Université Paris-Dauphine Année 2016–2017
L3 - Statistique (1) : Estimation

Examen Final 05/01/2017 - Durée 2h00 - Documents non autorisés

Important : Suivant les réglements en vigueur,

1. **les enseignants présents lors de l'épreuve ne peuvent communiquer que sur les fautes d'énoncé potentielles. Toute autre question durant la composition ne sera pas acceptée.**
2. **les étudiants sont tenus de se lever au moment de l'annonce de fin de la composition. En cas de refus, le responsable de l'UE sera fondé à ne pas prendre en compte la copie incriminée.**
3. **l'identification des copies et intercalaires doit se faire au moment de la remise de chaque copie par les enseignants et surveillants. Il ne sera pas accordé de délai pour cette raison en fin d'épreuve.**

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 gamma $\Gamma(a, b)$ a pour densité $f_{a,b}(x) = \mathbf{1}_{x>0} x^{a-1} e^{-bx} b^a / \Gamma(a)$
- la loi géométrique $\mathcal{G}(P)$ a pour densité sur $\mathbb{N}^*$ $\mathbb{P}(X = k) = p\,(1-p)^{k-1}$

Exercice 1 (6 pts)

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

1 Soit une loi normale $\mathcal{N}(\mu, 1/\nu)$ avec μ fixé. On prend la loi a priori $\nu \sim \Gamma(\alpha, \beta)$. La loi a posteriori de ν sachant une réalisation x de la loi normale est
(a) $\Gamma(\alpha+1, \beta+x)$ (b) $\Gamma(\alpha+x, \beta+1)$ (c) $\Gamma(\alpha+1/2, \beta+(\mu-x)^2/2)$ (d) $\Gamma(\alpha+(\mu-x)^2/2, \beta+1)$

2 Soit $X_1, \ldots, X_n$ un échantillon ordonné, $X_1 < X_2 < \cdots < X_n$. Soit $\theta(X_1, \ldots, X_n) = \min(X_1, \ldots, X_n)$ l'estimateur auquel on s'intéresse. Sa distribution bootstrap est
(a) $\mathbb{P}(\theta = X_i) = 1/n$ (b) $\mathbb{P}(\theta = X_i) = \binom{n}{i}(1-1/n)^i(1/n)^{n-i}$ (c) $\mathbb{P}(\theta = X_i) = \left(1 - \frac{i-1}{n}\right)^n - \left(1 - \frac{i}{n}\right)^n$ (d) $\mathbb{P}(\theta = X_i) = \binom{i+n-1}{i}(1/n)^i(1-1/n)^n$

3 Une loi a priori conjuguée pour la loi binomiale négative $\mathbb{P}(X = x|m, \pi) = \binom{x+m-1}{x}\pi^x(1-\pi)^m$ à m fixé est la
(a) gamma (b) beta (c) de Poisson (d) normale

4 Si $X_1, \ldots, X_n$ est un échantillon iid de loi de densité $a^{-1}xe^{-x^2/2a}\mathbb{I}_{x>0}$ ($a > 0$), l'estimateur du maximum de vraisemblance de a associé à l'échantillon $(x_1, \ldots, x_n)$ est donné par
(a) $1/n\sum_{i=1}^n x_i$ (b) $1/n\sum_{i=1}^n x_i^2$ (c) $1/2n\sum_{i=1}^n x_i^2$ (d) $1/2n\sum_{i=1}^n |x_i|$

5 Si $X_1, \ldots, X_n$ est un échantillon iid de loi géométrique $\mathcal{G}(p)$, l'estimateur du maximum de vraisemblance de p associé à l'échantillon $(x_1, \ldots, x_n)$ est donné par
(a) $1/n\sum_{i=1}^n x_i$ (b) $n/\sum_{i=1}^n x_i$ (c) $1/n\sum_{i=1}^n \ln(x_i)$ (d) $1/n\sum_{i=1}^n(1-x_i)$

6 Si $X_1, \ldots, X_n$ est un échantillon iid de loi uniforme sur $[0, \theta]$, l'estimateur du maximum de vraisemblance de θ associé à l'échantillon $(x_1, \ldots, x_n)$ est donné par
(a) $\max\limits_{i=1,..,n} x_i$ (b) $\min\limits_{i=1,..,n} x_i$ (c) $1/2(\max\limits_{i=1,..,n} x_i + \min\limits_{i=1,..,n} x_i)$ (d) $1/2\max\limits_{i=1,..,n} x_i$

Exercice 2 (9 pts)

Soit un échantillon aléatoire ($n \geq 2$)

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

1 **[1.5 pt]** Représenter ce modèle comme une famille exponentielle : donner la paramétrisation naturelle, la représentation minimale, et indiquer si la famille est régulière.

2 **[1.5 pt]** Donner une statistique exhaustive minimale T associée à l'échantillon et sa densité.

3 **[1 pt]** On suppose μ connu. Montrer que l'estimateur de la variance est

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

est sans biais et a une variance égale à $2\sigma^2/n$. En déduire la consistance de cet estimateur.

