Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2017 – 2018

Statistiques I

Contrôle continu n°2, décembre 2017

Examen de 1h30. Tout document ou calculatrice est interdit.

1. **(Sur 12 points)** Soit la variable X qui suit une loi dont la densité f_X par rapport à la mesure de Lebesgue sur $]0,1]$ est, avec β et $K \in \mathbb{R}$:
$$f_X(x) = K\, x^{\beta} \quad \text{pour tout } x \in]0,1],$$

   (a) Déterminer K en fonction de β **(0.5pts)**, en précisant quelle condition doit vérifier β **(0.5pts)**. En déduire $\mathbb{E}(X)$ **(0.5pts)** et $\mathrm{var}(X)$ **(1pt)**, en précisant également les conditions portant sur β.

   (b) On suppose que la suite $(X_i)_{i\in\mathbb{N}}$ est constituée de variables aléatoires indépendantes et identiquement distribuées suivant la même loi que X. Soit un échantillon observé $(X_1,\ldots,X_n)$. On désire estimer β à partir de cet échantillon. Quel est le modèle statistique **(0.5pts)**? Montrer que ce modèle appartient à la famille exponentielle **(0.5pts)**.

   (c) En déduire qu'il n'existe pas d'estimateur sans biais efficace de β **(1.5pts)**.

   (d) Montrer que $-\log(X)$ suit une loi connue dont on précisera le paramètre **(1.5pts)**. En déduire que $\widetilde{\beta}_n = -1-(n-1)\Big(\sum_{i=1}^n \log(X_i)\Big)^{-1}$ est un estimateur sans biais de β (utiliser la loi gamma...) pour $n \geq 2$ **(1.5pts)**, puis qu'il est de variance uniformément minimale parmi les estimateurs sans biais **(1pt)**.

   (e) Montrer que l'estimateur $\widehat{\beta}_n$ de β par maximum de vraisemblance diffère de $\widetilde{\beta}_n$ **(1pt)**. Montrer que $\widehat{\beta}_n$ est biaisé mais asymptotiquement non biaisé **(1pt)**. Montrer sans calcul que $\mathrm{var}(\widetilde{\beta}_n) < \mathrm{var}(\widehat{\beta}_n)$ **(1pts)**.

Proof. (a) On a pour $\beta \in]-1,\infty[$, $\int_0^1 K\,x^{\beta}\,dx = \frac{K}{\beta+1}$, donc $K = \beta+1$.
On en déduit que $\mathbb{E}(X) = (\beta+1)\int_0^1 x^{\beta+1}\,dx = \frac{1+\beta}{2+\beta}$, toujours pour $\beta > -1$.
De même $\mathrm{var}(X) = (\beta+1)\int_0^1 x^{\beta+2}\,dx - \Big(\frac{1+\beta}{2+\beta}\Big)^2 = \frac{1+\beta}{3+\beta} - \Big(\frac{1+\beta}{2+\beta}\Big)^2 = \frac{1+\beta}{(2+\beta)^2(3+\beta)}$, toujours pour $\beta > -1$.

(b) Le modèle statistique est $(]0,1]^n, \mathcal{B}(]0,1]^n), (\int f(x)dx)^{\otimes n})$, dominé par $\lambda_{]0,1]^n}$.
On peut écrire que la vraisemblance est $L_\beta(x_1,\ldots,x_n) = \exp\big(n\log(1+\beta)+\beta\sum_{i=1}^n\log(x_i)\big)$: le modèle est bien exponentiel.

(c) Le modèle est régulier et exponentiel, avec $\beta(\beta) = n\log(1+\beta)$, $\alpha_1(\beta) = \beta$ et $a_1(x_1,\ldots,x_n) = \sum_{i=1}^n\log(x_i)$. On en déduit que la fonction (à une transformation affine près) de β que l'on peut estimer sans biais et efficacement est: $g(\beta) = \frac{n}{1+\beta}$.

