Université Paris-Dauphine — Année 2014-2015

L3 - Statistique (1) : Estimation

Examen Final 06/01/2015 – solution

Exercice 1 (5 pts)

1 Un modèle statistique est associé au vecteur $(X_1, \ldots, X_n)$ en supposant que $X_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu, \sigma_i^2)$, avec μ et les σ_i supposés inconnus. Pour une observation $(x_1, \ldots, x_n)$, un estimateur du maximum de vraisemblance de μ est donné par
(a) $\hat\mu = x_2$ (b) $\hat\mu = (x_1 + \ldots + x_n)/n$ (c) $\sum_{i=1}^n (x_i - \hat\mu)^{-1} = 0$ (d) $|\hat\mu| = \min |x_i|$.
car prendre $\sigma_2 = 0$ conduit à une vraisemblance infinie

2 Si $X_1, \ldots, X_n$ et $Y_1, \ldots, Y_n$ sont des échantillons iid aléatoires de lois $\mathcal{E}(\lambda)$ et $\mathcal{E}(1/\lambda)$, respectivement, l'estimateur du maximum de vraisemblance de λ associé à l'ensemble des observations $(x_1, \ldots, x_n, y_1, \ldots, y_n)$ est donné par
(a) $\hat\lambda = \bar y/\bar x$ (b) $\hat\lambda = (\bar x + \bar y)/2$ (c) $\hat\lambda = \sqrt{\bar x \bar y}$ (d) $\hat\lambda = \sqrt{\bar y/\bar x}$ (e) $\hat\lambda = \sqrt{\bar x/\bar y}$.
par derivation de la vraisemblance

3 Si $X_1, \ldots, X_n$ est un échantillon iid de loi $\mathcal{B}eta(\lambda, 1)$, l'estimateur du maximum de vraisemblance de λ associé à l'échantillon $(x_1, \ldots, x_n)$ est donné par
(a) $\hat\lambda = -n \big/ \sum_{i=1}^n \log\{x_i\}$ (b) $\hat\lambda = 1/\bar x$ (c) $\hat\lambda = -\sum_{i=1}^n \log\{x_i\}/n$ (d) $\hat\lambda = -\log\left\{\prod_{i=1}^n (1 - x_i)\right\}$.
encore par simple derivation de la vraisemblance, en se rappelant que $\Gamma(\lambda + 1) = \lambda\Gamma(\lambda)$

4 Étant donné $X_1, \ldots, X_n$ un échantillon iid de loi $\mathcal{E}(\lambda)$, on observe $(Z_1, \ldots, Z_n) = (\mathbb{I}(X_1 \le \delta), \ldots, \mathbb{I}(X_n \le \delta))$, avec δ connu. Si $(z_1, \ldots, z_n) = (0, \ldots, 0)$, l'estimateur du maximum de vraisemblance de λ est donné par :
(a) $\hat\lambda = 1/\bar x$ (b) $\hat\lambda = \delta$ (c) $\hat\lambda = \bar x$ (d) $\hat\lambda = 1/\delta$ (e) $\hat\lambda = +\infty$ (f) $\hat\lambda = 0\}$,.
parce que la vraisemblance vaut $\exp\{-n\delta\lambda\}$ *et est maximale pour* $\lambda = 0$

5 Étant donné $X_1, \ldots, X_n$ un échantillon iid de loi $\mathcal{E}(\lambda)$, on observe $(Z_1, \ldots, Z_n) = (\mathbb{I}(X_1 \le \delta), \ldots, \mathbb{I}(X_n \le \delta))$, avec δ connu. Si $z_1 = 1$ et $(z_2, \ldots, z_n) = (0, \ldots, 0)$, l'estimateur du maximum de vraisemblance de λ est donné par :
(a) $\hat\lambda = -\log\{\frac{n-1}{n}\}/\delta$ (b) $\hat\lambda = \frac{(n-1)\delta}{n}$ (c) $\hat\lambda = \frac{\log\{\delta\}}{n-1}$ (d) $\hat\lambda = \frac{n+1}{\delta}$ (e) $\hat\lambda = 0$.
parce que la vraisemblance vaut $(1 - \exp\{-\delta\lambda\})\exp\{-(n-1)\delta\lambda\}$

Exercice 2 (7 pts)

Soit un échantillon aléatoire

$$X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(0, \theta^2), \quad \theta > 0$$

1 Représenter ce modèle comme une famille exponentielle : donner la paramétrisation naturelle, la représentation minimale, et indiquer si la famille est régulière.
Cours : ($T(x) = -x^2/2$, $\tau(\theta) = 1/\theta^2$, ce qui implique que la paramétrisation naturelle est en $1/\theta^2$

