M1 – Statistique – 2020/2021

Contrôle continu
Durée : 2h
Correction

Tout matériel qui ne ressemblerait pas à un simple stylo ou à une feuille blanche (calculatrice, portable, note de cours...) est interdit. Si vous vous demandez si vous avez le droit d'avoir un objet sur la table, c'est qu'il est certainement interdit.
L'énoncé est composé d'un seul grand exercice avec des parties plus ou moins dépendantes. ***Si un résultat n'a pas pu être démontré****, il pourra toutefois être admis pour les questions suivantes. Si un résultat n'est pas rigoureusement justifié, la totalité des points ne sera pas donnée.*

Contexte

Dans ce contrôle, nous étudions la famille de densités suivante :

$$f_\theta(x) = C_\theta x^{\theta-1}\mathbb{1}_{]0;1]}(x).$$

Nous supposons avoir un n-échantillon $(X_1, \ldots, X_n)$ d'observations iid suivant la loi de densités $f_{\theta^\star}$.

1. Montrer, en justifiant, que l'espace des paramètres Θ est $\mathbb{R}_+^\star$.

Comme nous avons une fonction de référence (intégrale de Riemann), il faut et il suffit que $\theta - 1 > -1$ ce qui équivaut à $\theta > 0$ pour que l'intégrale soit finie.

Un certain nombre d'étudiants ont juste montré que $\mathbb{R}_+^\star \subset \Theta$ ou $\mathbb{R}_- \not\subset \Theta$.

2. Calculer la constante C_θ.

Si $\theta = 1$, nous avons :

$$\begin{aligned}\int_0^1 f_\theta(x)\mathrm{d}x = 1 \quad &\Leftrightarrow \quad \int_0^1 C_\theta x^{1-1}\mathrm{d}x = 1\\ &\Leftrightarrow \quad C_\theta \int_0^1 \mathrm{d}x = 1\\ &\Leftrightarrow \quad C_\theta = 1\end{aligned}$$

Sinon, nous avons :

$$\begin{aligned}\int_0^1 f_\theta(x)\mathrm{d}x = 1 \quad &\Leftrightarrow \quad \int_0^1 C_\theta x^{\theta-1}\mathrm{d}x = 1\\ &\Leftrightarrow \quad \left[\frac{1}{\theta}x^\theta\right]_0^1 = \frac{1}{C_\theta}\\ &\Leftrightarrow \quad C_\theta = \theta\end{aligned}$$

Estimation par la méthode des moments

Dans cette partie, nous cherchons à estimer le paramètre $\theta^\star$ par la méthode des moments.

3. Pour tout $k \in \mathbb{N}^\star$, calculer le moment d'ordre k de la variable X_1.

Nous avons pour tout $k \in \mathbb{N}^\star$:

$$\begin{aligned}
\mathbb{E}_{\theta^\star}\left[X_1^k\right] &= \int_0^1 x^k \theta^\star x^{\theta^\star - 1}\mathrm{d}x \\
&= \int_0^1 \theta^\star x^{k+\theta^\star-1}\mathrm{d}x \\
&= \left[\frac{\theta^\star}{k+\theta^\star} x^{k+\theta^\star}\right]_0^1 \text{ car } \theta^\star + k > 1 \\
&= \frac{\theta^\star}{k+\theta^\star}.
\end{aligned}$$

4. En déduire que, pour tout $k \in \mathbb{N}^\star$, l'estimateur $\widehat{\theta}_{n,k}$ basé sur le moment d'ordre k est :

$$\widehat{\theta}_{n,k} = \frac{k\overline{X_n^k}}{1-\overline{X_n^k}} \text{ où } \overline{X_n^k} = \frac{1}{n}\sum_{i=1}^n X_i^k.$$

La méthode des moments propose de remplacer le paramètre d'intérêt par l'estimateur que nous estimions les moments par les estimateurs empiriques. Donc, nous avons :

$$\begin{aligned}
\overline{X_n^k} = \frac{\widehat{\theta}_{n,k}}{k+\widehat{\theta}_{n,k}} &\Leftrightarrow \overline{X_n^k}\left(k+\widehat{\theta}_{n,k}\right) = \widehat{\theta}_{n,k} \\
&\Leftrightarrow k\overline{X_n^k} = \widehat{\theta}_{n,k}\left(1-\overline{X_n^k}\right) \\
&\Leftrightarrow \widehat{\theta}_{n,k} = \frac{k\overline{X_n^k}}{1-\overline{X_n^k}}.
\end{aligned}$$

5. Pour tout $k \in \mathbb{N}^\star$, nous notons σ_k^2 la variance de X_1^k. Montrer que, lorsque k est grand, σ_k^2 est équivalent à $\theta^\star/(2k)$.

