MAP531 : Statistics
Exam

Stéphanie Allassonnière, Elodie Vernet

The three exercises are independent from each other.

You are allowed to use your handwritten notes from the class and PC sessions as well as the material that was provided to you on moodle.

You are allowed to use a calculator.

Cell phones in particular smart phones are forbidden.

1 Exercices

EXERCICE 1 - We consider $X_1, \dots, X_n$ i.i.d. observations distributed from a distribution with density

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

1. Write the statistical model associated with these data.
2. Give the likelihood function $\mathcal{L}(X_1, \dots, X_n|\theta)$, which depends on $\max_i X_i$.
3. Compute the maximum likelihood estimator $\hat{\theta}_{m.l.e.}$. Help : You can draw the graph of $\theta \in (0, +\infty) \mapsto \mathcal{L}(X_1, \dots, X_n|\theta)$.
4. Compute the following probability :
$$\mathbb{P}_\theta\left(\frac{\max_i X_i}{\theta} \leq t\right),$$
for $t \in \mathbb{R}$. You will pay attention to the cases where $t < 0$ and $t > 1$.
5. Give the density function associated to the distribution of $\frac{\max_i X_i}{\theta}$.
6. Using the previous questions, give the bias, variance and mean square error of the maximum likelihood estimator $\hat{\theta}_{m.l.e.}$.
7. Using the method of moments, give another estimator $\hat{\theta}_{mo}$ of θ.
8. Compute the bias, variance and mean square error of $\hat{\theta}_{mo}$.
9. Which estimator $\hat{\theta}_{m.l.e}$ or $\hat{\theta}_{mo}$, would you choose ? Why ?
10. Using Question 4, give a 95% confidence interval for θ.

EXERCICE 2 - We observe a unique observation X_1 distributed from a distribution $\mathbb{P}_\theta$ with density function :

$$f_\theta(x) = \frac{\theta}{(x+\theta)^2} \mathbb{1}_{\mathbb{R}_+}(x), \quad \theta > 0.$$

We want to test

$$H_0 : \theta = \theta_0 \quad \text{against} \quad H_1 : \theta = \theta_1,$$

where $\theta_0 < \theta_1$.

1. Write the statistical model.
2. Write the likelihood ratio. Is it an increasing or decreasing function of the observation ?
3. Give the likelihood ratio test at level $\alpha \in (0,1)$ for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$. The constants will be precised and will only depend on θ_0 and α.
4. Is this test $UMP(\alpha)$? Why ?
5. Compute the type II risk so that it only depends on α, θ_0 and θ_1.
6. We consider the particular case where $\alpha = 5\%$, $\theta_0 = 1$ and $\theta = 3$. We observe $x_1 = 10$, do we reject H_0 using the likelihood ratio test ?

2 Problem : introduction to the EM algorithm

This exercise will show you a very useful algorithm which has been introduced in the 1977 by Dempster, Laird and Rubin but which is still very often used. This algorithm called Expectation - Maximization (EM) is very powerful in order to estimate parameters from a mixture distribution.

Let us first recall the definition of a mixture of densities :

Def : *Let $f_1,\ f_2$ be two densities, and let $w_1,\ w_2 \in (0,1)$, such that $w_1 + w_2 = 1$. The mixture of densities $f_1,\ f_2$ with weights $w_1,\ w_2$ is the density given by $x \mapsto w_1 f_1(x) + w_2 f_2(x)$.*

We also recall that the mixture density has the following interpretation : let $X,\ Y$ have density $f_1,\ f_2$, and let $Z = 1$ (*resp.* $Z = 2$) with probability w_1 (resp. w_2). If $X,\ Y,\ Z$ are independent, then $X\mathbb{1}_{Z=1} + Y\mathbb{1}_{Z=2}$ has density $w_1 f_1 + w_2 f_2$.

1. Generalize the previous definition for a mixture of K densities.

Let us now describe the set of observations.

We are given a set of n observations which describe the relative spinal bone mineral density measurements on n North American adolescents. Each value is the difference taken on two consecutive visits, divided by the average. The observations are given by their histogram in Figure**??**.