2 Donner une statistique exhaustive minimale T associée à l'échantillon et sa densité.
Solution : $T(x_1, \dots, x_n) = \sum_{i=1}^n x_i^2$ *qui suit une loi du chi-deux* $\Gamma(n/2, 1/2\theta^2)$

3 Donner un estimateur sans biais de $\Phi(-1/\theta)$ fondé sur X_1, noté $\delta(X_1)$, où $\Phi(\cdot)$ est la fonction de répartiion de la loi normale $\mathcal{N}(0,1)$.
Solution : $\Phi(-1/\theta) = \mathbb{P}(Z \leq -1/\theta) = \mathbb{P}(\theta Z \leq -1) = \mathbb{P}(X_1 \leq -1)$, *donc* $\mathbb{I}(X_1 \leq 1)$ *est un estimateur sans biais de* $\Phi(-1/\theta)$

4 Donner la loi conditionnelle de X_1 sachant $T(X_1, \dots, X_n)$.
Quand $T(x_1, \dots, x_n) = \rho^2$, $((X_1, \dots, X_n)$ *sachant* $T = \rho$ *est uniforme sur la sphère et donc* $X_1|\rho$ *a la même loi que* $\rho \sin(2\pi U)$ *avec* U *uniforme* $\mathcal{U}(0,1)$.

5 En déduire $\delta^\star(X_1, \dots, X_n) = \mathbb{E}[\delta(X_1)|T(X_1, \dots, X_n)]$.
Solution : $\mathbb{P}(X_1 \leq -1|T = \rho) = \mathbb{P}(\rho \sin(2\pi U) \leq -1|T = \rho)$

Expliquer pourquoi la variance de $\delta^\star(X_1, \dots, X_n)$ est inférieure à celle de $\delta(X_1)$.
Théorème de Rao–Blackwell

Exercice 3 (8 pts)

Soit un échantillon aléatoire $X_1, \cdots, X_n$ de variables indépendantes et identiquement distribuées selon une loi $\mathcal{E}(\lambda)$, $\lambda > 0$. On observe une réalisation $(x_1, \dots, x_n)$.

1 Ecrire la fonction de vraisemblance de l'échantillon.
Solution : $\lambda^n \exp\{-\lambda \sum x_i\}$

2 Calculer le maximum de vraisemblance de λ.
Solution : $\hat{\lambda} = 1/\bar{x}_n$

Déterminer sa loi pour tout $n \geq 1$.
Comme $\bar{X}_n \sim \Gamma(n, n\lambda)$, $\hat{\lambda} \sim \Gamma^{-1}(n, n\lambda)$

En déduire l'estimateur du maximum de vraisemblance de $\exp\{-\lambda\}$.
Si $\eta = \exp\{-\lambda\}$, $\hat{\eta} = \exp\{-\hat{\lambda}\}$

3 On considère, dans une approche bayésienne, une loi de probabilité a priori Gamma sur λ, $\lambda \sim \Gamma(a, b)$.
Pour tout $(a, b) \in (\mathbb{R}^{+*})^2$, déterminer la loi a posteriori de λ.
Solution : $\lambda|x_1, \dots, x_n \sim \Gamma(a + n, b + n\bar{x}_n)$

En déduire que la famille des lois Gamma est une famille conjuguée pour le modèle $\mathcal{E}(\lambda)$.
Cours : C'est toujours une loi Gamma

4 Déterminer l'espérance a posteriori $\mathbb{E}^\pi[\lambda|X_1, \cdots, X_n]$, issue de la loi a posteriori de la question précédente. On notera $\hat{\lambda}_n^\pi$ cette quantité.
Solution : $\hat{\lambda}_n^\pi = {}^{a+n}/_{b+n\bar{x}_n}$

5 Montrer que $\sqrt{n}(\hat{\lambda}_n^\pi - \lambda)$ converge en loi vers une loi normale dont on déterminera l'espérance et la variance.
Par la delta-method, $\hat{\lambda}_n^\pi \approx \bar{x}_n^{-1} \sim \mathcal{N}(\lambda, \lambda^2)$

Exercice 4 (4 pts)

1 Donner la définition d'une statistique exhaustive.
Cf. cours

2 Dans une famille exponentielle

$$f_\theta(x) = h(x)e^{\theta^t T(x) - \psi(\theta)} \tag{1}$$

donner l'expression de l'espérance et de la variance de $T(X)$ en fonction des dérivées de ψ.
Cf. cours

3 Calculer l'information de Fisher en θ dans le modèle (1).
Cf. cours

4 Quelle est l'expression analytique de la variance associée à la distribution bootstrap de l'échantillon $(x_1, \ldots, x_n)$?
Solution : var$(X^*) = \mathbb{E}[(X^* - \mathbb{E}[X^*])^2] = \mathbb{E}[(X^* - \bar{x})^2] = {}^1\!/\!_n \sum_i (x_i - \bar{x}_n)^2$