Pour cela, nous avons besoin de calculer la variance de X_1^k (version brute) :

$$\begin{aligned}
\sigma_k^2 &= \mathbb{V}_{\theta^\star}\left[X_1^k\right] \\
&= \mathbb{E}_{\theta^\star}\left[\left(X_1^k\right)^2\right] - \mathbb{E}_{\theta^\star}\left[X_1^k\right]^2 \\
&= \mathbb{E}_{\theta^\star}\left[X_1^{2k}\right] - \mathbb{E}_{\theta^\star}\left[X_1^k\right]^2 \\
&= \frac{\theta^\star}{2k+\theta^\star} - \left(\frac{\theta^\star}{k+\theta^\star}\right)^2 \\
&= \theta^\star\left[\frac{1}{2k+\theta^\star} - \frac{\theta^\star}{(k+\theta^\star)^2}\right] \\
&= \theta^\star \times \frac{k^2+2k\theta^\star+\theta^{\star 2}-2k\theta^\star-\theta^{\star 2}}{(2k+\theta^\star)(k+\theta^\star)^2} \\
&= \theta^\star \times \frac{k^2}{(2k+\theta^\star)(k+\theta^\star)^2} \\
&\underset{k\to+\infty}{\sim} \frac{\theta^\star k^2}{2k^3} \\
&\underset{k\to+\infty}{\sim} \frac{\theta^\star}{2k}
\end{aligned}$$

Une version en utilisant les petits o marchait tout autant.

6. Pour tout $k \in \mathbb{N}^\star$, calculer la loi asymptotique de $\overline{X_n^k}$.

D'après le théorème de la limite centrale appliquées à la suite de variables indépendantes et de même loi $(X_n^k)_{n\in\mathbb{N}^\star}$, nous avons que :

$$\sqrt{n}\left(\overline{X_n^k} - \mathbb{E}_{\theta^\star}\left[\overline{X_n^k}\right]\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0,\sigma_k^2\right).$$

7. Parmi les estimateurs $\widehat{\theta}_{n,k}$, recommanderiez-vous de prendre ceux avec un k grand ?

Nous avons :

- $\sqrt{n}\left(\overline{X_n^k} - \mathbb{E}_{\theta^\star}\left[\overline{X_n^k}\right]\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0,\sigma_k^2\right)$
- $\sqrt{n} \to +\infty$
- Nous définissons la fonction $g_k(x) = \frac{kx}{1-x}$ dérivable sur $\mathbb{R}\backslash\{1\}$ donc en $\frac{\theta^\star}{\theta^\star+1}$ car $k \geq 1$.

Donc toutes les conditions sont réunies pour appliquer la méthode Delta et nous savons que $\widehat{\theta}_{n,k} = g_k\left(\overline{X_n^k}\right)$ et $g_k\left(\mathbb{E}_{\theta^\star}\left[\overline{X_n^k}\right]\right) = \theta^\star$. De plus :

$$\begin{aligned}
g_k'(x) &= \frac{k}{1-x} - \frac{-kx}{(1-x)^2} \\
&= \frac{k-kx+kx}{(1-x)^2} \\
&= \frac{k}{(1-x)^2}
\end{aligned}$$

et donc :

$$\begin{aligned} g_k'\left(\frac{\theta^\star}{k+\theta^\star}\right) &= \frac{k}{\left(1-\frac{\theta^\star}{k+\theta^\star}\right)^2} \\ &= \frac{k}{\left(\frac{k+\theta^\star-\theta^\star}{k+\theta^\star}\right)^2} \\ &= \frac{k\,(k+\theta^\star)^2}{k^2} \\ &= \frac{(k+\theta^\star)^2}{k}. \end{aligned}$$

et nous avons :

$$\begin{aligned} \sqrt{n}\left(\widehat{\theta}_{n,k}-\theta^\star\right) &\xrightarrow[n\to+\infty]{\mathcal{L}} \frac{(k+\theta^\star)^2}{k}\mathcal{N}\left(0,\sigma_k^2\right) \\ &\xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0,\frac{(k+\theta^\star)^4}{k^2}\sigma_k^2\right) \end{aligned}$$

D'après les questions 5 et 6, la variance de la loi limite est

$$\begin{aligned} \frac{(k+\theta^\star)^4}{k^2}\sigma_k^2 &\underset{k\to+\infty}{\sim} \frac{k^4}{k^2}\frac{\theta^\star}{2k} \\ &\underset{k\to+\infty}{\sim} \frac{k\theta^\star}{2} \end{aligned}$$

donc, quand k est très grand, la variance explose. Il vaut mieux prendre un k pas trop grand.

