Exercise Set 5

Goals

1) Practice working with estimators
2) Experience how one parameter can be estimated by different estimators, and that they have different efficiency
3) Compute the distribution function of a mean

1 Demonstrate an unbiased estimator for the variance

We have seen in the lecture that the sample variance is a biased estimator for the population variance, and it was stated that this problem can be removed by replacing $n \to n-1$. The aim of this exercise is to prove this statement.

Let X_1, X_2, ..., X_n be random variables (RV), that are independent and identically distributed (i.i.d.). They have an average μ and a variance σ^2. Let $\bar{X} = (X_1 + X_2 + ... + X_n)/n$ be the average of the samples. We define the "sum of Squares" $SS(X) = \sum_{i=1}^{n}(X_i - \bar{X})^2$ as the sum of the squares of the deviations $X_i - \bar{X}$. The statistic $s^2(X) = SS(X)/(n-1)$ is a estimator of the variance of X. Is this estimator biased?

2 Two different estimators of a macromolecule length measurement

A researcher has developed an experimental method to randomly cut a macromolecule and simultaneously measure the length of the molecule she cut off.

Before cutting, all molecules are of equal but unknown length we denote by θ. As the molecules are cut at any point with equal probability, the length distribution can be assumed to follow a uniform law $\mathcal{U}(0, \theta)$, i.e. a continuous distribution whose probability density $f(x)$ is constant and equal to c form 0 to θ and null everywhere else:

$$f(x) = \begin{cases} c & [0, \theta] \\ 0 &]-\infty, 0[\cup]\theta, \infty[\end{cases}$$

a) Compute the cumulative distribution function of this density and express the constant c as a function of θ.

b) The observations are noted X_1, X_2, ..., X_n and follow the law $\mathcal{U}(0, \theta)$. Two estimators for θ are proposed:

$$\begin{aligned} \theta_1 &= 2 \cdot \frac{X_1 + X_2 + ... + X_n}{n} \\ \theta_2 &= \max(X_1, X_2, ..., X_n) \end{aligned}$$

   i. Are the estimators biased? Prove your result.

ii. In the case that one of them is biased, compute its expected mean and suggest a solution to remove the bias.

c) Based on the previous results, compute the mean square error of the two unbiased estimators.

d) Which estimator performs better?

3 The distribution function of a mean

a) Let there be two independent normal/Gaussian distributions, both with a mean of 0. One has a standard deviation of 4 while the other has a standard deviation of 3. Derive the probability distribution for the mean of these two distributions (assuming that each was sampled once and they are given equal weight).

b) What is the mean and standard deviation of the resulting distribution? How does this compare to the result you expect from the rules for adding variances?