Université de Carthage
Ecole Supérieure de la Statistique
et de l'Analyse de l'Information

Année Universitaire
2017-2018
1ère Année

Devoir Surveillé
Statistique inférentielle 1

Février 2018

Aucun document autorisé. Toutes les réponses doivent être justifiées

Cours de Mme Héla Ouaili-Mallek — Durée : 1h 30mn

Exercice 1 *On dispose d'un n-échantillon* $(X_1, \dots, X_n)$ *de la variable aléatoire* X *de loi* P_θ *issue de la famille* $\{P_\theta\}$
où $\theta = (\theta_1, \dots, \theta_{k-1}) \in (]0\ ,\ 1[)^{k-1}$ *et* $X(\Omega) = \{\omega_1, \dots, \omega_k\}$.

P_θ *admet une densité par rapport à la mesure de comptage définie par*
$P_\theta[X = \omega_j] = \theta_j$, *pour* $1 \leq j \leq k$, *avec* $\theta_k = 1 - \sum_{j=1}^{k-1} \theta_j$.

1. *Ecrire la vraisemblance de ce modèle.*
2. *En déduire une statistique exhaustive pour* θ.
3. *Vérifier si le modèle appartient à la famille exponentielle.*
4. *Déterminer* $\widehat{\theta}_{MV}$, *l'estimateur du maximum de vraisemblance de* θ.
5. *Montrer que* $\widehat{\theta}_{MV}$ *coïncide avec l'estimateur de* θ *par substitution des fréquences.*

Exercice 2 *On dispose d'un n-échantillon* $(X_1, \dots, X_n)$ *de la variable aléatoire* X *de densité*

$$f(x, \theta) = \begin{cases} \frac{2}{\theta}\left(1 - \frac{x}{\theta}\right) & si\ 0 \leq x \leq \theta \qquad \theta > 0 \\ 0 & sinon \end{cases}$$

L'objectif est d'estimer le paramètre inconnu θ.

1. *Ecrire la vraisemblance de l'échantillon.*
2. *Calculer* $E_\theta[X]$ *et* $Var[X]$.
3. *En déduire un estimateur* $\widehat{\theta}_1$ *de* θ *par la méthode des moments. Vérifier qu'il est sans biais et convergent.*
4. *Est-il efficace?*
5. *Proposer un résumé exhaustif pour le modèle.*
6. *Peut-on proposer un estimateur du maximum de vraisemblance pour* θ?

Corrigé Exercice 1

$\theta = (\theta_1, ..., \theta_{k-1}) \in (]0\ ,\ 1[)^{k-1}$ $\quad P_\theta[X = \omega_j] = \theta_j$, pour $1 \leq j \leq k$, $\theta_k = 1 - \sum_{j=1}^{k-1} \theta_j$.

1. $P_\theta[X = x] = \prod_{j=1}^{k} \theta_j^{\mathbb{1}_{\{\omega_j\}}(x)} \mathbb{1}_{\{\omega_1,...,\omega_k\}}(x)$

$$\mathcal{L}(\underline{x}, \theta) = \prod_{i=1}^{n}\prod_{j=1}^{k} \theta_j^{\mathbb{1}_{\{\omega_j\}}(x_i)} \mathbb{1}_{\{\omega_1,...,\omega_k\}^n}(\underline{x}) = \prod_{j=1}^{k} \theta_j^{\sum_{i=1}^{n}\mathbb{1}_{\{\omega_j\}}(x_i)} \mathbb{1}_{\{\omega_1,...,\omega_k\}^n}(\underline{x})$$

$$\mathcal{L}(\underline{x}, \theta) = \prod_{j=1}^{k} \theta_j^{n_j} \mathbb{1}_{\{\omega_1,...,\omega_k\}^n}(\underline{x}) = \prod_{j=1}^{k-1} \theta_j^{n_j}\left(1 - \sum_{j=1}^{k-1}\theta_j\right)^{n-\sum_{j=1}^{k-1} n_j} \mathbb{1}_{\{\omega_1,...,\omega_k\}^n}(\underline{x})$$

2. $\mathcal{L}(\underline{x}, \theta) = g(T(\underline{x}), \theta) * h(\underline{x})$

   avec $g(T(\underline{x}), \theta) = \prod_{j=1}^{k-1} \theta_j^{n_j}\left(1 - \sum_{j=1}^{k-1}\theta_j\right)^{n-\sum_{j=1}^{k-1} n_j}$ et $h(\underline{x}) = \mathbb{1}_{\{\omega_1,...,\omega_k\}^n}(\underline{x})$.