Certains des étudiants n'ont pas fait la différence entre $\overline{X_n^k}$ de la question précédente et $\widehat{\theta}_{n,k}$.

Estimation par la méthode du maximum de vraisemblance

Dans cette partie, nous cherchons à estimer le paramètre $\theta^\star$ par la méthode du maximum de vraisemblance. Nous rappelons en fin de section les densités des lois gamma et inverse-gamma.

8. Calculer l'estimateur $\tilde{\theta}_n$ du maximum de vraisemblance de $\theta^\star$.

Nous avons :

$$\begin{aligned} \log V_{\mathbf{X}}(\theta) &= \sum_{i=1}^n \log\left(\theta X_i^{\theta-1}\right) \\ &= n\log\theta + (\theta-1)\sum_{i=1}^n \log X_i \end{aligned}$$

Ce qui donne :

$$\begin{aligned}\frac{\partial}{\mathrm{d}\theta}\log V_{\mathbf{X}}(\tilde{\theta}_n)=0 \quad &\Leftrightarrow \quad \frac{n}{\tilde{\theta}_n}+\sum_{i=1}^{n}\log X_i=0\\ &\Leftrightarrow \quad \tilde{\theta}_n=\frac{n}{\sum_{i=1}^{n}(-\log X_i)}\\ &\Leftrightarrow \quad \tilde{\theta}_n=\frac{1}{\overline{-\log \mathbf{X}}}\end{aligned}$$

9. Montrer que si X suit une loi de densité $f_{\theta^\star}$ alors $Y=-\log X$ suit une loi exponentielle $\mathcal{E}(\theta^\star)$.

Pour toute fonction h mesurable, nous avons :

$$\begin{aligned}\mathbb{E}\left[h(-\log X)\right] \quad &= \quad \int_0^1 h(-\log x)\theta^\star x^{\theta^\star-1}\mathrm{d}x\\ y=-\log x, \quad & \quad x=e^{-y} \text{ et } \frac{\mathrm{d}x}{\mathrm{d}y}=-e^{-y}\\ &= \quad \int_{+\infty}^{0} h(y)\theta^\star\left(e^{-y}\right)^{\theta^\star-1}\left(-e^{-y}\right)\mathrm{d}y\\ &= \quad \int_0^{+\infty} h(y)\theta^\star e^{-y\theta^\star+y-y}\mathrm{d}y\\ &= \quad \int_0^{+\infty} h(y)\theta^\star e^{-y\theta^\star}\mathrm{d}y\end{aligned}$$

Donc Y suit une loi exponentielle $\mathcal{E}(\theta^\star)$.

10. En admettant que la fonction caractéristique d'une loi exponentielle $\mathcal{E}(\lambda)$ de paramètre λ est définie pour tout $\xi\in\mathbb{R}$ par $\Phi_{\mathcal{E}(\lambda)}=\left(1-\frac{i\xi}{\lambda}\right)^{-1}$ et que la fonction caractéristique d'une loi gamma $\mathcal{G}amma(\alpha,\beta)$ est définie pour tout $\xi\in\mathbb{R}$ par $\Phi_{\mathcal{G}amma(\alpha,\beta)}=\left(1-\frac{i\xi}{\beta}\right)^{-\alpha}$, montrer que si $\mathbf{Y}=(Y_1,\dots,Y_n)$ est un n-échantillon de loi exponentielle $\mathcal{E}(\lambda)$ alors la variable $Z=Y_1+\ldots+Y_n$ suit une loi gamma $\mathcal{G}amma(n,\lambda)$.

