M1 – Statistique – 2018/2019

Examen final
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 de deux exercices indépendants pour un total de 9 questions. ***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. Si la réponse n'est pas correcte mais que le candidat ou la candidate s'en aperçoit et met un commentaire montrant un recul sur son travail, des points pourront éventuellement être accordés ; sinon, aucun point ne sera accordé à la question.

Intervalle de confiance

Dans cette partie, nous étudions un n-échantillon de loi exponentielle $\mathcal{E}(\theta^\star)$ avec $\theta^\star \in \mathbb{R}_+^\star$ dont nous rappelons la densité :

$$f_{\theta^\star}(x) = \theta^\star e^{-\theta^\star x}\mathbb{1}_{\mathbb{R}_+}(x).$$

1. En détaillant les calculs de l'espérance, calculer l'estimateur $\widehat{\theta}_n$ basé sur le moment d'ordre 1.

Nous avons :

$$\begin{aligned}
\mathbb{E}_{\theta^\star}[X] &= \int_0^{+\infty} x\theta^\star e^{-\theta^\star x}\mathrm{d}x \\
u(x) = x, &\quad u'(x) = 1, \\
v'(x) = \theta^\star e^{-\theta^\star x}, &\quad \text{et} \quad v(x) = -e^{-\theta^\star x}, \\
&= \left[-xe^{-\theta^\star x}\right]_0^{+\infty} + \int_0^{+\infty} e^{-\theta^\star x}\mathrm{d}x \\
&= 0 + \left[-\frac{1}{\theta^\star}e^{-\theta^\star x}\right]_0^{+\infty} \\
&= \frac{1}{\theta^\star}.
\end{aligned}$$

$$\overline{X_n} = \frac{1}{\widehat{\theta}_n} \Leftrightarrow \widehat{\theta}_n = \frac{1}{\overline{X_n}}.$$

2. En détaillant les calculs de la variance, calculer la loi limite de $\overline{X_n}$.

Nous avons :

$$\begin{array}{rcl}
\mathbb{E}_{\theta^\star}\left[X^2\right] & = & \int_0^{+\infty} x^2\theta^\star e^{-\theta^\star x}\mathrm{d}x \\
u(x) = x^2, & & u'(x) = 2x, \\
v'(x) = \theta^\star e^{-\theta^\star x}, & \text{et} & v(x) = -e^{-\theta^\star x}, \\
& = & \left[-x^2 e^{-\theta^\star x}\right]_0^{+\infty} + \int_0^{+\infty} 2x e^{-\theta^\star x}\mathrm{d}x \\
& = & 0 + \frac{2}{\theta^\star}\int_0^{+\infty} x\theta^\star e^{-\theta^\star x}\mathrm{d}x \\
& = & \frac{2}{\theta^\star} \times \frac{1}{\theta^\star} \\
& = & \frac{2}{\theta^{\star 2}}.
\end{array}$$

Nous obtenons ainsi :

$$\mathbb{V}_{\theta^\star}\left[X\right] = \mathbb{E}_{\theta^\star}\left[X^2\right] - \mathbb{E}_{\theta^\star}\left[X\right]^2 = \frac{2}{\theta^{\star 2}} - \frac{1}{\theta^{\star 2}} = \frac{1}{\theta^{\star 2}}.$$

Donc, d'après le théorème central limite, nous avons :

$$\sqrt{n}\left(\overline{X_n} - \frac{1}{\theta^\star}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{\theta^{\star 2}}\right).$$

3. Calculer la loi limite de $\widehat{\theta}_n$.

Nous voyons que $\sqrt{n}$ tend vers l'infini avec n et nous posons $\ell(x) = 1/x$ qui est dérivable en $1/\theta^\star$ car dérivable sur $\mathbb{R}_+^\star$ avec pour dérivée :

$$\ell'(x) = -\frac{1}{x^2}.$$

Nous pouvons donc utiliser la méthode Delta :

$$\begin{array}{rl}
& \sqrt{n}\left[\ell\left(\overline{X_n}\right) - \ell\left(\frac{1}{\theta^\star}\right)\right] \xrightarrow[n\to+\infty]{\mathcal{L}} \ell'\left(1/\theta^\star\right)\mathcal{N}\left(0, \frac{1}{\theta^{\star 2}}\right) \\
\Leftrightarrow & \sqrt{n}\left(\widehat{\theta}_n - \theta^\star\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \theta^{\star 2}\mathcal{N}\left(0, \frac{1}{\theta^{\star 2}}\right)
\end{array}$$

$$\Leftrightarrow \quad \sqrt{n}\left(\widehat{\theta}_n - \theta^\star\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \theta^{\star 2}\right).$$

4. En isolant $\theta^\star$, déduire du résultat précédent un intervalle de confiance asymptotique de $\theta^\star$ de niveau $1-\alpha$ avec $\alpha \in]0,1[$.

