ESILV ENGINEERING SCHOOL DE VINCI PARIS	**Advanced Statistics : Estimation, confidence interval and test**	**A4 ESILV** **Tutorial 1**

Exercise 1. EMPIRICAL MEAN AND VARIANCE

We consider a sample $X_1, \dots, X_n$ of independent and identically distributed random variables with expectation m and variance σ^2. We call empirical mean and empirical variance

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

1. Are these estimators unbiased ?
2. Propose an unbiased estimator S_n^{*2} of the variance.
3. **In this question only**, we assume that $X_1, \dots, X_n$ follow a Gaussian distribution $\mathcal{N}(0, \sigma^2)$. Calculate the mean squared error of S_n^2 and S_n^{*2}. Which estimator is better ?
4. Are these estimators consistent ?
5. Show that the three estimators are asymptotically normal estimators.
6. Assume that σ^2 is **an unknown parameter**. Construct an asymptotic confidence interval of level $1-\alpha$ for m depending on S_n^2.

Exercise 2. POISSON MODEL

Let $X_1, X_2, \dots, X_n$ be a sample of i.i.d. random variables with distribution $\mathcal{P}(\lambda)$ where $\lambda \in \mathbb{R}_+^*$ is an unknown parameter.

1. Give an estimator of λ based on method of moment and maximum likelihood.
2. State the Central Limit Theorem for $\overline{X}_n$, the empirical mean.
3. Find a function $g : \mathbb{R}_+ \to \mathbb{R}$ such that $\sqrt{n}(g(\overline{X}_n) - g(\lambda))$ converges in distribution to a standard Gaussian variable.
4. Deduce an asymptotic confidence interval of level $1-\alpha$ for λ.

Exercise 3. DISCRETE DISTRIBUTION

Let $\theta \in (0,1)$ be an unknown parameter and we denote X a random variable with distribution defined for all $k \in \mathbb{N}$ by

$$\mathbb{P}_\theta(X = k) = (k+1)(1-\theta)^2\theta^k.$$

1. Show that $\mathbb{E}_\theta[X] = \frac{2\theta}{1-\theta}$ and $\mathrm{Var}_\theta[X] = \frac{2\theta}{(1-\theta)^2}$.
2. Give an estimator $\hat{\theta}_n$ of parameter θ by using method of moment.
3. Study the consistency of $\hat{\theta}_n$.
4. Give the asymptotic distribution of $\hat{\theta}_n$.

Exercise 4. Gaussian Mixing
Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables such that the distribution admits a density f defined by

$$f(x) = pf_{0,1}(x) + (1-p)f_{0,4}(x),$$

where parameter $p \in (0,1)$ is unknown and f_{m,σ^2} is the density function of gaussian distribution $\mathcal{N}(m, \sigma^2)$.

1. Which is the difficulty to estimate parameter p according to the maximum likelihood.
2. Give the estimator $\hat{p}_n$ of parameter p based on method of moment. We are going to use the second moment, why ?
3. Show that $\hat{p}_n$ is consistent and give the asymptotic distribution of $\sqrt{n}(\hat{p}_n - p)$.
4. **Bonus:** How can we simulate a random variable X with density f according to the simulation of two independent random variables $Z \sim \mathcal{N}(0,1)$ and $U \sim \mathrm{U}([0,1])$.

Exercise 5. Uniform Distribution
Let $X_1, X_2, \ldots, X_n$ be a n-sample (i.i.d. random variables) with distribution $\mathrm{U}([0,\theta])$ where $\theta \in \mathbb{R}_+^*$ unknown.

1. (a) Provide $\hat{\theta}_n$, the maximum likelihood estimator for θ and $\widetilde{\theta}_n$ the estimator by the moment method.
   (b) We denote $\overline{\theta}_n = 2x_{1/2}(n) = 2X_{(\lceil n/2 \rceil)}$. Why can we suggest $\overline{\theta}_n$ to estimate θ ?
   (c) Study the consistency and the asymptotic normality of $\hat{\theta}_n$ and $\widetilde{\theta}_n$. Compare both.
2. By using the distribution of $\hat{\theta}_n$, construct a confidence interval with exact level $1-\alpha$
3. Construct a test with exact level α by taking H_0: $\theta = \theta_0$ against $H_1 : \theta \neq \theta_0$.

Exercise 6. Exponential Model
Let $X_1, X_2, \ldots, X_n$ be a sample of i.i.d. random variables with exponential distribution $\varepsilon(\lambda)$ where $\lambda \in \mathbb{R}_+^*$ is an unknown parameter. In this exercise We want to estimate λ.

1. Construct an estimator of λ based on moment method and maximum likelihood and check that it is well defined. We denote this estimator Y_n.
2. Determine the limiting distribution of $\sqrt{n}(Y_n - \lambda)$.
3. Give the law of $\sum_{i=1}^{n} X_i$ and deduce the value of $\mathbb{E}[(Y_n - \lambda)^2]$.
4. In this question we suggest to use $Z_n = \frac{n-1}{n} Y_n$. Do Z_n satisfy the same properties as Y_n ?
5. Which of these two estimators would you choose to estimate λ ?
6. Let $\alpha \in (0,1)$. Provide an asymptotic confidence interval of level $1-\alpha$ for λ.
7. Propose an asymptotic α-level test for $H_0 : \lambda \geq 1$ against $H_1 : \lambda < 1$.
8. Propose an asymptotic α-level test for $H_0 : \lambda = 1$ against $H_1 : \lambda \neq 1$.

Exercise 7. Scaled Beta Model

We observe a sample $X_1, X_2, \dots, X_n$ of i.i.d. random variables with the following density :

$$f_\theta(x) = \frac{2}{\theta^2} x \mathbf{1}_{[0,\theta]},$$

where θ is unknown parameter.

1. We denote by $X_{(n)} = \max_{1 \leq i \leq n} X_i$. Give the density of $X_{(n)}$.
2. Compute the mean and variance of $X_{(n)}$.
3. Show that $X_{(n)}$ is a consistent estimator of parameter θ. Is it strongly consistent ?
4. Give another estimator of λ based on moment method.
5. Which of these two estimators is better to estimate θ in the sense of mean squared error ?
6. For all $x \in [0, \theta]$, show that
$$\mathbb{P}(x \leq X_{(n)} \leq \theta) = 1 - \frac{x^{2n}}{\theta^{2n}},$$
and construct a confidence interval of level $1 - \alpha$ by choosing x appropriately in the previous identity.
7. Construct an α-level test for the null hypothesis $\theta = 1$ against the alternative $\theta \neq 1$. Give the p-value when we observe $n = 20$ and $x_{(20)} = 0,85$.