Université Paris-Dauphine Année 2014-2015
L3 - Statistique (1) : Estimation

Examen Partiel 04/11/2014 - Durée 2h00 - Documents autorisés

Problem 1 (6 pts)

Define $N \sim \mathcal{N}(\mu, 1)$ ($\mu \in \mathbb{R}$), $X = \exp(N)$ and a sample of size n $(X_1, \ldots, X_n)$ from X.

1) (2 pts) Show that the density of X is given by

$$f_X(x;\mu) = \frac{1}{\sqrt{2\pi}x} \exp\left\{-\frac{1}{2}(\log(x) - \mu)^2\right\} \mathbb{I}_{x>0}\,.$$

2) (2 pts) Compute $\mathbb{E}(X)$ (for instance via a change of variable $y = \log(x)$ and the identification of a relevant Gaussian density).

3) (2 pts) Set $T_n = \log\left({}^1\!/\!{}_n \sum_{i=1}^n X_i\right) - 1/2$. Show that T_n converges in probability to μ.

Problem 2 (4pts)

Consider an exponential family with probability density

$$f_\theta(x) = h(x)e^{\theta T(x)}e^{-\psi(\theta)}, \quad \theta \in \Theta = \{\theta \in \mathbb{R}; \int_{\mathbb{R}} h(x)e^{\theta T(x)} d\mu(x) < +\infty\} \tag{1}$$

with respect to the σ-finite measure ν on $\mathbb{R}$. We assume Θ is an open set of $\mathbb{R}$.

1) (1 pt) Establish that the Binomial $\mathcal{B}(m, \mu)$ distribution is an exponential family when the parameter is μ. Determine the natural (canonical) parameterisation, θ, and the natural parameter space Θ. Show that this exponential family is regular.

2) (1 pt) Establish that the Beta $\mathcal{B}e(\alpha\mu, \alpha(1-\mu))$ distribution is an exponential family when the parameter is μ. Determine the natural (canonical) parameterisation, θ, and the natural parameter space Θ. Show that this exponential family is regular.
Hint The density of the Beta $\mathcal{B}e(\alpha\mu, \alpha(1-\mu))$ distribution against the Lebesgue measure is equal to

$$f_{\alpha,\mu}(x) = \mathbf{1}_{(0,1)}(x) \frac{\Gamma(\alpha)}{\Gamma(\alpha\mu)\Gamma(\alpha(1-\mu))} x^{\alpha\mu - 1}(1-x)^{\alpha(1-\mu)-1}$$

3) (2 pts) Let $X_1, \cdots, X_n$ n be independent random variables with identical distribution $\mathcal{B}e(\alpha\mu, (1-\alpha)\mu)$. Derive the density (against the Lebesgue measure on $\mathbb{R}^n$) of $X^{(n)} = (X_1, \cdots, X_n)$ and show that $X^{(n)}$ has a distribution that also belongs to an exponential family. Identify all terms in (1).

Problem 3 (12 pts)

Given two independent samples $X^{(n_1)} = (X_1, \ldots, X_{n_1})$ and $Y^{(n_2)} = (Y_1, \ldots, Y_{n_2})$, with respective distributions F et G, set

$$\widehat{F}_{X^{(n_1)}}(x) = {}^1\!/_{n_1} \sum_{i=1}^{n_1} \mathbb{I}_{\{X_i \leq x\}}, \quad \text{and} \quad \widehat{G}_{Y^{(n_2)}}(x) = {}^1\!/_{n_2} \sum_{i=1}^{n_2} \mathbb{I}_{\{Y_i \leq x\}}$$

as their respective empirical cdfs.

A test of the potential equality $F = G$ is based on the decision rule : compute the distance

$$T_{n_1,n_2} = \sup_{x \in \mathbb{R}} \left| \widehat{F}_{X^{(n_1)}}(x) - \widehat{G}_{Y^{(n_2)}}(x) \right|$$

and reject the hypothesis that $F = G$ when T_{n_1,n_2} is too large.

1) (2 pts) Show that, if F also denotes the cdf of the random variable X and if we set

$$F^{-}(u) = \inf\{x;\, F(x) \geq u\}$$

then, when $U \sim \mathcal{U}(0,1)$, $F^{-}(U) \sim F$.

2) (1 pt) Check the above result when F is an exponential distribution $\mathcal{E}(\lambda)$ and when F is a Poisson distribution $\mathcal{P}(\lambda)$.

3) (1 pt) Provide an example of *and* a counter-example to the statement that, when $X \sim F$, then $F(X) \sim \mathcal{U}(0,1)$.

4) (2 pts) Show that, when F is strictly increasing, T_{n_1,n_2} is distributed as

$$\tilde{T}_{n_1,n_2} = \sup_{x \in \mathbb{R}} \left| {}^1\!/_{n_1} \sum_{i=1}^{n_1} \mathbb{I}_{\{U_i \leq F(x)\}} - {}^1\!/_{n_2} \sum_{i=1}^{n_2} \mathbb{I}_{\{V_i \leq G(x)\}} \right|,$$

where the U_i's and the V_i's are two independent uniform $\mathcal{U}[0,1]$ samples.

5) (2 pts) Deduce that, when $F = G$, the distribution of T_{n_1,n_2} does not depend on F.

6) (2 pt) Propose a Monte Carlo technique to approximate the distribution of T_{n_1,n_2}.

7) (2 pts) Show that, if the X_i's and the Y_i's are jointly sorted into an increasing vector

$$Z^{(n_1+n_2)} = (Z_1, \ldots, Z_{n_1+n_2}) \quad \text{with } Z_i \leq Z_{i+1} \; i = 1, \ldots, n_1 + n_2 - 1,$$

then, defining $S_k = {}^1\!/_{n_1}$ when Z_k is an element of $X^{(n_1)}$ and $S_k = {}^{-1}\!/_{n_2}$ when Z_k is an element of $Y^{(n_2)}$, the following representation

$$T_{n_1,n_2} = \max_{1 \leq k \leq n_1+n_2} \left| \sum_{\ell=1}^{k} S_\ell \right|$$

holds.