Comme $\sqrt{n}\frac{\widehat{\theta}_n - \theta^\star}{\theta^\star}$ suit asymptotiquement une loi gaussienne centrée réduite et en note q_α le quantile d'ordre α de la gaussienne centrée réduite, nous avons :

$$\begin{aligned}
&\mathbb{P}_{\theta^\star}\left(q_{\alpha/2} \leq \sqrt{n}\frac{\widehat{\theta}_n - \theta^\star}{\theta^\star} \leq q_{1-\alpha/2}\right) \underset{n\to+\infty}{\longrightarrow} 1-\alpha \\
\Leftrightarrow \quad &\mathbb{P}_{\theta^\star}\left(-\frac{q_{1-\alpha/2}}{\sqrt{n}} \leq \frac{\widehat{\theta}_n}{\theta^\star} - 1 \leq \frac{q_{1-\alpha/2}}{\sqrt{n}}\right) \underset{n\to+\infty}{\longrightarrow} 1-\alpha \\
\Leftrightarrow \quad &\mathbb{P}_{\theta^\star}\left(1-\frac{q_{1-\alpha/2}}{\sqrt{n}} \leq \frac{\widehat{\theta}_n}{\theta^\star} \leq 1+\frac{q_{1-\alpha/2}}{\sqrt{n}}\right) \underset{n\to+\infty}{\longrightarrow} 1-\alpha \\
\Leftrightarrow \quad &\mathbb{P}_{\theta^\star}\left(\frac{1}{1-\frac{q_{1-\alpha/2}}{\sqrt{n}}} \geq \frac{\theta^\star}{\widehat{\theta}_n} \geq \frac{1}{1+\frac{q_{1-\alpha/2}}{\sqrt{n}}}\right) \underset{n\to+\infty}{\longrightarrow} 1-\alpha \\
&\text{pour } n \text{ assez grand pour que } 1 > \frac{q_{1-\alpha/2}}{\sqrt{n}} \\
\Leftrightarrow \quad &\mathbb{P}_{\theta^\star}\left(\frac{\widehat{\theta}_n}{1-\frac{q_{1-\alpha/2}}{\sqrt{n}}} \geq \theta^\star \geq \frac{\widehat{\theta}_n}{1+\frac{q_{1-\alpha/2}}{\sqrt{n}}}\right) \underset{n\to+\infty}{\longrightarrow} 1-\alpha.
\end{aligned}$$

Donc, l'intervalle de confiance asymptotique de niveau $1-\alpha$ vaut :

$$IC_{1-\alpha}(\theta^\star) = \left[\frac{\widehat{\theta}_n}{1+\frac{q_{1-\alpha/2}}{\sqrt{n}}}, \frac{\widehat{\theta}_n}{1-\frac{q_{1-\alpha/2}}{\sqrt{n}}}\right]$$

si $1 > \frac{q_{1-\alpha/2}}{\sqrt{n}}$. Sinon, c'est simplement :

$$IC_{1-\alpha}(\theta^\star) = \left[\frac{\widehat{\theta}_n}{1+\frac{q_{1-\alpha/2}}{\sqrt{n}}}, +\infty\right[.$$

Test

Dans cette partie, nous étudions des lancers d'une pièce et nous cherchons à savoir si la pièce est équilibrée ou non ; c'est-à-dire si nous avons autant de chance d'obtenir *pile* ou *face*. Pour ce faire, nous modélisons les lancers par un n-échantillon de loi de Bernoulli $\mathcal{B}(\theta^\star)$ avec $\theta^\star \in [0, 1]$ où 0 représente *face* et 1 *pile*.

5. Expliquer pourquoi nous pouvons modéliser cette problématique par un test comparant les deux hypothèses suivantes :

$$\mathcal{H}_0 : \theta^\star = 1/2 \text{ contre } \mathcal{H}_1 : \theta^\star \neq 1/2.$$

Donner les deux sous-ensembles correspondants.

