Topic 6: Introduction to sampling. Estimators

A. Class exercises

VARIABLE **The random variable X is distributed according to the density function:**

$$f(x,\theta)=\frac{\theta^2}{2x\theta},\ 0\le x\le\frac{2}{\theta},$$

where $E(X)=\theta$ and $Var(X)=1-\theta^2$. In order to estimate θ, a simple random sample of size *n* has been drawn, and the estimator $T_1=\bar{X}+3$ has been proposed:

a) Obtain and interpret the bias of the above estimator.

b) Calculate the variance of the estimator.

c) The estimator T_2, which is unbiased and has variance $\frac{1-\theta^2}{n}$, is considered as an alternative to T_1. Which of the two estimators is more efficient to estimate θ?

d) If an average of 3.5 has been observed in the sample, provide an estimate for θ.

ESTIMATORS FOR THE MEAN **In order to estimate the mean (μ) of a given population a simple random sample of size 4 is to be drawn, and the following two estimators are considered:**

$$T_1=\bar{X}_4,\qquad T_2=\frac{X_1+2X_3+X_4}{4}$$

Which of the two estimators is more efficient?

UNIVERSITY **A researcher wishes to ascertain the proportion (p) of university students in a certain population. Some studies indicate that this proportion is 25% while others say it is 30%. In a random sample of 20 people drawn from from the population under study, 4 people claimed to be university students. With this information, justify what is the best estimate for p.**

JUMPS **The monthly number of null jumps made by an athlete (X) is a random variable with the following probability function:**

$$P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!},\ x=0,\ 1,\ 2,\ \dots$$

a) Obtain the likelihood function for a simple random sample of size n.

b) Derive the maximum likelihood estimator for λ. If an average of 1.75 null jumps per month has been observed in the sample, what is the maximum likelihood estimate for λ?

c) Obtain a method of moments estimate for λ.

CITY A city council is studying the duration (X, in months) of the bulbs installed in the lamps of the street lighting, which is a random variable with:

$$f(x) = \frac{1}{\theta} e^{-\frac{(x-10)}{\theta}} \qquad x > 10; \theta > 0$$

E (X) = θ + 10, Var (X) = θ^2. To estimate θ it is decided to observe the duration of a random sample of 10 bulbs.

a) Derive the expression of the maximum likelihood estimator for θ.

b) Analyse whether the obtained estimator is unbiased and calculate its mean squared error.

ESTIMATORS

In order to approximate parameter θ $(\theta > 0)$ a s.r.s of size 4 has been selected, and the following two estimators have been proposed:

Estimator	Expectation	Variance
T_A	$E(T_A) = \theta$	$Var(T_A) = \frac{\theta}{4}$
T_B	$E(T_B) = 5\frac{\theta}{4}$	$Var(T_B) = 15\frac{\theta}{16}$

a) Calculate and interpret the bias of the above estimators.

b) Which one is more efficient to estimate θ ?

A. Proposed exercises

SCHOLARSHIP The ministry of education wishes to estimate the expected monthly spending per student, with a view to elaborate a scholarship program. Information on the monthly spending of 4 students selected at random has been obtained through a survey. Justify which of the following estimators

$$T_a = \sum_{i=1}^{4} \frac{X_i}{4}$$

$$T_b = \frac{X_1 + 0.5X_2 - 0.5X_3 + 3X_4}{3}$$

is more efficient to estimate the expected monthly cost:

Results:

$$ECM_{T_a}(\mu) = \frac{\sigma^2}{4} \qquad ECM_{T_b}(\mu) = \left(\frac{\mu}{3}\right)^2 + \frac{10.5\sigma^2}{9}$$

RESTAURANT The waiting times (X, in minutes) until the customers of a restaurant are given a table are uniformly distributed on the interval $[0,\theta]$.

a) Study the bias of the sample mean as an estimator for θ .

b) Derive an estimator for θ by using the method of moments.

c) Compare the efficiency of the above two estimators.

[From *Análisis de datos económicos II. Métodos inferenciales,* Problem 5.5, p. 299]

Results

a)

$$B_{\bar{X}}(\theta) = -\frac{\theta}{2}$$

b)

$$\hat{\theta} = 2\bar{X}$$

c)

$$ECM_{\bar{X}}(\theta) = \frac{\theta^2(3n+1)}{12n}$$

$$ECM_{2\bar{X}}(\theta) = \frac{4\theta^2}{12n}$$

SUPERMARKET A supermarket is conducting a survey in order to investigate the proportion (p) of customers interested in an extension of its business hours. A s.r.s. of n customers will be investigated.

a) Justify how to define the variables in the sample, studying their probability distribution.

b) Derive the likelihood function. What does it mean?

c) Assume that the sample size is 5 and consider statistic T defined as:

$$T = \frac{2X_1 + X_2 - X_3 + X_4 + 2X_5 + 10}{5}$$

Check whether T is an unbiased estimator for p and calculate its mean squared error.

Results

a) $X_i \sim Be(p)$

b)

$$L(x_1, x_2, \ldots x_n, p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

c)

$E(T) \neq p$

$ECM_T(p) = 4 + \dfrac{11pq}{25}$

CALL The duration (X, in minutes) of the telephone calls made in a company is a random variable with:

$$f(x) = \frac{1}{\theta} e^{\frac{-x}{\theta}} \quad for \quad x>0; \quad \theta>0, \; with \quad E(X) = \theta$$

a) Derive the maximum likelihood estimator for θ.

b) Analyse the bias of the above estimator.

c) Derive a method of moments estimator for θ.

Results

a)

$\hat{\theta} = \overline{X}$

b)

$B_{\hat{\theta}} = 0$

c)

MM estimator: $\hat{\theta} = \overline{X}$