   D'après le théorème de factorisation, la statistique $T(\underline{X}) = (N_1, ..., N_{k-1})$ est exhaustive pour θ.

3. $\mathcal{L}(\underline{x}, \theta) = \exp\left(\sum_{j=1}^{k} n_j \ln\theta_j + \left(n - \sum_{j=1}^{k-1} n_j\right)\ln\left(1 - \sum_{j=1}^{k-1}\theta_j\right)\right)\mathbb{1}_{\{\omega_1,...,\omega_k\}^n}(\underline{x})$

$$\mathcal{L}(\underline{x}, \theta) = \exp\left(\sum_{j=1}^{k} n_j \ln\theta_j + n\ln\left(1 - \sum_{j=1}^{k-1}\theta_j\right) - \sum_{j=1}^{k-1} n_j \ln\left(1 - \sum_{j=1}^{k-1}\theta_j\right)\right)\mathbb{1}_{\{\omega_1,...,\omega_k\}^n}(\underline{x})$$

$$\mathcal{L}(\underline{x}, \theta) = \exp\left(\sum_{j=1}^{k} n_j\left(\ln\theta_j - \ln\left(1 - \sum_{j=1}^{k-1}\theta_j\right)\right) + n\ln\left(1 - \sum_{j=1}^{k-1}\theta_j\right)\right)\mathbb{1}_{\{\omega_1,...,\omega_k\}^n}(\underline{x})$$

$$\mathcal{L}(\underline{x}, \theta) = \exp\left(\sum_{j=1}^{k} N_j \ln\left(\frac{\theta_j}{1 - \sum_{j=1}^{k-1}\theta_j}\right) - + n\ln\left(1 - \sum_{j=1}^{k-1}\theta_j\right)\right)\mathbb{1}_{\{\omega_1,...,\omega_k\}^n}(\underline{x})$$

   Le modèle appartient à la famille exponentielle à $k-1$ paramètres avec

   $c_j(\theta) = \ln\frac{\theta_j}{1 - \sum_{j=1}^{k-1}\theta_j}$, $\quad 1 \leq j \leq k-1 \qquad d(\theta) = n\ln\left(1 - \sum_{j=1}^{k-1}\theta_j\right)$

   $T_j(\underline{X}) = N_j$; $\ 1 \leq j \leq k-1 \qquad A = \{\omega_1, ..., \omega_k\}^n$ indépendant de θ.

4. L'application $c : (\theta_1, ..., \theta_{k-1}) \longrightarrow \left(\ln \frac{\theta_1}{1 - \sum_{j=1}^{k-1} \theta_j}, ... \ln \frac{\theta_{k-1}}{1 - \sum_{j=1}^{k-1} \theta_j} \right)$ est injective et de classe C^2.

   L'application $d \longmapsto n \ln \left(1 - \sum_{j=1}^{k-1} \theta_j\right)$ est de classe C^2. $\Theta = (]0\ ,\ 1[)^{k-1}$ est un ouvert.

   $E_\theta [T(\underline{X})] = T(\underline{x}) \iff E_\theta [(N_1, ..., N_{k-1})] = (n_1, ..., n_{k-1}) \iff E_\theta [N_j] = n_j,\ 1 \leq j \leq k-1$

   $E_\theta [N_j] = E_\theta \left[\sum_{i=1}^{n} \mathbb{1}_{\{\omega_j\}}(X_i) \right] = \sum_{i=1}^{n} E_\theta \left[\mathbb{1}_{\{\omega_j\}}(X_i)\right] = \sum_{i=1}^{n} P[X_i = \omega_j] = n\theta_j$

   $E_\theta [N_j] = n_j,\ 1 \leq j \leq k-1 \iff n\theta_j = n_j \Longrightarrow \widehat{\theta}_j = \frac{N_j}{n},\ 1 \leq j \leq k-1$

   L'emv est $\widehat{\theta}_{MV} = \left(\widehat{\theta}_1, ..., \widehat{\theta}_{k-1}\right)$.

