EPFL, Spring 2023
Guillaume Obozinski

EE-209 Eléments de Statistiques pour les Data Sciences

Feuille d'exercices 5

Exercise 5.1 Consistency and bias of estimators

We consider in this exercise a real-valued random variable X such that $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[X^2] < \infty$ and with expectation $\theta := \mathbb{E}[X] > 0$ and variance $v > 0$. We consider $(X_i)_{i=1..n}$ i.i.d. copies of the random variable X.

(a) Relate each estimator of θ presented below with its corresponding properties.

Estimator			Properties
$\widehat{\theta}_1 := X_1$	•	•	Biased and consistent
$\widehat{\theta}_2 := \frac{1}{n}\sum_{i=1}^{n} X_i$	•	•	Biased and non-consistent
$\widehat{\theta}_3 := \frac{1}{n+1}\sum_{i=1}^{n} X_i$	•	•	Unbiased and consistent
$\widehat{\theta}_4 := \frac{X_1}{2}$	•	•	Unbiased and non-consistent

Exercise 5.2 The unbiased variance estimator

In exercice 4.3 from last week, we considered the *moment estimator for the variance estimator,* based on the method of moments. More precisely, given an i.i.d. sample $X_1, \dots, X_n$ from a distribution with expected value $\mu := \mathbb{E}[X_1]$ and variance $\sigma^2 := \mathrm{Var}(X_1) = \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2$, we are interested in estimating σ^2. Clearly, the moment estimator for σ^2 is

$$\widehat{\sigma}^2 = \overline{X^2} - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big)^2.$$

The purpose of the exercise of last week was to show that we always have

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^2.$$

This week we are interested in determining the bias of this estimator.

(a) Compute the expectation of $\widehat{\sigma}^2$.

(b) Deduce from the previous question the value of the bias of $\widehat{\sigma}^2$.

(c) What is the advantage of the estimator

$$S := \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \overline{X})^2?$$

Exercise 5.3 Size of an agricultural field

A farmer has a **perfectly square** agricultural field and wants to estimate its area. Unfortunately, he forgot the size of his field... When he measures a side of his field, he knows (one of his statistician friend has confirmed this) that the measurement error follows a normal distribution with zero expectation and standard deviation $\sigma > 0$. He decides to carry out two measurements which he assumes to be independent and tries to find which estimator he can build from the two measurements.

We denote by X a random variable corresponding to a length obtained after one measurement. We have $X \sim \mathcal{N}(\mu, \sigma^2)$ where μ is the true length of the side and σ is the measurement error. Thus the value θ that he tries to estimate is μ^2.

Since the farmer only wants to make two measurements, he only has a sample of size two for the construction of an estimator. He has three ideas for estimators from these two measurements:

$$T_1 = X_1 \times X_2, \qquad T_2 = \frac{1}{2}(X_1^2 + X_2^2,) \qquad T_3 = \left(\frac{X_1 + X_2}{2}\right)^2.$$

(a) Compute the expectation and the variance of the estimators T_1, T_2 and T_3. Deduce the mean square error (MSE) of these estimators.

To avoid cumbersome computations, you can use directly the fact that for any normal random variable Z with expectation m and variance v, we have

$$\mathbb{E}[Z^4] = m^4 + 6m^2 v + 3v^2.$$

Hint: for T_3, it might be wiser to ask yourself what distribution $\frac{X_1+X_2}{2}$ follows before jumping into the calculations.

(b) Based on your computations, which estimator among T_1, T_2 or T_3 do you suggest to the farmer?

Exercise 5.4 Estimators for the parameter of the Pareto distribution

We say that X is a Pareto random variable with parameter $(1, \alpha)$ and write $X \sim \mathcal{P}(1, \alpha)$, if its probability density function is proportional to

$$g_\alpha : x \mapsto \frac{1}{x^{\alpha+1}} 1_{\{x \geq 1\}}.$$

(a) Determine the constant $c > 0$ such that $c\, g_\alpha$ is the probability density function of a Pareto distribution with parameters $(1, \alpha)$.

(b) Compute the expected value of $X \sim \mathcal{P}(1, \alpha)$ random variable for $\alpha > 1$. What happens if $\alpha \leq 1$?

(c) We now assume that $X_1, X_2, \ldots, X_n$ are i.i.d. $\sim \mathcal{P}(1, \alpha)$. Use the method of moments to obtain an estimator of α, assuming that $\alpha > 1$.

(d) Give an estimator of α using the maximum likelihood principle.

(e) ($*$) Can you prove that $\widehat{\alpha}_{\text{MLE}}$ is consistent for any value of $\alpha > 0$? (Without using the result from the course saying that the MLE is consistent under broad conditions, but by proving it directly?) *Hint:* Compute the expectation of its inverse.