Mathematical Statistics 2021/2022, Problem sets 5 and 6

Estimator properties

1. The size of organisms from a given population may be described by a distribution with density $f_\beta(x) = \frac{x}{\beta^2} e^{-\frac{x}{\beta}}$ for $x > 0$ (and 0 otherwise). A sample of n organisms is drawn from the population. For the m.l.e. estimator of β (see Problem 3/Set 4), calculate the MSE, the bias and the estimator variance.

   Hint: The expected value for this distribution is 2β, and the variance is $2\beta^2$.

2. A population is characterized by a density function of $f_\lambda(x) = \lambda e^{-\lambda x}$ for $x > 0$. Check that $\hat{v} = \frac{1}{2n} \sum_{i=1}^{n} X_i^2$ is an unbiased estimator of the population variance. What is the variance of this estimator? Calculate the MSE of the estimator.

   Hint: The first four moments of the exponential distribution are equal to $\frac{1}{\lambda}, \frac{2}{\lambda^2}, \frac{6}{\lambda^3}, \frac{24}{\lambda^4}$.

3. 30 measurements of an unknown value μ were taken. Method A was used for the first 20 measurements, and the results – $X_1, X_2, \ldots, X_{20}$ – are random variables from a normal distribution $N(\mu, 3^2)$. A different method B was used for the following 10 measurements, and the results – $X_{21}, \ldots, X_{30}$ – are random variables from a normal distribution $N(\mu, 2^2)$. All measurements are independent. Let

$$\bar{X}_A = \frac{1}{20} \sum_{i=1}^{20} X_i, \quad \bar{X}_B = \frac{1}{10} \sum_{i=21}^{30} X_i.$$

   Find a and b such that the estimator $\hat{\mu} = a\bar{X}_A + b\bar{X}_B$ is unbiased with minimum variance.

4. Assume that the amount (in USD) a random consumer is willing to spend yearly on water consumption follows a uniform distribution on the interval $[0, \theta]$, where $\theta > 0$ is an upper bound unknown to the researcher. The researcher surveys n independent individuals and records their yearly expenses $X_1, X_2, \ldots, X_n$.

   (a) Find the m.l.e. of θ and calculate its bias and variance.

   (b) Construct an unbiased estimator on the basis of the m.l.e. of θ. Calculate the variance of this estimator.

   (c) For which value of a will $\hat{\theta}_a = \frac{a}{n} \sum_{i=1}^{n} X_i$ be an unbiased estimator of the parameter θ? Determine the variance of this estimator.

   (d) Compare the three estimators above on the base of the MSE.

   (e) Construct the method of quantiles estimator for θ, based on the median. Verify whether this estimator is unbiased (you may assume that n is odd, i.e. that $n = 2l + 1$).

   Hints: The m.l.e. of θ for a sample from a uniform distribution on $[0, \theta]$ is $\hat{\theta} = \max\{X_1, \ldots X_n\}$. The distribution of a k-th order statistic of a sample from a distribution with density $f(x)$ and cumulative distribution $F(x)$ has a density function

$$f_{X_{k:n}}(x) = n\binom{n-1}{k-1} f(x) F(x)^{k-1} (1 - F(x))^{n-k}$$

   .

5. Let $X_1, X_2, \ldots, X_n$ denote the prices (in EUR) of a given article in different shops. We assume these observations are independent, from a normal distribution with unknown mean μ and variance σ^2. Previous research suggests that the average price level is around 50, so the researcher uses the following "conservative" estimator of the parameter μ:

$$\hat{\mu} = \frac{50 + \bar{X}}{2}.$$

   (a) Verify whether this estimator is unbiased. Calculate the MSE.

   (b) Verify whether this estimator is consistent.

6. Let $\theta \in (0, 1)$ denote the probability that a random client entering a shop will buy a box of chocolates. Let $X_1, X_2, \ldots, X_{2n}$ denote the outcomes (1 – purchase, 0 – otherwise) for $2n$ independent consumers ($2n > 20$).

   (a) Denote by $\hat{\theta}_{MLE}$ the m.l.e. of θ on the base of the sample $X_1, X_2, \ldots, X_{2n}$, by $\hat{\theta}_2$ the m.l.e. on the base of odd observations only, and by $\hat{\theta}_3$ the m.l.e. on the base of the first 20 observations (i.e. $X_1, \ldots, X_{20}$).

   (b) Check that these estimators are unbiased.

   (c) Verify whether these estimators are efficient.

   (d) Verify whether these estimators are consistent.