Bachelor in Computer Science and Engineering

Statistics Problems

VI Inference

1. The random variable X, tree height is distributed following a model N(μ, σ). Height data are collected in four areas of the forest generating a simple random sample of size 4. As estimators of parameter μ, mean height of the trees in the forest, the following are proposed:

$$\hat{\mu}_1 = \frac{X_1 + 2X_2 + 3X_3}{6}$$

$$\hat{\mu}_2 = \frac{X_3 - 4X_2}{-3}$$

$$\hat{\mu}_3 = \frac{X_1 + X_2 + X_3 + X_4}{4}$$

It is asked:

a) Check if they are unbiased estimators
b) See which one has the smallest variance
c) Assessing the mean squared error, which one is more efficient

2. Let X_1, X_2, …., X_n, be a random sample of independent variables from a population with mean μ and variance σ^2. Considering the following estimators for μ:

$$\widehat{\mu_1} = \frac{X_1 + X_2 + \cdots + X_7}{7}$$

$$\widehat{\mu_2} = \frac{2X_2 - X_3 + X_5}{2}$$

a) Are both estimators unbiased?
b) Which one is the best estimator?
c) Calculate the mean squared error for the best estimator identified in b)

3. Let X and Y be two independent normal random variables, X~N(3α, σ^2); Y~N(5α,σ^2). We take a simple random sample of size n from X and a simple random sample of size n from Y. To estimate parameter α we use the estimator:

$$\hat{\alpha} = a\bar{X} - b\bar{Y}$$

where a and b are two constants.

a) What conditions must the constants a and b meet for the estimator to be unbiased?
b) Now consider that a=2. B=1, α=1 and n= 5, which is the mean squared error for the estimator of α?

4. The measure of a quantitative characteristic of a population with mean μ and standard deviation σ has a density function f(x) = ʎɵ x(1-x) when x ε (0,1) where ɵ and μ are parameters greater than 0. The following estimators have been proposed for parameter ʎ, for samples with size equal or greater than 3:

$$\hat{\lambda}_1 = \frac{2X_1+X_2+X_3}{3}; \qquad \hat{\lambda}_2 = \frac{X_1+2X_2+X_3}{4}$$

a) Calculate the value of ɵ, in both cases, for both estimators to be unbiased
b) Calculate the bias for both estimators for any value of ɵ.
c) Which estimator has smallest variance?

5. Consider the computation time required by an emergent computational algorithm to solve a certain class of complex problems. Let us denote by T the random variable in question and let μ and σ^2 be its expected value and variance. 100 independent scenarios are simulated with the same simulator and the sample average $\overline{t_{100}}$ of the observed computation times is calculated. Please comment if the following statements are true or false and why.

a) By the central limit theorem, we can assume that the distribution of $\overline{t_{100}}$ is approximately N(0,1).
b) The mean squared error for $\overline{t_{100}}$ is equal to σ^2/n
c) Suppose that we have a second sample average $\overline{t_{50}}$ from 50 independent scenarios out of the 100 already mentioned, then

$$\hat{\mu}_1 = (\bar{t}_{100} + \bar{t}_{50})/_2$$

Is an unbiased estimator of μ.

d) Consider the unbiased estimator:

$$\hat{\mu}_2 = \frac{2}{3}\bar{t}_{100} + \frac{1}{3}\bar{t}_{50}$$

Then $\hat{\mu}_1$ is more efficient than $\hat{\mu}_2$ because $\hat{\mu}_2$ has a greater variance.

6. Given a population X with exponential density function $f(x) = \frac{1}{9}e^{\frac{-x}{\theta}}$, x>0, ɵ>0. To estimate parameter ɵ a simple random sample of size n=2. As estimator of ɵ the following statistics are proposed:

$$\hat{\theta}_1 = \frac{2X_1+4X_2}{6}$$

$$\hat{\theta}_2 = \frac{5X_1 + 6X_2}{12}$$

a) Check if both estimators are unbiased
b) Which is the best estimator?