Pour cela, il suffit de calculer la fonction de répartition de la somme. Pour tout $\xi\in\mathbb{R}$, nous avons :

$$\begin{aligned}\Phi_Z(\xi) \quad &= \quad \mathbb{E}\left[e^{-i\xi Z}\right]\\ &= \quad \mathbb{E}\left[\prod_{i=1}^{n}e^{-i\xi Y_i}\right]\\ &= \quad \prod_{i=1}^{n}\mathbb{E}\left[e^{-i\xi Y_i}\right] \text{ par indépendance car } n\text{-échantillon}\\ &= \quad \prod_{i=1}^{n}\Phi_{Y_i}(\xi)\\ &= \quad \prod_{i=1}^{n}\left(1-\frac{i\xi}{\lambda}\right)^{-1} \text{ par hypothèse}\\ &= \quad \left(1-\frac{i\xi}{\lambda}\right)^{-n}\end{aligned}$$

et nous retrouvons la fonction caractéristique de la loi gamma $\Gamma(n, \lambda)$.

11. Si Z suit une loi $\mathcal{G}amma(\alpha, \beta)$ avec $\alpha > 0$ et $\beta > 0$, nous disons que $W = 1/Z$ suit une loi inverse gamma $\mathcal{I}nv\mathcal{G}amma(\alpha, \beta)$. Dans le cas où $\alpha > 1$, montrer que l'espérance de W vaut $\beta/(\alpha - 1)$.

Nous avons :

$$\begin{aligned}
\mathbb{E}[W] &= \int_0^{+\infty} w\beta^\alpha/\Gamma(\alpha)w^{-\alpha-1}e^{-\beta/w}\mathrm{d}w \\
&= \int_0^{+\infty} \beta^\alpha/\Gamma(\alpha)w^{-\alpha}e^{-\beta/w}\mathrm{d}w \\
u(w) = e^{-\beta/w}, &\qquad u'(w) = \frac{\beta}{w^2}e^{-\beta/w}, \\
v'(w) = w^{-\alpha} \quad &\text{et} \quad v(w) = \frac{1}{-\alpha+1}w^{-\alpha+1} \\
&= \underbrace{\left[\beta^\alpha/\Gamma(\alpha) \times \frac{1}{-\alpha+1}w^{-\alpha+1}e^{-\beta/w}\right]_0^{+\infty}}_{=0 \text{ car } \alpha>1} - \int_0^{+\infty} \beta^\alpha/\Gamma(\alpha)\frac{1}{-\alpha+1}w^{-\alpha+1}\frac{\beta}{w^2}e^{-\beta/w}\mathrm{d}w \\
&= \frac{\beta}{\alpha-1}\underbrace{\int_0^{+\infty} \beta^\alpha/\Gamma(\alpha)w^{-\alpha-1}e^{-\beta/w}\mathrm{d}w}_{=1 \text{ par définition}}
\end{aligned}$$

D'où le résultat.

12. L'estimateur $\tilde{\theta}_n$ est-il biaisé et/ou asymptotiquement biaisé ?

Par la question 8, nous savons que les variables $(-\log X_1, \ldots, -\log X_n)$ sont iid de même loi $\mathcal{E}(\theta^\star)$; par la question 9, nous savons que la variable $\sum_{i=1}^n (-\log X_i)$ suit donc une loi gamma $\Gamma(n, \theta^\star)$ et par la question 10, nous savons que $\tilde{\theta}_n/n$ suit une loi inverse gamma $\mathcal{I}nv\Gamma(n, \theta^\star)$ donc nous avons :

$$\begin{aligned}
\mathbb{E}_{\theta^\star}\left[\tilde{\theta}_n\right] &= n\mathbb{E}_{\theta^\star}\left[\tilde{\theta}_n/n\right] \\
&= n \times \frac{\theta^\star}{n-1} \text{ par la question 10,} \\
&= \frac{n}{n-1}\theta^\star \\
&\underset{n\to+\infty}{\longrightarrow} \theta^\star.
\end{aligned}$$

Donc l'estimateur $\tilde{\theta}_n$ est biaisé mais asymptotiquement non biaisé.

Nous rappelons que

- *la loi $\mathcal{G}amma(\alpha, \beta)$ a pour densité :*

$$f_Z(z) = \beta^\alpha/\Gamma(\alpha)z^{\alpha-1}e^{-\beta z}\mathbb{1}_{\mathbb{R}_+^\star}(z),$$

- *la loi inverse gamma $\mathcal{I}nv\mathcal{G}amma(\alpha, \beta)$ a pour densité :*

$$f_W(w) = \beta^\alpha/\Gamma(\alpha)w^{-\alpha-1}e^{-\beta/w}\mathbb{1}_{\mathbb{R}_+^\star}(w),$$

- *la fonction Γ est définie pour $x \in \mathbb{R}_+^\star$ par :*

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