Homework MAA204

January 8, 2023

Exercice 1

We recall that an exponential random variable X with parameter $\lambda = \frac{1}{m} > 0$ has a density function given by:

$$f_X(x) = \frac{1}{m} e^{-\frac{x}{m}},\ \forall x \geq 0$$

Question 1. Compute the mean and the variance of an exponential random variable X with general parameter $\lambda = \frac{1}{m} > 0$.

The dataset contained in the file `exponential.csv` is represented by a sample of size $N = 500$ of independent observations $(x_i)_{1 \leq i \leq N}$ drawn according to exponential distribution of parameter $\frac{1}{m} \in]0, +\infty[$. As we ignore the value of the parameter m, our goal is to estimate it. For that purpose, we shall study 2 estimators

1. Estimation with the Maximum Likelihood (ML).
2. the estimator $Y_n = n \min(X_1, X_2, \cdots, X_n)$.

Question 2. Maximum Likelihood estimation: compute the likelihood function that we denote $L(\lambda, x_1, x_2, \cdots, x_N)$ and deduce the maximum likelihood parameter $\hat{\lambda} = \frac{1}{\hat{m}_{\text{ML}}}$. What do you notice ?

Question 3. Is the estimator $\hat{m}_{\text{ML}}$ biased ? compute its quadratic risk.

Question 4. On R, read the file `exponential.csv`. Then compare through a plot the distribution of the dataset observations and the distribution of the exponential random variable with the estimated parameter $\hat{m}_{\text{ML}}$.

Question 5. We consider the second estimator Y_n. We set $M_n = \min(X_1, X_2, \cdots, X_n)$. What is the law of M_n ?

Question 6. Is the estiamtor Y_n a biased estimator ? compute its quadratic risk.

Question 7. Given these informations, what do you choose between these two estimators ? justify.

In what follows, we work with the Maximum likelihood estimator $\hat{m}_{\text{ML}}$ and we want to analyze how close is the value of $\hat{m}_{\text{ML}}$ to the real value of the parameter m in terms of probability.

We consider the empirical variance $\bar{V}_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the empirical mean. For the remaining part of this exercice, we recall some properties that will be useful:

P.1 From Slutsky's theorem: Let $n \in \mathbb{N}$ and X_n, Y_n be random variables. if X_n converges in law to a random variable X $\left(\text{we use the notation } X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X\right)$ and Y_n converges in probability to a constant $c \in \mathbb{R}$ $\left(\text{we use the notation } Y_n \xrightarrow[n\to+\infty]{\mathbb{P}} c\right)$ then $X_nY_n \xrightarrow[n\to+\infty]{\mathcal{L}} cX$ and $\frac{X_n}{Y_n} \xrightarrow[n\to+\infty]{\mathcal{L}} \frac{X}{c}$ if $c \neq 0$.

P.2 Property from continuous mapping theorem: Let X_n, X be random variables defined on a metric space S and assume that $f : S \to S'$ (S' another metric space) such that g is continuous then

- if $X_n \xrightarrow[n \to +\infty]{\mathcal{L}} X$ then $g(X_n) \xrightarrow[n \to +\infty]{\mathcal{L}} g(X)$.
- if $X_n \xrightarrow[n \to +\infty]{\mathbb{P}} X$ then $g(X_n) \xrightarrow[n \to +\infty]{\mathbb{P}} g(X)$.

Question 8. Show that $\bar{V}_n$ converges in probability to the variance of X_1, $\mathrm{Var}(X_1)$.

Question 9. What is the asymptotic distribution of the random variable $Z_n = \frac{\sqrt{n}(\hat{m}_{\mathrm{ML}} - m)}{\sqrt{\bar{V}_n}}$ as $n \to +\infty$?

Question 10. For large $n = 10^5$, simulate the random variable Z_n and compare it to its asymptotic distribution (you can choose any value of m)

Question 11. Let $\alpha \in [0, 1]$. By considering a large n, give the interval I such that:

$$P(m \in I) = 1 - \alpha$$

Exercice 2

In this exercice, we will prove the Box Muller method for generating Gaussian random variables. Then, we will investigate the method in practice.

Question 12. Let U follow a uniform distribution $\mathcal{U}([0, 1])$. Let $\phi(x) = \sqrt{-2\log(x)}$. Find the cumulative distribution function of the random variable $R = \phi(U)$. Deduce the density of R

Question 13. Now, let V be a second uniformly distributed random variable between 0 and 1, independent from U. Give the density of the random variable $\Theta = 2\pi V$. Deduce the joint density of R, Θ (remember that U and V are independent).

We now have two independent random variables of known joint distribution. We may consider those two variables as the polar coordinates of a point in a plane. The following fact allows to compute the cartesian coordinates of the same point.

Question 14. Let $(a, b) = g(R, \theta) = (R\cos(\theta), R\sin(\theta))$. Find a way to compute R given only a and b.

Let R, θ be two random variables, and let $X, y = g(R, \theta) = (R\cos(\theta), R\sin(\theta)$. Then, $f_{X,Y}(a, b) = \frac{f_{R,\theta}(g^{-1}(a,b))}{r}$.

Question 15. In the previous question, we only computed the first term of $g^{-1}(a, b)$. Why is the second term unnecessary? Find the analytical expression of $f_{X,Y}$.

Question 16. Are X and Y independent? What are their respective law?

Question 17. We recall that the Box-Muller method consists of generating normal distributions by observing samples of U and V two uniform random variables between 0 and 1, and applying either the transformation $X = \sqrt{-2\log(U)}\cos 2\pi V$ or $Y = \sqrt{-2\log(U)}\sin 2\pi V$. Conclude on the validity of the method.

Question 18. We will now empirically test the Bow-Muller method. Generate 3000 samples of U and V i.i.d. following a uniform law in $[0, 1]$, then X and Y according to the Box-Muller method. Plot the histograms of both functions and compute the covariance between X and Y. Generate 3000 samples using the rnorm function of R. Comment each results.

Question 19. A common way to compare two distributions is through a quantile-quantile diagram. Generate $N1, N2$ two vectors of 3000 sample of a normal distribution using the funntion rnorm, and $E1, E2$ a vector of 3000 samples of an exponential distribution using the function rexp. Plot the quantile-quantile diagrams of $N1$ and $N2$, then of $E1$ and $E2$. What do you observe? Now plot the diagram of $N1$ and $E1$. How do you interpret the changes? (you may type help(qqplot) in the console to get information on the syntax and the procedure).

Question 20. Draw the qqplot of the samples generated through the Box Muller method with the exponentially distributed samples first, then the normally distributed ones. Conclude.