CENTRALESUPÉLEC

First Year 2018-2019

Test I - Statistics and Learning

Without document.

Exercise 1

In order to evaluate the production potential of a wind turbine plant, wind speed is modelled by a random variable with Weibull distribution, the density of which is given by

$$f_\beta(x) = 2\beta^{-2} x \exp(-x^2/\beta^2)\mathbb{I}_{\mathbb{R}+}(x),$$

where β is a strictly positive parameter called the scale parameter.
We will admit that if $X \sim f_1$, then :

$$\mathbb{E}_1(X) = \int_{\mathbb{R}} x\, f_1(x)\, dx = \frac{\sqrt{\pi}}{2},\ \mathbb{E}_1(X^2) = \int_{\mathbb{R}} x^2 f_1(x)\, dx = 1 \text{ et } \ \mathbb{E}_1(X^4) = \int_{\mathbb{R}} x^4 f_1(x)\, dx = 2$$

We have a sample $X_1, \dots, X_N$ of i.i.d. random variables with density function f_{β^*} where β^* is the true value of the parameter.

1. Show that $\forall r \in \mathbb{N}^*$, $\mathbb{E}_\beta(X^r) = \beta^r \mathbb{E}_1(X^r)$. Deduce then $\mathbb{E}_\beta(X)$, $\mathbb{V}_\beta(X)$ and $\mathbb{V}_\beta(X^2)$ for all $\beta > 0$.
(N.B. : $\mathbb{E}_\beta$ and $\mathbb{V}_\beta$ denote the expectation and variance under the density function f_β).

2. a) Provide an estimator of β^* by using the method of moments based on the first moment. $\hat{\beta}_1$ denotes this estimator.

b) Calculate its bias and mean square error.

c) Show that $\hat{\beta}_1$ is convergent.

d) Determine the asymptotic distribution of $\sqrt{N}(\hat{\beta}_1 - \beta^*)$.

e) Make an asymptotically pivotal function for β^*. Deduce an asymptotic confidence interval at level 98% for β^* as a function of $q_{0.99}$ where $q_{0.99}$ is the $0.99-$quantile for the standard normal distribution.

3. a) Define the likelihood of the parameter β based on the sample $(x_1, \dots, x_N)$.

b) Show that the maximum likelihood estimator exists, that it is unique and that it is expressed as a function of the second order empirical moment of the sample. We will denote it by $\hat{\beta}_2$.

c) Show that $\hat{\beta}_2$ is convergent.

d) Based on $\hat{\beta}_2$ and by using the Delta method, determine a confidence interval for β at asymptotic level 0.98 as a function of $q_{0.99}$, the 0.99−quantile for the standard normal distribution.

Exercise 2

Let $(X_1, X_2, \ldots, X_N)$ be a sample of independent random variables, identically distributed from a statistical model parameterized by $\theta \in \Theta \subset \mathbb{R}$. In this exercise, we consider the likelihood ratio test for parametric type tests defined as following :

$$H_0 : \theta \in \Theta_0 \text{ against } H_1 : \theta \in \Theta_0^c$$

where $\Theta_0 \subsetneq \Theta$ is given and Θ_0^c is the complementary set of Θ_0 in Θ.

For this purpose, a test statistic is constructed :

$$\lambda(X_1, \ldots, X_N) = \frac{\sup_{\Theta} \mathcal{L}(\theta; X_1, \ldots, X_N)}{\sup_{\Theta_0} \mathcal{L}(\theta; X_1, \ldots, X_N)}$$

where $\mathcal{L}(\theta; X_1, \ldots, X_N)$ represents the likelihood function of the parameter θ for the sample $(X_1, \ldots, X_N)$. We are then led to a rejection zone defined by :

$$R_\alpha = \{(x_1, \ldots, x_N) \text{ such that } \lambda(x_1, \ldots, x_N) > c_\alpha\}$$

where c_α is chosen such that :

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta \left((X_1, \ldots, X_N) \in R_\alpha \right) = \alpha \ .$$

1) Let $X = (X_1, X_2, \ldots, X_N)$ be a sample of independent random variables, identically distributed from a normal distribution $\mathcal{N}(\mu, 1)$.

a) Calculate $\hat{\mu}$ the maximum likelihood estimator for μ defined on $\Theta = \mathbb{R}$ and give its distribution.

b) We want to perform a parametric hypothesis test :

$$H_0 : \mu = \mu_0 \text{ against } H_1 : \mu \neq \mu_0$$

where μ_0 is given.
Determine the likelihood ratio $\lambda(X)$ and deduce a simplified form of the rejection region.
Determine c_α.