4 **[1.5 pt]** On considère à présent μ inconnu. Montrer que l'estimateur du maximum de vraisemblance de la variance est

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

est à présent biaisé. Montrer que la correction

$$\tilde{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

est sans biais.

5 **[1 pt]** Montrer que la variance de l'estimateur $\hat{\sigma}_n^2$ est inférieure à la variance de l'estimateur $\tilde{\sigma}_n^2$. (*Indication :* On montrera que $n\hat{\sigma}_n^2$ suit une loi Gamma $\mathrm{Ga}({}^{n-1}/_2, {}^1/_{2\sigma^2})$.)

6 **[1 pt]** En déduire que l'erreur quadratique de l'estimateur du maximum de vraisemblance de la variance est toujours inférieure à l'erreur quadratique de l'estimateur sans biais de la variance :

$$\mathbb{E}[(\hat{\sigma}_n^2 - \sigma^2)^2] \leq \mathbb{E}[(\tilde{\sigma}_n^2 - \sigma^2)^2]$$

7 **[1.5 pt]** Montrer que l'estimateur

$$\bar{\sigma}_n^2 = \frac{1}{n+1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

a une erreur quadratique minimale parmi les estimateurs de la forme $\alpha\hat{\sigma}_n^2$.

Exercice 3 (6 pts)

Soit un problème numérique où la quantité d'intérêt est la probabilité, p, qu'une variable aléatoire de loi de Cauchy $\mathcal{C}(0,1)$ soit plus grande que 2,

$$p = \int_2^{+\infty} \frac{1}{\pi(1+x^2)}\,\mathrm{d}x\ .$$

On cherche à évaluer p par simulation.

1 **[1 pt]** Prenant un échantillon iid $X_1, \ldots, X_m \sim \mathcal{C}(0,1)$ et la moyenne empirique

$$\hat{p}_1 = \frac{1}{m}\sum_{j=1}^{m}\mathbb{I}_{X_j>2}$$

montrer que sa variance est égale à $p(1-p)/m$ (soit $0.125/m$ puisque $p \approx 0.15$).

2 **[.5 pt]** Montrer que l'on peut réduire cette variance en considérant que la loi $\mathcal{C}(0,1)$ est symétrique, en établissant que la moyenne

$$\hat{p}_2 = \frac{1}{2m}\sum_{j=1}^{m}\mathbb{I}_{|X_j|>2}$$

est sans biais et a pour variance $p(1-2p)/2m$ (soit $0.052/m$).

3 **[1 pt]** Montrer que p peut s'écrire

$$p = \frac{1}{2} - \int_0^2 \frac{1}{\pi(1+x^2)}\,\mathrm{d}x\ ,$$

et en déduire que, si $U_1, \ldots, U_m \sim \mathcal{U}_{[0,2]}$, un autre estimateur sans biais de p est

$$\hat{p}_3 = \frac{1}{2} - \frac{1}{m}\sum_{j=1}^{m} h(U_j)$$

où $h(x) = 2/\pi(1+x^2)$.

4 **[1.5 pt]** Montrer par une integration par partie que

$$\int_0^2 \frac{1}{(1+u^2)^2}\mathrm{d}u = {}^1\!/_2 \int_0^2 \frac{1}{(1+u^2)^2}\mathrm{d}u + {}^1\!/_5$$

et en déduire que la variance de $\hat{p}_3$ vaut ${}^{1-2p}\!/_{2\pi} + {}^2\!/_{5\pi^2} - {}^{(1-2p)^2}\!/_4$ (soit approximativement $0.0285/m$).

4 **[1 pt]** Montrer que p peut encore s'écrire

$$p = \int_0^{1/2} \frac{y^{-2}}{\pi(1+y^{-2})}\,dy\ .$$

En déduire que l'intégrale est l'espérance de ${}^1\!/_4\, h(Y) = 1/2\pi(1+Y^2)$ quand Y est distribuée suivant la loi uniforme sur $[0, {}^1\!/_2]$. Montrer que l'estimateur de p

$$\hat{p}_4 = \frac{1}{4m}\sum_{j=1}^{m} h(Y_j)$$

où $Y_1, \ldots, Y_m \sim \mathcal{U}_{[0,1/2]}$ est sans biais et de variance ${}^p\!/_{4\pi} + {}^1\!/_{10\pi^2} - p^2$ (soit approximativement $0.95\ 10^{-4}/m$).

5 **[1 pt]** On rappelle qu'une variable Z de loi de Cauchy $\mathcal{C}(0,1)$ peut s'écrire comme le rapport de deux variables normales et indépendantes, $X, Y \sim \mathcal{N}(0,1)$. Montrer que l'utilisation de la représentation $Z = X/Y$ dans la simulation ne conduit pas à une amélioration par rapport à $\hat{p}_1$.

French – English Lexicon

échantillon – sample
famille exponentielle – exponential family
loi *a posteriori* – posterior distribution
loi *a priori* – prior distribution
statistique exhaustive – sufficient statistic
vraisemblance – likelihood