ASDS - Applied Statistics with R

Fall 2018, YSU

Homework No. 07

Due time/date: 21:20, 4 December, 2018

Note: Please use R only in the case the statement of the problem contains (R) at the beginning. Otherwise, show your calculations on the paper. Supplementary Problems will not be graded, but you are very advised to solve them and to discuss later with TA or Instructor.

Problem 1. Assume we have a random sample

$$X_1, X_2, ..., X_n \sim Bernoulli(p)$$

where $p \in [0, 1]$ is our unknown parameter. Which of the followings are estimators for p?

a. $\hat{p}_n = \dfrac{X_1 + X_2 + ... + X_n}{2}$;

b. $\hat{p}_n = X_1 \cdot X_n - 3X_2 \cdot X_{n-1}$;

c. $\hat{p}_n = 250 \cdot X_3$;

d. $\hat{p}_n = \dfrac{X_1 + X_n}{2} - p$;

e. $\hat{p}_n = \sqrt[n]{X_1 \cdot X_2 \cdot ... \cdot X_n}$.

Problem 2. Assume we know that our random sample comes from one of the distributions of the family $\{\mathcal{N}(2, \sigma^2) : \sigma^2 \in (0, +\infty)\}$:

$$X_1, X_2, ..., X_n \sim \mathcal{N}(2, \sigma^2),$$

and we want to estimate σ^2.

a. We take

$$\widehat{\sigma^2} = \frac{1}{n-1} \cdot \sum_{k=1}^{n} (X_k - \overline{X})^2,$$

where

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n}.$$

Calculate the bias of this estimator for σ^2. Is $\widehat{\sigma^2}$ unbiased?

b. Now we take

$$\widehat{\sigma^2} = \frac{1}{n-1} \cdot \sum_{k=1}^{n} (X_k - 2)^2,$$

Note the difference: instead of the sample mean $\overline{X}$ here we have the theoretical mean $\mu = 2$. Calculate the bias of this estimator for σ^2. Is $\widehat{\sigma^2}$ unbiased? What if we will take $\frac{1}{n}$ instead of $\frac{1}{n-1}$?

Problem 3. Assume $X_1, ..., X_n$ are IID from some distribution with an unknown mean μ and known variance σ^2. We want to estimate the mean μ. We consider two estimators for μ:

$$\widehat{\mu}_n = \frac{X_1 + 3X_n}{4} \qquad \text{and} \qquad \tilde{\mu}_n = \frac{X_1 + X_2 + ... + X_{n-1}}{n}.$$

Calculate the biases for each estimators. Which of the estimators is unbiased?

Problem 4. (**R**) Assume we want to test if the coin is fair. To this end we want to use the statistics

$$\hat{p} = \frac{X_1 + X_2 + ... + X_{10}}{10}.$$

The attached Excel file *Hw07p04.xlsx* contains results for 150 experiments of tossing a coin 10 times (150 realizations/experiments for X_1,...,X_{10}, 0 means Tails, 1 means Heads, the first column is the number of the experiment). Construct the distribution (histogram and KDE) of $\hat{p}$ using **R**, calculate the mean of all estimates $\hat{p}$. What do you think, was the coin fair? Explain your reasoning.

Problem 5. Assume we have a random sample $X_1, X_2, ..., X_n$ from the Poisson distribution:

$$X_1, X_2, ..., X_n \sim Poisson(\lambda),$$

where $\lambda > 0$ is our unknown parameter. We consider the following estimator:

$$\widehat{\lambda}_n = \frac{X_1 + X_2 + ... + X_n}{n+1}.$$

a. Calculate the bias and check if $\widehat{\lambda}_n$ is an unbiased estimator for λ.

b. Check if $\widehat{\lambda}_n$ is an asymptotically unbiased estimator.

c. Assume we have the following observation after making the experiment:

$$4,\ 2,\ 5,\ 4,\ 2,\ 4,\ 11,\ 1,\ 5,\ 0$$

Calculate the estimate for λ using the estimator $\widehat{\lambda}_n$.

d. (**R**) We have 20 observations from our random sample, please find them in the attached Excel file *Hw07p05.xlsx* (each row is a realization of $X_1, ..., X_{10}$). Construct the histogram and KDE for the estimator $\widehat{\lambda}_{10}$.

Problem 6. Assume $X_1, X_2, ..., X_n$ are IID from some distribution with CDF $F(x)$.

a. Find the CDF for order statistics $X_{(1)}$ and $X_{(n)}$. Here, as for datasets,

$$X_{(1)} = \min\{X_1, ..., X_n\} \qquad \text{and} \qquad X_{(n)} = \max\{X_1, ..., X_n\}$$

b. (**R**) Here we want to find the distributions of $X_{(1)}$ and $X_{(n)}$ by simulations. Assuming $X_k \sim Unif[-1, 1]$, generate observations $x_1, ..., x_n$ many times and calculate the corresponding values of $X_{(1)}$ and $X_{(n)}$. Then draw the density histogram, KDE, and the theoretical PDF obtained from the above calculations on the same plot, both for $X_{(1)}$ and $X_{(n)}$.

Problem 7. Assume we are estimating some parameter using two estimators: $\hat{\theta}_n$ and $\tilde{\theta}_n$. Here n is the size of the sample. Assume the bias of $\hat{\theta}_n$ is $\frac{2}{n}$ and the variance is $\frac{4}{n}$. The statistics $\tilde{\theta}_n$ is unbiased, and the variance is $\frac{25}{n^2}$. Which of these estimators is preferable?

Problem 8. Assume we have an unbiased estimator $\hat{\theta}_n$ (defined for any n) for the parameter θ. Prove that if

$$Var(\hat{\theta}_n) \to 0,$$

then $\hat{\theta}_n$ is consistent.

Problem 9. Let $X_1, X_2, ..., X_n$ be IID from a family of distributions with unknown mean μ and known variance σ^2. We want to estimate μ, based on the observation.

a. To estimate μ, we consider two estimators:

$$\hat{\mu}_n^1 = \frac{3}{4}X_1 + \frac{1}{4}X_n \qquad \text{and} \qquad \hat{\mu}_n^2 = \frac{1}{7}X_1 + \frac{6}{7}X_n.$$

Show that these two estimators are unbiased, and choose the preferable one from these two.

b. Among all estimators of the form

$$\hat{\mu}_a = aX_1 + (1-a)X_n, \qquad a \in [0, 1],$$

find the one with the minimal risk.

Problem 10. We are using a random sample form the distribution with unknown mean μ to estimate that μ. We will use the estimator

$$\hat{\mu}_n = \frac{X_2 + X_3 + ... + X_{n-1}}{n}$$

to estimate μ. Show that $\hat{\mu}_n$ is biased, but asymptotically unbiased.

Problem 11. Consider the Poisson Distribution: $Poisson(\lambda)$, where $\lambda > 0$ is our unknown parameter.

a. Calculate the Fisher Information $I(\lambda)$;

b. Assume we have a random sample $X_1, X_2, ..., X_n$ from that distribution. Show that the estimator $\hat{\lambda} = \overline{X}$ is efficient.