(d) Comme X prend ses valeurs dans $]0,1]$, alors $-\log(X)$ prend ses valeurs dans $[0,\infty[$. Pour $x \geq 0$, on a $\mathbb{P}\big(-\log(X) \leq x\big) = \mathbb{P}\big(X \geq e^{-x}\big) = 1-(1+\beta)\int_0^{e^{-x}} t^{\beta}dt = 1-e^{-(1+\beta)x}$. Donc la densité de $\log(X)$ est $(1+\beta)e^{-(1+\beta)x}\mathbb{I}_{x\geq 0}$: $-\log(X)$ suit une loi $\mathcal{E}(1+\beta)$.
La loi de $-\sum_{i=1}^n\log(X_i)$ est donc une loi $\Gamma(n,(1+\beta))$. Donc $\mathbb{E}(-1/\sum_{i=1}^n\log(X_i)) = \frac{(1+\beta)^n}{\Gamma(n)}\int_0^\infty x^{-1}x^{n-1}e^{-(1+\beta)x}dx = \frac{(1\beta)}{\Gamma(n)}\int_0^\infty y^{n-2}e^{-y}dy = \frac{(1\beta)}{\Gamma(n)}\Gamma(n-1) = \frac{(1\beta)}{(n-1)}$. En conséquence, $\mathbb{E}(\widetilde{\beta}_n) = -1+(1+\beta) = \beta$: l'estimateur est sans biais.
On a un modèle exponentiel et l'intérieur de $\alpha(\Theta)$ est $]-1,\infty[$ donc non vide. On en déduit que la statistique $\sum_{i=1}^n\log(X_i)$ est exhaustive complète pour le modèle. D'après le Théorème de Lehmann-Scheffé, on sait qu'il existe une unique fonction de cette statistique sans biais de variance unformément minimale: c'est donc $\widetilde{\beta}_n$.

(e) On montre en dérivant la log-vraisemblance par β que $\widehat{\beta}_n = -1-n\Big(\sum_{i=1}^n\log(X_i)\Big)^{-1}$, sa dérivée seconde étant $-n(1+\beta)^{-2}$ donc négative.
A partir du calcul du biais de $\widetilde{\beta}_n$, on a $\mathbb{E}(\widehat{\beta}_n) = -1+\frac{n}{n-1}(1+\beta) = \beta+\frac{1+\beta}{n-1}$, d'où $\widehat{\beta}_n$ estimateur biaisé mais asymptotiquement non biaisé.
On a $\widetilde{\beta}_n = \frac{1}{n}+\frac{n-1}{n}\widehat{\beta}_n$, d'où le résultat.

□

2. **(Sur 20 points)** Soit une suite $(X_i)_{n\geq 1}$ de variables aléatoires indépendantes et identiquement distribuées suivant une loi uniforme sur $\{-m, 1-m, \dots, m-1, m\}$ où $m \in \mathbf{N}$ est inconnu. On observe $(X_1, \dots, X_n)$ où $n \in \mathbf{N}^*$.

   (a) Rappeler la loi de X_1 ainsi que son espérance **(1pt)**.

   (b) Déterminer le modèle statistique induit par $(X_1, \dots, X_n)$ **(0.5pts)**. Est-ce un modèle dominé **(0.5pts)**? Exponentiel **(0.5pts)**?

   (c) Montrer que $\widehat{T} = (\min_{1\leq i\leq n} X_i\,,\, \max_{1\leq i\leq n} X_i)$ est une statistique exhaustive pour le modèle **(1pt)**.

   (d) Montrer que $\widehat{T}' = \max_{1\leq i\leq n} |X_i|$ est aussi une statistique exhaustive **(1pt)**. Montrer qu'elle est minimale **(1.5pts)** alors que T ne l'est pas **(1pt)**.

   (e) Déterminer la loi de $|X_1|$ **(0.5pts)**, et en déduire que $\mathbb{P}(\widehat{T}' = j) = \frac{(2j+1)^n - (2j-1)^n}{(2m+1)^n}$ pour $j \in \{1, 2, \leq, m\}$ **(2.5pts)**.

