$$

$$= \sum_{x \in A_k} (\varphi^*(x) - \theta)^2$$

Since the probability of $y \in A_k$ under $\mathbb{P}_\theta$ is $\theta^k (1-\theta)^{n-k}$

$$\boxed{\mathrm{Var}_\theta\, \varphi =} \mathbb{E}_\theta[(\varphi - \theta)^2] = \sum_{k=0}^{n} \theta^k (1-\theta)^{n-k} \sum_{y \in A_k} (\varphi(y) - \theta)^2$$

$$\overset{①}{\geq} \sum_{k=0}^{n} \theta^k (1-\theta)^{n-k} \sum_{x \in A_k} (\varphi^*(x) - \theta)^2 = \mathbb{E}_\theta[(\varphi^* - \theta)^2] = \mathrm{Var}_\theta(\varphi^*)$$

$\square$

Step 3 to show $\varphi^* = \bar{X}_n$ $\rightarrow$ $\square$

Def: A function $T : \mathcal{X} \to \mathbb{R}^r$ is called a sufficient statistic if the function

$T(X) = x_1 + \dots + x_n = k$

$$\boxed{\theta \mapsto \mathbb{P}_\theta[X = x | \boxed{T(X) = t}]} \tag{9}$$

is constant for all $x \in \mathcal{X}$ and for all $t \in \mathbb{R}^r$,i.e.,

$$\mathbb{P}_{\theta_1}[X = x | T(X) = t] = \mathbb{P}_{\theta_2}[X = x | T(X) = t] \tag{10}$$

for all $t \in \mathbb{R}^r$ and all $\theta_1, \theta_2 \in \Theta$ with $\mathbb{P}_{\theta_1}[T(X) = t] \neq 0$ and $\mathrm{P}_{\theta_2}[T(X) = t] \neq 0$

Example

Example

Sufficient statistic

Lemma: Let $T_{\mathcal{X}} \to \mathbb{R}^r$ be a sufficient statistic and let $g : Im(T) \to \mathbb{R}^k$ an injective function. Then the concatenation

$$g \circ T : \mathcal{X} \to \mathbb{R}^k, \quad x \mapsto g(T(x)) \tag{11}$$

a sufficient statistic.

Proposition: Let $(\mathbb{P}_\theta)_{\theta \in \Theta}$ a family of probability measures on the sample space $(\mathcal{X}, \mathcal{A})$, where $\Theta \subset \mathbb{R}$ is an interval. Furthermore let

- $T : \mathcal{X} \to \mathbb{R}^m$ a sufficient statistic and
- $\hat{\theta} : \mathcal{X} \to \mathcal{R}$ an unbiased estimator of θ with $\mathbb{E}_\theta[\theta^2] \leq \infty$ for all $\theta \in \Theta$.

φ

Define $\tilde{\theta} := \mathbb{E}_\theta[\hat{\theta} | T]$. Then $\tilde{\theta}$ is an unbiased estimator of θ and the following holds

φ^*

$$Var_\theta \tilde{\theta} \leq Var_\theta \hat{\theta} \tag{12}$$

for all $\theta \in \Theta$.