Comme nous cherchons à savoir si la pièce est équilibrée, nous voulons savoir si le paramètre vaut 1/2. L'alternative est alors toutes les autres possibilités. Les deux ensembles sont :

$$\Theta_0 = 1/2 \text{ et } \Theta_1 = [0, 1/2[\cup]1/2, 1].$$

6. Calculer l'estimateur du maximum de vraisemblance du modèle.

Commençons par calculer la vraisemblance en fonction de θ :

$$\begin{aligned} V_{\mathbf{X}}(\theta) &= \prod_{i=1}^{n} \theta^{\star X_i} (1-\theta^\star)^{1-X_i} \\ &= \theta^{\star \sum_{i=1}^n X_i} (1-\theta^\star)^{n-\sum_{i=1}^n X_i}. \end{aligned}$$

Puis calculons le maximum de vraisemblance sur $[0, 1]$:

$$\begin{aligned} \ln(V_{\mathbf{X}}(\theta^\star)) &= \ln(\theta^\star) \sum_{i=1}^n X_i + \left(n - \sum_{i=1}^n X_i\right) \ln(1-\theta^\star), \\ \frac{\partial}{\mathrm{d}\theta} \ln(V_{\mathbf{X}}(\theta)) = 0 &\Leftrightarrow \frac{1}{\theta} \sum_{i=1}^n X_i - \left(n - \sum_{i=1}^n X_i\right) \frac{1}{1-\theta} = 0 \\ &\Leftrightarrow (1-\theta) \sum_{i=1}^n X_i - \left(n - \sum_{i=1}^n X_i\right) \theta = 0 \\ &\Leftrightarrow \sum_{i=1}^n X_i - \theta \sum_{i=1}^n X_i - n\theta - \theta \sum_{i=1}^n X_i = 0 \\ &\Leftrightarrow \theta = \overline{X_n}, \\ \frac{\partial^2}{\mathrm{d}\theta^2} \ln(V_{\mathbf{X}}(\theta)) &= -\frac{1}{\theta^2} \sum_{i=1}^n X_i - \left(n - \sum_{i=1}^n X_i\right) \frac{1}{(1-\theta)^2} \end{aligned}$$

$$< \quad 0.$$

Donc l'estimateur du maximum de vraisemblance est $\overline{X_n}$.

7. Montrer que la statistique du rapport de vraisemblance $h(\mathbf{X})$ est égale à $g_n(\overline{X_n})$ avec

$$\begin{array}{rcl} g_n :]0,1[& \to & \mathbb{R}_+^\star \\ x & \mapsto & 2^n x^{nx} (1-x)^{n(1-x)}. \end{array}$$

Or, nous savons que la statistique du rapport de vraisemblance vaut :

$$\begin{aligned} h(\mathbf{X}) &= \frac{\sup_{\theta\in\Theta_1} V_{\mathbf{X}}(\theta)}{\sup_{\theta\in\Theta_0} V_{\mathbf{X}}(\theta)} \\ &= \frac{\sup_{\theta\neq 1/2} V_{\mathbf{X}}(\theta)}{\sup_{\theta=1/2} V_{\mathbf{X}}(\theta)} \\ &= \frac{\sup_{\theta\neq 1/2} V_{\mathbf{X}}(\theta)}{V_{\mathbf{X}}(1/2)}. \end{aligned}$$

Pour le numérateur, comme la vraisemblance est croissante jusqu'à $\overline{X_n}$ et décroissante après, si $\overline{X_n} \neq 1/2$, le maximum appartient donc à l'ensemble, sinon, même si l'abscisse du maximum n'est pas dans l'ensemble, le sup est également $V_{\mathbf{X}}(\overline{X_n})$. Au final, la statistique du rapport de vraisemblance est donc :