   (f) Soit $h : \mathbf{N} \to \mathbf{R}$, quelconque. Déterminer explicitement $\mathbb{E}[h(\widehat{T}')]$ **(0.5pts)**. En déduire que si $\mathbb{E}[h(\widehat{T}')] = 0$ pour tout $m \in \mathbf{N}$, alors $h \equiv 0$ (on pourra commencer par le cas $m = 0$,...) **(2pts)**. Conséquence **(0.5pts)**?

   (g) Déterminer $\widehat{m}_n$ l'estimateur du maximum de vraisemblance de m **(1pt)**, puis montrer que $\mathbb{E}(\widehat{m}_n) = m - b_n(m)$ avec $b_n(m) = \frac{1}{(2m+1)^n} \sum_{i=1}^{m-1} (2i+1)^n$ **(2pts)**.

   (h) A l'aide d'une comparaison avec une intégrale, montrer que $b_n(m) \leq (2m+1)/(n+1)$ **(2pts)**. En déduire que $\mathbb{E}[|\widehat{m}_n - m|] \underset{n\to+\infty}{\longrightarrow} 0$ **(1pt)**, puis que $\widehat{m}_n$ est un estimateur convergent de m **(0.5pts)**. A quelle vitesse converge-t-il **(0.5pts)**?

Proof. (a) On a une loi discrète à valeurs dans $\{-m, 1-m, \dots, m-1, m\}$ et pour tout $j \in \{-m, 1-m, \dots, m-1, m\}$, on a $\mathbb{P}(X_1 = j) = \frac{1}{(2m+1)}$.
On a facilement $\mathbb{E}(X_1) = 0$

(b) Le modèle statistique est $\left(\mathbf{Z}^n, \mathcal{P}(\mathbf{Z}^n), \mathbb{P}_m^{\otimes n}\right)_{m\in\mathbf{N}}$.
Le modèle est dominé par la mesure $\left(\sum_{i\in\mathbf{Z}} \delta_i\right)^{\otimes n}$), mesure de comptage sur $\mathbf{Z}^n$.
La vraisemblance du modèle est $L_m(x_1, \dots, x_n) = \frac{1}{(2m+1)^n} \mathbb{I}_{\min_i(X_i)\geq -m \cap \max_i(X_i)\leq m}$. Ce n'est pas un modèle exponentiel car cette vraisemblance peut s'annuler.

(c) D'après la forme de la vraisemblance, on a $L_m(x_1, \dots, x_n) = g_m(\widehat{T})$, donc d'après la factorisation de Neyman, on en déduit que $\widehat{T}$ est une statistique exhaustive.

(d) On peut écrire que
$\big\{\min_i(X_i) \geq -m \cap \max_i(X_i) \leq m\big\} \iff \big\{\max\big(-\min_i(X_i)\,,\, \max_i(X_i)\big) \leq m\big\} \iff \big\{\max\big(\max_i(-X_i)\,,\, \max_i(X_i)\big) \leq m\big\} \iff \big\{\max_i(|X_i|) \leq m\big\}$. Donc $L_m(x_1, \dots, x_n) = \frac{1}{(2m+1)^n} \mathbb{I}_{\max_i(|X_i|)\leq m}$: la statistique $\widehat{T}'$ est donc aussi exhaustive pour le modèle.
On a $\frac{L_m(x_1,\dots,x_n)}{L_m(y_1,\dots,y_n)} = \frac{\mathbb{I}_{\max_i(|x_i|)\leq m}}{\mathbb{I}_{\max_i(|y_i|)\leq m}}$ et pour que ce rapport ne dépende pas de m, il est nécessaire que $\max_i(|x_i|) = \max_i(|y_i|)$, donc $\widehat{T}'$ est bien minimale.
Si $\widehat{T}$ était minimale, alors on aurait $\widehat{T}$ qui pourrait s'écrire comme une fonction de $\widehat{T}'$. Mais $\min_i(X_i)$ et $\max_i(X_i)$ ne peuvent se déduire de $\max_i(|X_i|)$ dès que $n \geq 2$. Par exemple, si $(X_1, X_2) = (3, -1)$ alors $\widehat{T}' = 3$ alors que $\widehat{T} = (-1, 3)$.

