ECUE21.1 : Science des données **QCM 3 – Construction d'estimateurs**

29 avril 2020

Question 1. Soit $(x_1,x_2,\dots,x_n)$ un échantillon d'une variable aléatoire X. On suppose que X suit une loi paramétrisée par γ. La vraisemblance de $(x_1,x_2,\dots,x_n)$ est donnée par

- ☐ $\mathbb{P}(x_1,x_2,\dots,x_n,\gamma)$
- ☐ $\mathbb{P}(x_1,x_2,\dots,x_n|\gamma)$
- ☐ $\mathbb{P}(\gamma|x_1,x_2,\dots,x_n)$
- ☐ $\prod_{i=1}^n \mathbb{P}(x_i|\gamma)$
- ☐ $\prod_{i=1}^n \mathbb{P}(\gamma|x_i)$

Question 2. Soit X une loi exponentielle de paramètre λ. L'estimateur par maximum de vraisemblance de λ est donné par

- ☐ $L_n = n\ln(\lambda) - \lambda\sum_{i=1}^n X_i$, où $(X_1,X_2,\dots,X_n)$ est un échantillon aléatoire de X
- ☐ $\widehat{\lambda} = n\ln(\lambda) - \lambda\sum_{i=1}^n x_i$, où $(x_1,x_2,\dots,x_n)$ est un échantillon aléatoire de X
- ☐ $L_n = \frac{n}{\sum_{i=1}^n X_i}$, où $(X_1,X_2,\dots,X_n)$ est un échantillon aléatoire de X
- ☐ $\widehat{\lambda} = \frac{n}{\sum_{i=1}^n x_i}$, où $(x_1,x_2,\dots,x_n)$ est un échantillon aléatoire de X.

Question 3. ★ L'estimateur de Bayes est plus proche de l'espérance a priori que de l'estimateur par maximum de vraisemblance quand la taille de l'échantillon est

- ☐ grande
- ☐ petite
- ☐ ça dépend.

Solution

Question 1. Par définition (cf. équation (3.7) du poly),

$$L(x_1,x_2,\dots,x_n;\gamma) = \mathbb{P}(x_1,x_2,\dots,x_n|\gamma) = \prod_{i=1}^{n} \mathbb{P}(x_i|\gamma).$$

Question 2. Par définition la vraisemblance d'un échantillon $(x_1,x_2,\dots,x_n)$ est donnée par

$$L(x_1,x_2,\dots,x_n;\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n \prod_{i=1}^{n} e^{-\lambda x_i},$$

et donc sa *log-vraisemblance* vaut

$$\ell(x_1,x_2,\dots,x_n;\lambda) = \ln\left(\lambda^n \prod_{i=1}^{n} e^{-\lambda x_i}\right) = n\ln(\lambda) - \lambda \sum_{i=1}^{n} x_i.$$

La fonction $\lambda \mapsto n\ln(\lambda) - \lambda \sum_{i=1}^{n} x_i$ est concave sur $]0,+\infty[\to \mathbb{R}$ et on peut donc la maximiser en annulant sa dérivée.

On obtient *l'estimation par maximum de vraisemblance* de λ suivante :

$$\widehat{\lambda}_{\text{MLE}} = \frac{n}{\sum_{i=1}^{n} x_i}$$

et, si on appelle $(X_1,X_2,\dots,X_n)$ un échantillon aléatoire de X, on obtient *l'estimateur par maximum de vraisemblance* de λ :

$$L_n = \frac{n}{\sum_{i=1}^{n} X_n}.$$

Question 3. La tendance que nous avons observée sur l'exemple de la section 3.6 (cf. « Remarque importante ») se vérifie en général : plus on observe d'échantillons, plus on s'éloigne de l'a priori pour se rapprocher d'un estimateur issu uniquement des données.