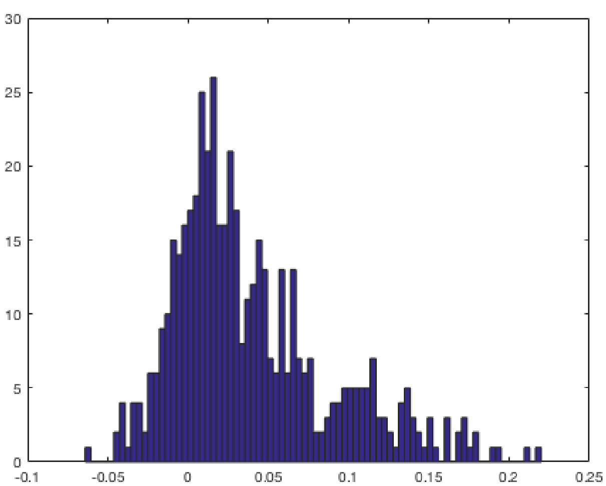

FIGURE 1 – Histogram of spinal bone mineral density measurements

2. Propose two possible mixture models for these data. Provide in particular the chosen forms of the densities f_k for $1 \leq k \leq K$ and the parameters of these distributions.

The problem now is to estimate the parameters of this statistical model thanks to the observations.

3. Which estimator you have seen in this course would you rather use for this question ? Why ?

Let us now fix some notations : in the following, we denote $(X_1, ..., X_n)$ the sequence of i.i.d random variables from which we observe a realisation which consists in our observations (the spinal bone mineral density measurements). We denote Z_i the random variable defining the cluster of observation i which is not observed (this random variable is hidden to the statistician). Given that $Z = k$, we model the measurements of cluster k by a Gaussian distribution with mean m_k and variance σ_k^2.

4. Write the probability distribution of Z as a finite sum of Dirac measures and the conditional distribution of X given $Z = k$. We recall that a Dirac measure at some point a, usually denoted by δ_a, puts all the mass on a :
$$\delta_a(A) = \begin{cases} 1 & \text{if } a \in A \\ 0 & \text{otherwise.} \end{cases}$$
5. Write the statistical model : space of observations, parameter θ, set of possible parameters Θ and the parametric family of distributions $\mathcal{P}_\theta$.
6. Write the likelihood of the observation set.

Unfortunately this likelihood is not straightforward to maximize. In order to do it, we use an iterative algorithm which will first rely on the following quantity :

$\log \mathcal{L}_c^{(\theta)}(x_1, ..., x_n, z_1, ..., z_n)$ the log −likelihood of what is called the complete data $(x_i, z_i)_{1\leq i\leq n}$ where z_i represents the cluster of observation i.

7. Prove the following equality :
$$\log \mathcal{L}_c^{(\theta)}(x_1, ..., x_n, z_1, ..., z_n) = \sum_{i=1}^{n} \left[\sum_{k=1}^{K} \mathbb{1}_{\{z_i=k\}} \log\left(f_{\theta_k}(x_i) w_k\right) \right]$$
where we have set $\theta_k = (m_k, \sigma_k^2)$.
8. Justify that $\mathbb{P}(Z_i = k|x_1, ..., x_n) = \mathbb{P}(Z_i = k|x_i)$.
9. Using the Bayes rule, prove that
$$\mathbb{P}_\theta(Z_i|x_i) = \sum_{k=1}^{K} \mathbb{1}_{\{z_i=k\}} \frac{w_k f_{\theta_k}(x_i)}{f_\theta(x_i)} \tag{1}$$
where $f_\theta(x_i)$ is the probability density function of x_i depending on the whole set of parameters $\theta = (\theta_k, w_k)_{1\leq k\leq K}$. We now denote $\mathbb{P}_\theta$ this conditional distribution.
10. Let us define $p_{i,k} \triangleq \mathbb{E}_{\mathbb{P}_\theta}[\mathbb{1}_{Z_i=k}]$, the conditional expectation of $\mathbb{1}_{Z_i=k}$ given the observation X_i for a parameter θ. Compute the following quantity :
$$Q(\theta, \theta') = \mathbb{E}_{\mathbb{P}_{\theta'}}[\log(\mathcal{L}_c^{(\theta)})]. \tag{2}$$
as a function of θ and the $p'_{i,k} = \mathbb{E}_{\mathbb{P}'_\theta}[\mathbb{1}_{Z_i=k}]$ (for another parameter θ').
11. Compute now $\underset{\theta}{\operatorname{argmax}}\, Q(\theta, \theta')$ by differentiation the previous equation with respect to all the parameters.

In order to construct the sequence $(\theta_t)_{t\in\mathbb{N}}$ which will converge to the maximum likelihood estimate of θ given our set of observations, we iterate the following steps which define the EM algorithm :

Step 0 : Initialization : We initialize θ_0 at random and $t = 0$. Then iterate :

Step E : Expectation : Compute $Q(\theta, \theta_t)$.

Step M : Maximization : Compute $\theta_{t+1} = arg\max_\theta Q(\theta, \theta_t)$ and $t = t + 1$.