(e) La loi de $|X_1|$ est discrète à valeurs dans $\{0, 1, \dots, m\}$, et $\mathbb{P}(|X_1| = j) = \frac{2}{2m+1}$ pour $j \in \{1, 2, \dots, m\}$ et $\mathbb{P}(|X_1| = 0) = \frac{1}{2m+1}$.
On a pour $j \in \{0, 1, \dots, m\}$, $\mathbb{P}(\max_i(|X_i|) \leq j) = \mathbb{P}(|X_1| \leq j)^n$ puisque on considère des vaiid. Mais $\mathbb{P}(|X_1| \leq j) = \frac{2j+1}{2m+1}$.
Donc $\mathbb{P}(\max_i(|X_i|) \leq j) = \left(\frac{2j+1}{2m+1}\right)^n$. Enfin $\mathbb{P}(|X_1| = j) = \mathbb{P}(|X_1| \leq j) - \mathbb{P}(|X_1| \leq j-1)$, d'où le résultat.

(f) On a $\mathbb{E}(h(\widehat{T}')) = \sum_{j=0}^m h(j)\,\mathbb{P}(|X_1| = j)$. Si on a $\mathbb{E}(h(\widehat{T}')) = 0$ pour tout $m \in \mathbf{N}$, cela est vrai pour $m = 0$, pour lequel $\mathbb{E}(h(\widehat{T}')) = h(0)\,\mathbb{P}(|X_1| = 0)$ et donc $h(0) = 0$. De même pour $m = 1$, donc $h(1) = 0$ puisque $h(0) = 0$. Par itération, on en déduit que $h(j) = 0$ pour tout $j \in \mathbf{N}$, donc $h = 0$: la statistique est complète.
En conséquence, on sait qu'un estimateur sans biais uniformément de variance minimale pourra être obtenu comme une fonction unique de $\widehat{T}'$ (Théorème de Lehmann-Scheffé).

(g) Comme $L_m(x_1, \dots, x_n) = \frac{1}{(2m+1)^n} \mathbb{I}_{\widehat{T}'\leq m}$, on en déduit que la vraisemblance sera maximale si elle est non nulle et si m est le plus petit possible. Cela est obtenu pour $m = \widehat{T}'$, donc $\widehat{m}_n = \max_i(|X_i|)$.
On a $\mathbb{E}(\widehat{m}_n) = \sum_{j=0}^m j\,\mathbb{P}(|X_1| = j) = (1+2m)^{-m} \sum_{j=1}^m j\big((2j+1)^n - (2j-1)^n\big)$. La série se simplifie par téléscopage et on a $\mathbb{E}(\widehat{m}_n) = (1+2m)^{-m}\Big(\sum_{j=1}^m j(2j+1)^n - \sum_{j=0}^{m-1}(j+1)(2j+1)^n\Big) = (1+2m)^{-m}\Big(m(2m+1)^n - \sum_{j=0}^{m-1}(2j+1)^n\Big) = m - \sum_{j=0}^{m-1}\left(\frac{2j+1}{2m+1}\right)^n$.

(h) Comme la fonction $x \mapsto x^n$ est croissante, on en déduit que pour tout $x \in [2j+1, 2j+3]$, alors $\left(\frac{2j+1}{2m+1}\right)^n \leq \left(\frac{x}{2m+1}\right)^n$. Donc $b_n \leq (2m+1)^{-n} \sum_{j=1}^{m-1} \int_j^{j+1} (2x+1)^n dx \leq (2m+1)^{-n} \int_0^m (2x+1)^n dx \leq \frac{2m+1}{n+1}$.
Par construction, comme $\widehat{m}_n \leq m$, on en déduit que $|\widehat{m}_n - m| = m - \widehat{m}_n$ donc $\mathbb{E}\big[|\widehat{m}_n - m|\big] = b_n \underset{n\to+\infty}{\longrightarrow} 0$.

D'après l'inégalité de Markov, on a tout de suite la convergence en probabilité de $\widehat{m}_n$ vers m.
La vitesse de convergence est inférieure à $\frac{2m+1}{n+1}$, ce qui est mieux qu'avec un modèle régulier...

$\square$