FINAL EXAM

An electronic calculator, the distributions tables as well as a sheet of personal notes are allowed for the exam. Any kind of dictionary is permitted

Exercice 1 (6 points)

A company makes wires whose diameters are normally distributed (in mm). Two machines produce the wires and we consider $X \sim \mathcal{N}(\mu_1, \sigma_1)$ the random variable which gives the diameter of the wires from machine A and $Y \sim \mathcal{N}(\mu_2, \sigma_2)$ the random variable corresponding to the diameters of the wires produced by machine B. We collect the following observations :

x_i	10.2	9.9	10.1	9.5	10
y_i	10	9.7	10.1	9.6	9.8

1. Give a point estimate of the mean and standard deviation of X and Y.

2. Construct a confidence interval for μ_1 at the level of confidence of 95%.

3. Test at a level of significance of 5% the equality of variances.

4. Assuming the variances are equal, give a point estimate of the common value.

5. Test if the mean of X and Y are equal (level of significance 5%).

6. We propose to test $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 = \mu_2 + 0.01$. Give the β-risk for this alternative hypothesis in terms of a probability involving the t-distribution (you are not asked to actually perform the calculation).

Exercice 2 The Pareto distribution (14 points)

The Pareto distribution of parameter $k \in \mathbb{R}_+^*$ is defined over $[1; +\infty)$ (i.e. the density is zero for $x \in (-\infty; 1)$) by the following density function :

$$f(x) = \frac{k}{x^{k+1}} \text{ for any } x \in [1, +\infty)$$

The corresponding cumulative distribution function is $F(x) = 1 - \frac{1}{x^k}, \forall x \in [1; +\infty)$.

In this problem X follows a Pareto distribution of parameter k. It will be denoted by $X \sim Par(k)$ and $(X_1, \ldots, X_n)$ will be a random sample of size n where $X_i \sim X$ and the $X_i's$ are independent.

1. The distribution $Par(k)$ with $k > 1$:

a. Show that $E(X) = \frac{k}{k-1}$. What would you say about the expectation of X if $k = 1$?

b. Show that $k = \frac{E(X)}{E(X) - 1}$ and propose an estimator of k using the statistics $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

2. The Pareto distribution can be used as a model for incomes distribution. For instance the random variable $X \sim Par(k)$ represents the monthly amount of money earned by a random employee of a given firm. The lower possible salary would be $1K€$ (kilo-euro) per month and therefore X takes value in $[1, +\infty)$.

a. An observation of the random sample $(X_1, \ldots, X_n)$ gives the following data : $\overline{x} = 2K€$. Give a point estimate of k.

b. For $k = 2$ what is the proportion of people who earn more than $3K€$?

c. Find the value α such that $P(X \geqslant \alpha) = 0.2$.

d. Prove that $\int_{\alpha}^{+\infty} xf(x)dx \approx E(X) \times 0.45$. Why can we say that in this model the best paid 20% of the employees share 45% of the total incomes ?

3. An estimator of $\frac{1}{k}$: we assume $X \sim Par(k)$ and let $Y = \ln(X)$,

a. Show that for all $t \in \mathbb{R}_+$, $F_Y(t) = F_X(e^t)$.

b. Deduce that Y follows an exponential distribution of parameter k, $Y \sim \mathcal{E}(k)$.

c. We consider $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, with $Y_i = \ln(X_i) \sim \mathcal{E}(k)$. What is the expectation of $\overline{Y}$? What is its variance ?

d. Deduce that $\overline{Y}$ is a convergent estimator of $\frac{1}{k}$

e. Use the central limit theorem to prove that $\sqrt{n}(k\overline{Y} - 1) \approx \mathcal{N}(0, 1)$.

4. Confidence interval and test of parameter : we assume n is a large number.

a. Use the confidence variable $Z = \sqrt{n}(k\overline{Y} - 1)$ and build a confidence interval I_k for the parameter k with level of confidence 95%.

b. Propose a procedure to test $H_0 : k = 2$ versus $k \neq 2$ with a level of significance of 5% (find the domain of acceptance and explain how you would perform the test from an observation $(x_1, \ldots, x_n)$).

5. An other estimator of k :

a. Let $(x_1, \ldots, x_n)$ be an observation of the random sample $(X_1, \ldots, X_n)$. We recall the maximum of likelihood function is defined by

$$L(k) = f(x_1) \ldots f(x_n)$$

Compute the maximum of $\ln(L(k))$.

b. Conclude that the estimator of k, by the maximum of likelihood method, is

$$\hat{k} = \frac{n}{\ln(X_1) + \cdots + \ln(X_n)} = \frac{1}{\overline{Y}}$$

c. **[extra points]** The density function of $T = \ln(X_1) + \cdots + \ln(X_n)$ is $g(x) = \frac{k^n x^{n-1} e^{-kx}}{(n-1)!}$ if $x \geqslant 0$ and $g(x) = 0$ if $x < 0$. Show that $E(\frac{1}{T}) = \frac{nk}{n-1}$ and deduce that $\frac{n-1}{n}\hat{k}$ is an unbiased estimator of k.