Probability and Statistics
Y-DATA School of Data Science
P&P 4
Due: 29.11.2022

PROBLEM 1. Let $X_1,...,X_n \overset{\text{i.i.d.}}{\sim} Geo(\theta)$, $\theta \in [0,1]$. Find the MLE for θ.

PROBLEM 2. Let $X_1,...,X_n$ be an i.i.d. sample from the density

$$f_\theta(x) = \frac{1}{x}e^{-\pi(log(x)-\theta)^2}$$

Compute the MLE for θ.

PROBLEM 3. Let $X_1,...,X_n \overset{\text{i.i.d.}}{\sim} N(\mu,\sigma^2)$, where μ and σ^2 are the unknown parameters. Find the MLE of μ and σ^2.

PROBLEM 4. Let $X_1,...,X_n \overset{\text{i.i.d.}}{\sim} U(\theta+2,\theta+10)$ (continuous).
(1) Find $\hat{\theta}_{MOM}$ (method of moments estimator for θ).
(2) Evaluate $\hat{\theta}_{MOM}$ for the sample

$$12.3, 17.5, 15.1, 14.7$$

PROBLEM 5. It is assumed that the daily amount of rain (in mm) that falls in London during January is distributed $N(\mu,25)$. We are interested in estimating $P(X>75)$. Two approaches were suggested:
A Estimate μ using the method of moments, and then estimate the probability using $\hat{\mu}_{MOM}$ instead of μ in the normal distribution.
B Don't assume normality. Estimate the probability by calculating the proportion of observations that are greater than 75.

In a random sample of 10 observations, the following results were received:

$$68.49, 63.61, 71.22, 76.38, 75.99, 78.66, 59.08, 68.82, 75.47, 64.56$$

(1) Estimate the required probability using both methods and compare the results.
(2) Estimate the probability $P(X>72)$ using both methods and compare the results.

PROBLEM 6. Let $X_1,...X_n \overset{\text{i.i.d.}}{\sim} Pois(\lambda)$.
(1) Compute the MSE of the MLE for λ.
(2) A researcher believes that λ is approximately 3, so he suggests to use the estimator which is the average between the MLE and 3: $T = \frac{\bar{X}_n+3}{2}$. Compute the MSE of T.
(3) Compare the bias and the variance of the estimators as functions of λ.
(4) Compare the MSE of the estimators as a function of λ and find for which values of λ each estimator is better than the other. Note that the range of λ might depend on n.

PROBLEM 7. The weight of students in some university is normally distributed. A sample of 12 students is drawn with the following results (in Kg):

$$53.8, 67.34, 51.7, 52, 58.9, 74, 45.3, 53, 62.5, 48.87, 49, 55.6$$

(1) Assuming that the variance is known and equals 1.5 Kg, calculate the confidence interval for the expected value of the weight with confidence level 95%.

(2) Repeat part 1, this time for a confidence level of 90%. What can you say about the difference between the results?
(3) Assuming that the variance is known and equals 2 Kg, calculate the confidence interval for the expected value of the weight with confidence level 95%. What can you conclude from the result?
(4) Repeat part 1, assuming that the variance is unknown.

PROBLEM 8. Let $X \sim N(\mu, \sigma^2)$ (both parameters are unknown). In a random sample of 10 observations we received that

$$\sum_{i=1}^{n} x_i = 15, \sum_{i=1}^{n} x_i^2 = 27$$

and the CI for μ is $[1.09, 1.91]$. What is the confidence level of this confidence interval?

PROBLEM 9. In a random sample of 100 students, it was found that 30 like Bamba.

(1) Compute an asymptotic confidence interval for the proportion of Bamba lovers among the students.
(2) Find the minimal sample size n for which the length of the CI will be at most 0.02.