$$\begin{aligned} h(\mathbf{X}) &= \frac{V_{\mathbf{X}}(\overline{X_n})}{V_{\mathbf{X}}(1/2)} \\ &= \frac{\overline{X_n}^{\sum_{i=1}^n X_i} \left(1-\overline{X_n}\right)^{n-\sum_{i=1}^n X_i}}{\left(\frac{1}{2}\right)^{\sum_{i=1}^n X_i} \left(1-\frac{1}{2}\right)^{n-\sum_{i=1}^n X_i}} \\ &= \frac{\overline{X_n}^{\sum_{i=1}^n X_i} \left(1-\overline{X_n}\right)^{n-\sum_{i=1}^n X_i}}{\left(\frac{1}{2}\right)^{\sum_{i=1}^n X_i} \left(\frac{1}{2}\right)^{n-\sum_{i=1}^n X_i}} \\ &= 2^n \overline{X_n}^{n\overline{X_n}} \left(1-\overline{X_n}\right)^{n(1-\overline{X_n})} \\ &= g_n(\overline{X_n}). \end{aligned}$$

8. En étudiant le sens de variation du logarithme de la fonction g_n, montrer que le test du rapport de vraisemblance est équivalent au test :

$$\phi(\mathbf{X}) = \mathbb{1}_{\{\overline{X_n}<\ell_\alpha^1 \text{ ou } \ell_\alpha^2<\overline{X_n}\}}$$

avec $\ell_\alpha^1 < \ell_\alpha^2$ qui seront définis dans la question suivante.

Commençons par étudier le sens de variation de $\ln(g_n(x))$:

$$\begin{aligned}
\ln(g_n(x)) &= n\ln(2) + nx\ln(x) + n(1-x)\ln(1-x), \\
\frac{\partial}{\mathrm{d}x}\ln(g_n(x)) = 0 &\Leftrightarrow n\ln(x) + n - n\ln(1-x) - n = 0 \\
&\Leftrightarrow n\ln(x) - n\ln(1-x) = 0 \\
&\Leftrightarrow \ln(x) = \ln(1-x) \\
&\Leftrightarrow x = 1-x \\
&\Leftrightarrow x = 1/2, \\
\frac{\partial^2}{\mathrm{d}x^2}\ln(g_n(x)) &= \frac{n}{x} + \frac{n}{1-x} \\
&> 0.
\end{aligned}$$

Donc la fonction g_n est décroissant sur $]0, 1/2[$ et croissante sur $]1/2, 1[$. Or, nous savons que le test du rapport de vraisemblance s'écrit :

$$\phi(\mathbf{X}) = \mathbb{1}_{\{h(\overline{X_n}) > k_\alpha\}}$$

et la fonction est plus grande qu'une certaine valeur k_α que si $\overline{X_n}$ est plus petit qu'un certain ℓ_α^1 ou plus grand qu'un certain ℓ_α^2.

9. En étudiant la loi asymptotique de $\overline{X_n}$ sous l'hypothèse $\mathcal{H}_0$, proposer des valeurs pour ℓ_α^1 et ℓ_α^2 de telle sorte que le test soit de niveau α.

Le théorème de limite centrale permet de dire que :

$$\sqrt{n}\left(\overline{X_n} - 1/2\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{4}\right)$$

donc, en notant q_α le quantile de la loi gaussienne centrée réduite, nous avons :

$$\begin{aligned}
&\mathbb{P}_{1/2}\left(q_{\alpha/2} \le 2\sqrt{n}\left(\overline{X_n} - 1/2\right) \le q_{1-\alpha/2}\right) \xrightarrow[n\to+\infty]{} 1-\alpha \\
\Leftrightarrow\ &\mathbb{P}_{1/2}\left(-\frac{q_{1-\alpha/2}}{2\sqrt{n}} \le \overline{X_n} - 1/2 \le \frac{q_{1-\alpha/2}}{2\sqrt{n}}\right) \xrightarrow[n\to+\infty]{} 1-\alpha \\
\Leftrightarrow\ &\mathbb{P}_{1/2}\left(\frac{1}{2} - \frac{q_{1-\alpha/2}}{2\sqrt{n}} \le \overline{X_n} \le \frac{1}{2} + \frac{q_{1-\alpha/2}}{2\sqrt{n}}\right) \xrightarrow[n\to+\infty]{} 1-\alpha.
\end{aligned}$$

Donc, si nous prenons $\ell_\alpha^1 = \frac{1}{2} - \frac{q_{1-\alpha/2}}{2\sqrt{n}}$ et $\ell_\alpha^2 = \frac{1}{2} + \frac{q_{1-\alpha/2}}{2\sqrt{n}}$ alors nous nous tromperons asymptotiquement avec une probabilité de α.