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

Examen Partiel Octobre 2016 - Durée 2h00 - Sans document

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.**

Exercice 1 (9 pts)

Étant donné un échantillon i.i.d., $(X_1, \dots, X_n)$, de variables aléatoires de fonction de répartition F, avec $n \geq 2$, on note

$$X_{(k)} = \min_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} \max(X_{i_1}, \dots, X_{i_k}).$$

1 (1 pt) Montrer que $X_{(1)} = \min\{X_1, \dots, X_n\}$, $X_{(n)} = \max\{X_1, \dots, X_n\}$, et que, si $n = 2m+1$, $X_{(m)} = \text{median}\{X_1, \dots, X_n\}$.

2 (1.5 pts) Si $\hat{F}_n$ dénote la fonction de répartition empirique associée à l'échantillon i.i.d. $(X_1, \dots, X_n)$, montrer que $n\hat{F}_n(t)$ suit une loi binomiale $\mathcal{B}(n, F(t))$ pour une valeur donnée de $t \in \mathbb{R}$. En déduire que la fonction de répartition de $X_{(k)}$ est donnée par

$$\mathbb{P}(X_{(k)} \leq t) = \sum_{i=k}^{n} \binom{n}{i} F(t)^i (1 - F(t))^{n-i}$$

3 (3 pts) Si F admet une densité f, en déduire que la densité de $X_{(k)}$ est

$$f_{(k)}(x) = f(x)\, k \binom{n}{k} F(x)^{k-1} \left[1 - F(x)\right]^{n-k}.$$

Construire la densité de la loi conditionnelle de $(X_1, \dots, X_n)$ sachant $(X_{(1)}, \dots, X_{(n)})$ et la densité de la loi jointe de $(X_{(1)}, \dots, X_{(n)})$.

4 (2 pts) Si F est la fonction de répartition d'une loi exponentielle $\mathcal{E}(\lambda)$, avec $\lambda > 0$, donner la densité de $X_{(1)}$ et celle de $X_{(n)}$. Montrer que la loi de $X_{(1)}$ est la loi exponentielle $\mathcal{E}(n\lambda)$. Montrer que la loi de $X_{(n)} - X_{(1)}$ est la loi du maximum de $(n-1)$ variables i.i.d. $\mathcal{E}(\lambda)$.

5 (1 pt) Si $n > 2$ et F est la fonction de répartition d'une loi de Cauchy, de densité

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2} \qquad x \in \mathbb{R},$$

donner la densité de $X_{(2)}$ et en déduire que $X_{(2)}$ admet un moment d'ordre 1. *On rappelle que la fonction de répartition d'une loi de Cauchy est* ${}^{\operatorname{arctg}(x)}/_{\pi} + {}^{1}/_{2}$ *est que son équivalent en* $-\infty$ *est* $-{}^{1}/_{x}$.

Exercice 2 (6 pts)

On rappelle que pour $p > 0$, la densité d'une variable aléatoire suivant une loi Gamma de paramètre (p, θ), notée $Ga\,(p, \theta)$, est :

$$t \mapsto \frac{\theta^p}{\Gamma\,(p)} e^{-\theta t} t^{p-1} \mathbf{1}_{t>0}$$

avec

$$\Gamma\,(p) = \int_0^{+\infty} e^{-t} t^{p-1}\, \mathrm{d}t\,.$$

1 (1 pt) Calculer la fonction caractéristique d'une variable aléatoire $Z \sim Ga\,(p, \theta)$. En déduire :
(i) $\mathbb{E}\,(Z)$ et $\mathbb{V}\,(Z)$.
(ii) Si $W \sim Ga\,(q, \theta)$, $q > 0$, avec W indépendante de Z alors $Z + W \sim Ga\,(p + q, \theta)$.

2 (1 pt) Soit X_1 suivant une loi normale $\mathcal{N}\,(0, 1)$. Calculer la densité de la variable aléatoire X_1^2 et en déduire qu'elle suit une loi gamma dont on précisera les paramètres.

3 (1pt) Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires identiquement et indépendamment distribuées de loi $\mathcal{N}\,(0, 1)$. Donner la densité de $Y_n = \sum_{i=1}^n X_i^2$. On pourra utiliser le résultat de la question 1 (ii).

4 (1pt) Montrer que la loi de Y_n appartient à une famille exponentielle (le paramètre est n). Préciser le paramètre naturel et l'espace de cette dernière.