2) Now, we consider a family of exponential distributions with densities of the form :

$$f(x; \theta) = \begin{cases} e^{-(x-\theta)} & x \geq \theta \\ 0 & x < \theta \end{cases}$$

Let $X = (X_1, X_2, \ldots, X_N)$ be a sample of independent random variables, identically distributed according to $f(x;\theta)$. We test :

$$H_0 : \theta \leq \theta_0 \text{ against } H_1 : \theta > \theta_0$$

where θ_0 is given.

a) Calculate $\hat{\theta}$ the maximum likelihood estimator for θ on $\Theta = \mathbb{R}$ by bringing out $X_{(1)} := \min_{1\leq i\leq N} X_i$.

b) Determine $\lambda(X)$ for the considered parametric test and deduce a simplified form of the rejection region. Determine c_α.

Exercise 3

We recall the definition of a multinomial distribution of order K. Let $N \in \mathbb{N}^*$, $p \in]0;1[^K$ such that $\sum_{i=1}^K p_i = 1$. The probability mass function of the multinomial distribution with parameter (N,p) is :

$$P(x_1,\ldots,x_K) = \begin{cases} \dfrac{N!}{\prod_{i=1}^K x_i!} \displaystyle\prod_{i=1}^K p_i^{x_i}, \text{ if } (x_1,\ldots,x_K) \in \{0;1;\ldots;N\}^K \text{ such that } \Sigma_{i=1}^K x_j = N \\ \\ 0 \qquad \text{otherwise.} \end{cases}$$

Let us denote $X \sim M(N,p)$.
We also recall the definition of a Dirichlet distribution of order K.
Let $a = (a_1,\ldots,a_K) \in \left(\mathbb{R}_+^*\right)^K$. The Dirichlet distribution of order K with parameter a and the support $\mathcal{S} = \left\{x \in [0;1]^K : \sum_{i=1}^K x_i = 1\right\}$ has a probability density function defined as :

$$p(x_1,\ldots,x_K) = \begin{cases} \dfrac{1}{\beta(a)} \displaystyle\prod_{i=1}^K {x_i}^{a_i-1} \text{ if } (x_1,\ldots,x_K) \in \mathcal{S} \\ 0 \qquad \text{otherwise} \end{cases}$$

It is denoted by $\text{Dir}(a)$. The Dirichlet distribution of ordre 2 is the Beta distribution, $\text{Dir}(a_1,a_2) = \text{Beta}(a_1,a_2)$.
1) Without performing any calculation, tell how to determine the function $a \mapsto \beta(a)$. We will assume that it is known for the followings.

2) Let Y be a random variable that follows a multinomial distribution of order K, $K \geq 3$ with parameters (N,θ), where N is known and $\theta = (\theta_1,\ldots,\theta_K)$ is unknown. Let $y = (y_1,\ldots,y_K)$ be an observation of the variable Y. We focus on the Bayesian estimation for θ and assume a prior distribution $\pi = \text{Dir}(a)$, with $a = (a_1,\ldots,a_K) \in \left(\mathbb{R}_+^*\right)^K$. Determine the posterior distribution $p(\theta|y)$.

3) a) Show that if $(X_1, \dots, X_{K-1}, X_K)$ has a Dirichlet distribution with parameter $(a_1, \dots, a_{K-1}, a_K)$, then $(X_1, \dots, X_{K-2}, X_{K-1} + X_K)$ has a Dirichlet distribution with parameter $(a_1, \dots, a_{K-2}, a_{K-1} + a_K)$.

b) We denote $a_r = \sum_{i=3}^{K} a_i$ and $y_r = \sum_{i=3}^{K} y_i$. Deduce from the previous question that

$$p(\theta_1, \theta_2 | y) \propto \theta_1^{a_1+y_1-1} \theta_2^{a_2+y_2-1} (1 - \theta_1 - \theta_2)^{a_r+y_r-1} \ .$$

4) We carry out the variable change ϕ :

$$(\alpha_1, \alpha_2) = \left(\frac{\theta_1}{\theta_1 + \theta_2}, \theta_1 + \theta_2 \right) = \phi(\theta_1, \theta_2).$$

a) Show that ϕ is a $\mathcal{C}^1$-Diffeomorphism of $]0; +\infty[^2$ onto $]0; 1[\times]0; +\infty[$.

b) Deduce the conditional density $p(\alpha_1, \alpha_2 | y)$ up to a normalizing constant.

c) Finally, deduce the probability law of the conditional density $p(\alpha_1 | y)$.