Université Paris-Dauphine — Année 2015-2016

L3 - Statistique (1) : Estimation

Examen Final 08/01/2016 — Durée 2h00 — Documents non autorisés

On rappelle que
- la loi Normale $\mathcal{N}(\mu, \sigma^2)$ a pour densité $f_{\mu,\sigma^2}(x) = e^{-(x-\mu)^2/(2\sigma^2)}/\sqrt{2\pi}\sigma$
- la loi Exponentielle $\mathcal{E}(\lambda)$ a pour densité $f_\lambda(x) = \lambda e^{-\lambda x}\mathbf{1}_{x>0}$
- la loi Beta $\mathcal{B}eta(a, b)$ a pour densité $f_{a,b}(x) = x^{a-1}(1-x)^{b-1}\mathbf{1}_{x\in(0,1)}\Gamma(a+b)/\Gamma(a)\Gamma(b)$
- la loi Gamma $\text{Ga}(a, b)$ a pour densité $f_{a,b}(x) = \mathbf{1}_{x>0}x^{a-1}e^{-bx}b^a/\Gamma(a)$
- la loi de Pareto $\mathcal{P}(\alpha, \mu)$ a pour densité $f_{\alpha,\mu}(x) = \frac{\alpha\mu^\alpha}{x^{\alpha+1}}\mathbf{1}_{x\geq\mu}$

Exercice 1 (6 pts)

Dans cet exercice il vous est demandé de donner la bonne réponse, seules les réponses justifiées dans la copie seront validées. Il n'y a pas de points négatifs.

1 Un modèle statistique est associé au vecteur $(X_1, \ldots, X_n)$ en supposant que $X_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu_i, \sigma)$, avec les μ_i et σ supposés inconnus. Pour une observation $(x_1, \ldots, x_n)$, un estimateur du maximum de vraisemblance de σ est donné par
(a) $\hat{\sigma} = 1$ (b) $\hat{\sigma} = 0$ (c) $\hat{\sigma} = +\infty$ (d) $\hat{\sigma} = \sqrt{\sum_{i=1}^n (x_i - \bar{x}_n)^2/n}$ (e) $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n x_i^2$.

2 Si $X_1, \ldots, X_n$ et $Y_1, \ldots, Y_n$ sont des échantillons iid aléatoires de lois $\mathcal{E}(\lambda)$ et $\mathcal{E}(2\lambda)$, 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) $\hat{\lambda} = 2\bar{y} - \bar{x}$ (b) $\hat{\lambda} = \bar{x}^2\bar{y}$ (c) $\hat{\lambda} = \sqrt{\bar{x} + \bar{y}^2}$ (d) $\hat{\lambda} = \sqrt{\bar{y} + \bar{x}^2}$ (e) $\hat{\lambda} = 2\{\bar{x} + 2\bar{y}\}^{-1}$.

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

4 Étant donné $X_1, \ldots, X_n$ un échantillon iid de loi $\mathcal{U}(a, b)$, l'estimateur du maximum de vraisemblance de $\theta = \{a + b\}/2$ est donné par, si $x_{(n/2)}$ dénote la médiane de l'échantillon
(a) $\hat{\theta} = \min(\bar{x}_n - a, b - \bar{x}_n)$ (b) $\hat{\theta} = \{\min(x_i) + \max(x_i)\}/2$ (c) $\hat{\theta} = \bar{x}_n$
(d) $\hat{\theta} = x_{(n/2)}$ (e) $\hat{\theta} = \max(x_i)$ (f) $\hat{\theta} = 0$.

Exercice 2 (7 pts)

Dans cet exercice, on considère la simulation de variables normales.

1 Si X_1 et X_2 sont deux variables $\text{N}(0, 1)$ indépendantes, montrez que

$$X_1^2 + X_2^2 \sim \mathcal{E}(1/2) \quad \text{et} \quad \theta = \arctan(X_1/X_2) \sim \mathcal{U}(-\pi/2, \pi/2)$$

2 Montrer ou déduire de la question 1. que générer $U_1, U_2 \overset{\text{i.i.d.}}{\sim} \mathcal{U}(0,1)$ et construire

$$X_1 = \sqrt{-2\log(U_1)}\ \cos(2\pi U_2)\ , \qquad X_2 = \sqrt{-2\log(U_1)}\ \sin(2\pi U_2)\ ,$$

produit deux variables $\mathrm{N}(0,1)$ indépendantes.

3 Soit (X_1, X_2) un vecteur normal $\mathrm{N}_2\left((0,0), \Sigma\right)$ avec

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Montrer que $\mathrm{cov}(X_1 - X_2, X_1 + X_2) = 0$ et en déduire que les variables $X_1 - X_2$ et $X_1 + X_2$ sont normales et indépendantes. On précisera les paramètres des deux lois.

4 Déduire des questions 2. et 3. une méthode de simulation du vecteur $(X_1, X_2) \sim \mathrm{N}_2\left((0,0), \Sigma\right)$ reposant sur deux variables $U_1, U_2 \overset{\text{i.i.d.}}{\sim} \mathcal{U}(0,1)$.

5 Montrer que, si $(X_1, X_2) \sim \mathrm{N}_2\left((0,0), \Sigma\right)$, $X_1 X_2$ a la même distribution que

$$1/2\left\{(1+\rho)Z_1^2 - (1-\rho)Z_2^2\right\}\ ,$$

où $Z_1, Z_2 \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0,1)$.

6 En déduire que simuler la loi de XY revient à générer $U \sim \mathcal{U}(0,1)$, $R \sim \mathcal{E}(1)$ et à prendre

$$XY = R(\cos(4\pi U) + \rho)\,.$$

Exercice 3 (5 pts)

Soient $X_1, \ldots, X_n$ des variables aléatoires indépendentes et identiquement distribuées de loi Gamma $\Gamma(\alpha, \beta)$.

Dans tout cet exercice, on supposera α connu.

1 Montrer que la loi de $(X_1, \ldots, X_n)$ appartient à une famille exponentielle régulière. En déduire que $S = \sum_{i=1}^n X_i$ est une statistique exhaustive.

2 Calculer l'information de Fisher apportée par $X_1, \ldots, X_n$ sur β et en justifier l'existence.

3 Calculer $\hat{\beta}$ l'estimateur du maximum de vraisemblance.

4 Calculer le biais et la variance de $\hat{\beta}$.

5 Construire à partir de $\hat{\beta}$ un estimateur β^* sans biais de β. L'estimateur β^* est-il efficace ?

Exercice 5 (4 pts)

Etant donné $X_1, \ldots, X_n$ un échantillon iid de loi de Pareto $\mathcal{P}(\alpha, \mu)$,

1 On suppose μ connu et on choisit une loi Gamma a priori sur α donnée par

$$\alpha \sim \Gamma(a, b).$$

Déterminer la loi a posteriori sur α en fonction de a et b.

2 Dans le cadre de la question précédente, calculer l'espérance a posteriori et montrer que c'est un estimateur convergent presque sûrement vers α.

3 On suppose que μ est inconnu et α est connu et on choisit une loi a priori sur μ uniforme sur $[0, 1]$

$$\mu \sim \mathcal{U}(0, M)$$

Déterminer la loi a posteriori sur μ en fonction de M. Etudier la limite de la loi a posteriori lorsque M tend vers l'infini.

French – English Lexicon

bayésien(ne) – Bayesian
échantillon – sample
famille exponentielle – exponential family
loi *a posteriori* – posterior distribution
loi *a priori* – prior distribution
statistique libre – ancilary statistics
statistique exhaustive – sufficient statistic
vraisemblance – likelihood