5 (1pt) Déduire des propriétés des familles exponentielles les expressions de $\mathbb{E}\,(Y_n)$ et $\mathbb{V}\,(Y_n)$. Comparer avec les résultats de la question 1 (i).

6 (1pt) Montrer que ${}^{Y_n - n}/_{\sqrt{2n}} \xrightarrow{d} \mathrm{N}(0, 1)$.

Exercice 3 (6 pts)

Soient $x_1, \ldots, x_n$, n réalisations d'une variable aléatoire X de loi $\mathbb{P}_X$ ayant pour densité $f(x, y) = {}^1/_4 \mathbf{1}_{((x,y)\in[-1,1]\times[-1,1])}$. Il s'agit de la loi uniforme sur le carré $[-1, 1] \times [-1, 1]$. On note

$$\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$$

la distribution empirique avec $\delta_x(A) = 1$ si $x \in A$ et 0 sinon. On note $\mathcal{C}_r$ le cercle de centre 0 et de rayon r plus petit que 1. Dans toute la suite $\mathbb{P}_n(\mathcal{C}_r)$ sera notée $p_n(r)$ et $\mathbb{P}_X(\mathcal{C}_r)$ sera notée $p(r)$.

1 (1 pt)
(i) Calculer $p(r)$.
(ii) Trouver la loi de $np_n(r)$.

2 (1 pt) Montrer que $p_n(r)$ tend presque sûrement vers une constante que l'on déterminera.

3 (1 pt) Montrer que $\sqrt{n}\,(p_n(1) - p(1))$ converge en loi et expliciter la limite.

4 (1 pt) Montrer que $2\sqrt{n}\left(\sin(p_n(1)) - \frac{1}{\sqrt{2}}\right)$ converge en loi et expliciter la limite.

5 (2 pts) Montrer que $Y_n = np_n\,({}^1\!/\!\sqrt{n})$ converge en loi vers Y une variable aléatoire de Poisson de paramètre $\pi/4$.

Exercice 4 (4 pts)

[1]

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

1 Pour simuler une variable aléatoire X suivant une loi exponentielle $\mathcal{E}(\lambda)$, avec $\lambda > 0$, de densité $f(x) = \lambda \exp\{-\lambda x\}$ sur $\mathbb{R}^+$, à partir d'une seule simulation uniforme $\mathcal{U}(0,1)$, U, il faut utiliser la transformation
(a) $X = \lambda \log(U)$; (b) $X = -\log[(1-U)^\lambda]$; (c) $X = -\log(1-U)/\lambda$; $X = \log[\lambda(1-U)]$; (d) $X = -\log(U/\lambda)$; (e) $X = \log(U)^\lambda$.

2 Le nombre moyen $\mathbb{E}[N]$ de variables uniformes iid $U_1, U_2, \ldots \sim \mathcal{U}(0,1)$ qu'il faut simuler pour que leur somme cumulée $U_1 + \cdots + U_N$ dépasse 1 (un) est
(a) 2 ; (b) π ; (c) 3 ; (d) π^2 ; (e) e ; (f) $1+\sqrt{5}/2$; (g) $4/3$; (h) $\sqrt{5}-1/2$; (i) $5/3$.

3 Si X_1 est une simulation suivant une loi de fdr F_1 et X_2 une simulation de fdr F_2, simuler suivant la loi de fdr $F_1 \times F_2$ s'obtient en prenant
(a) $Y = X_1 \times X_2$; (b) $Y = X_1 + X_2$; (c) $Y = \min\{X_1, X_2\}$; (d) $Y = \max\{X_1, X_2\}$

4 Pour simuler une loi de Pareto de densité $f(x) = n\theta^n x^{-n-1}$ sur $(\theta, +\infty)$, à partir d'une simulation uniforme $\mathcal{U}(0,1)$, U, il suffit d'utiliser la transformation
(a) $Y = \theta U^n$; (b) $Y = \theta U^{-n}$; (c) $Y = 1/\theta U^n$; (d) $Y = \theta U^{-1/n}$; (e) $Y = 1/[\theta U]^{1/n}$

French – English Lexicon

échantillon – sample
famille exponentielle – exponential family
fonction de répartition (fdr) – cumulative distribution function (cdf)

1. Solutions : c – e – d – d.