5. L'estimateur par substitution de fréquences de θ est

6. $\widehat{\theta} = \left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{\omega_1\}}(X_i), ..., \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{\omega_{k-1}\}}(X_i) \right)$

   $\widehat{\theta} = \left(\frac{1}{n} N_1, ..., \frac{1}{n} N_{k-1}\right) = \widehat{\theta}_{MV}$

Corrigé Exercice 1

1. $f(x) = \frac{2}{\theta}\left(1 - \frac{x}{\theta}\right) \mathbb{1}_{\{\min x_i \geq 0\}} \mathbb{1}_{\{\max x_i \leq \theta\}}$

   $\mathcal{L}(\underline{x}, \theta) = \left(\frac{2}{\theta}\right)^n \prod_{i=1}^{n} \left(1 - \frac{x_i}{\theta}\right) \mathbb{1}_{\{\min x_i \geq 0\}} \mathbb{1}_{\{\max x_i \leq \theta\}}$

2. $E_\theta [X] = \int_0^\theta \frac{2x}{\theta}\left(1 - \frac{x}{\theta}\right) dx = \int_0^\theta \frac{2x}{\theta} - \frac{2x^2}{\theta^2} dx$

   $E_\theta [X] = \left[\frac{x^2}{\theta} - \frac{2x^3}{3\theta^2}\right]_0^\theta = \theta - \frac{2}{3}\theta = \frac{\theta}{3}$

   $E_\theta [X^2] = \int_0^\theta \frac{2x^2}{\theta}\left(1 - \frac{x}{\theta}\right) dx = \int_0^\theta \frac{2x^2}{\theta} - \frac{2x^3}{\theta^2} dx$

   $E_\theta [X^2] = \left[\frac{2x^3}{3\theta} - \frac{x^4}{2\theta^2}\right]_0^\theta = \frac{2\theta^2}{3} - \frac{\theta^2}{2} = \frac{\theta^2}{6}$

   $Var[X] = \frac{\theta^2}{18}$

3. $E_\theta[X] = \dfrac{\theta}{3} \Longleftrightarrow \theta = 3\mu_1(\theta) \Longrightarrow \widehat{\theta}_1 = 3\widehat{m}_1 = 3\overline{X}_n$

   $E_\theta\left[\widehat{\theta}_1\right] = E_\theta\left[3\overline{X}_n\right] = \theta$

   D'après la loi des grands nombres, $\overline{X}_n$ converge presque sûrement vers $E_\theta[X]$. D'où $\widehat{\theta}$ converge presque sûrement vers $3E_\theta[X] = \theta$.

4. L'hypothèse H_3 (on peut dériver au moins 2 fois par rapport à θ sous le signe intégrale la quantité $\int_0^\theta \ln f(x,\theta)\,dx$) ne peut être valide car la borne sup de l'intégrale dépend de θ.

5. $\mathcal{L}(\underline{x},\theta) = \left(\dfrac{2}{\theta}\right)^n \prod_{i=1}^{n}\left(1 - \dfrac{x_i}{\theta}\right) \mathbb{1}_{\{\min x_i \geq 0\}} \mathbb{1}_{\{\max x_i \leq \theta\}}$

   $\mathcal{L}(\underline{x},\theta) = \exp\left(-2n\ln\theta + n\ln 2 + \sum_{i=1}^{n}\ln(\theta - x_i)\right) \mathbb{1}_{\{\min x_i \geq 0\}} \mathbb{1}_{\{\max x_i \leq \theta\}}$

   Le théorème de factorisation ne s'applique pas. Le seul résumé exhaustif est la statistique triviale.

6. Le modèle n'admet pas de statistique exhaustive. Or l'emv, lorsqu'il existe, dépend seulement de la statistique exhaustive.