Correction du TD 4

Exercice 1 Dans tout l'exercice, le modèle statistique considéré est $\mathcal{P} = \{f_\theta \cdot \lambda_{\mathbb{R}}, \theta \in \mathbb{R}\}$, qui est dominé par la mesure de Lebesgue sur $\mathbb{R}$, notée $\lambda_{\mathbb{R}}$, et chaque loi du modèle indicée par θ a pour densité f_θ.

1. La log vraisemblance s'écrit

$$\begin{aligned}\ell_n(\theta) &= \sum_{i=1}^n \ln f(X_i - \theta) \\ &= -(n/2)\ln(2\pi\sigma^2) - 1/(2\sigma^2)\sum_{i=1}^n (X_i-\theta)^2.\end{aligned}$$

D'après l'exercice 1 du TD 3, cette fonction est maximale pour $\theta = \overline{X}_n$. On peut également remarquer que cette fonction est strictement concave en θ, elle admet donc un unique maximum, en son unique point critique.

2. Cette fois, la log-vraisemblance s'écrit

$$\ell_n(\theta) = -n\ln(2) - \sum_{i=1}^n |X_i - \theta|,$$

qui est maximale pour $\theta = F_n^{-1}(1/2)$, toujours d'après l'exercice 1 question 2)b. du TD 3 (on notera que $\underset{\theta}{\operatorname{argmin}}\ 1/n\sum_{i=1}^n |X_i-\theta| = \underset{\theta}{\operatorname{argmin}}\ \sum_{i=1}^n |X_i-\theta| = \underset{\theta}{\operatorname{argmax}}\ -\sum_{i=1}^n |X_i-\theta|$). On retrouve ici la médiane empirique.

3. La vraisemblance $L_n(\theta)$ est non nulle si et seulement si pour tout i, $-1 < X_i - \theta < 1$, c'est-à-dire si et seulement si $\theta \in]X_{(n)} - 1, X_{(1)} + 1[$ où $X_{(n)}$ est le maximum des $X_i, i = 1, \ldots, n$ et $X_{(1)}$ leur minimum. Ainsi, $L_n(\theta) = 0$ pour $\theta \notin]X_{(n)} - 1, X_{(1)} + 1[$. De plus pour $\theta \in]X_{(n)} - 1, X_{(1)} + 1[$,

$$\ell_n(\theta) = n\ln(3/4) + \sum_{i=1}^n \ln(1 - (X_i - \theta)^2).$$

L'EMV existe car c'est une fonction continue sur l'intervalle ouvert et de limite $-\infty$ aux bornes de cet intervalle, donc elle atteint son maximum. Cependant, la recherche de points critiques revient à trouver les racines d'un polynôme de degré $2n-1$. En effet, en annulant la dérivée par rapport à θ, on obtient

$$\sum_{i=1}^n \frac{(X_i - \theta)}{1 - (X_i - \theta)^2} = 0,$$

qui, en réduisant au même dénominateur, donne

$$\frac{1}{\prod_{i=1}^n (1 - (X_i-\theta)^2)} \left(\sum_{i=1}^n (X_i - \theta) \underbrace{\prod_{j\neq i} \left(1 - (X_j - \theta)^2\right)}_{\text{degré}=2(n-1)} \right) = 0.$$

Donc l'EMV n'est pas explicite. Nous pouvons cependant l'approcher numériquement.

Exercice 2

1. Comme $\mathbb{E}X_1 = 2\theta/(1-\theta)$, on a $\theta = \mathbb{E}X_1/(2+\mathbb{E}X_1)$. Ainsi, l'estimateur empirique associé est
$$\widehat{\theta}_n = \frac{\overline{X}_n}{2+\overline{X}_n}.$$
Notons que $\widehat{\theta}_n < 1$ mais on peut avoir $\widehat{\theta}_n = 0$ avec probabilité strictement positive, égale à $(1-\theta)^{2n}$.
2. Ici, le modèle statistique considéré est $\mathcal{P} = \{f_\theta \cdot \nu_{\mathbb{N}}, \theta \in]0,1[\}$, où pour tout $x \in \mathbb{N}$,
$$f_\theta(x) = \mathbb{P}_\theta(X = x) = (x+1)(1-\theta)^2\theta^x.$$
Ce modèle est dominé par la mesure de comptage sur $\mathbb{N}$, notée $\nu_{\mathbb{N}}$, et chaque loi du modèle indicée par θ a pour densité f_θ. Ainsi, nous avons pour tout $\theta \in]0,1[$,
$$\begin{aligned}\ell_n(\theta) &= \sum_{i=1}^n \ln f_\theta(X_i)\\ &= 2n\ln(1-\theta) + n\overline{X}_n\ln(\theta) + \sum_{i=1}^n \ln(X_i+1).\end{aligned}$$
Donc la dérivée vaut
$$-2n(1-\theta)^{-1} + n\overline{X}_n\theta^{-1},$$
qui est positive si et seulement si $\theta \leq \widehat{\theta}_n$, où $\widehat{\theta}_n$ est défini à la question précédente. Ainsi, si $\widehat{\theta}_n > 0$ (i.e. $\overline{X}_n > 0$), alors la fonction $\theta \mapsto L_n(\theta)$ atteint son maximum sur $]0,1[$ en $\widehat{\theta}_n$ et l'EMV vaut $\widetilde{\theta}_n = \widehat{\theta}_n$. Si $\widehat{\theta}_n = 0$ (i.e. $\overline{X}_n = 0$), alors la vraisemblance est strictement décroissante sur $]0,1[$ donc l'EMV n'est pas défini.
3. La LFGN appliquée aux X_i i.i.d. intégrables donne que $\overline{X}_n$ converge p.s. vers $\mathbb{E}(X_1)$, donc par le théorème de continuité $\widehat{\theta}_n$ converge p.s. vers θ. Ainsi, $\widehat{\theta}_n$ est consistant. Le TCL appliqué aux X_i i.i.d. de carré intégrable donne
$$\sqrt{n}(\overline{X}_n - 2\theta/(1-\theta)) \rightsquigarrow \mathcal{N}(0, 2\theta/(1-\theta)^2).$$
La méthode delta appliquée avec la fonction $g(x) = x/(2+x) = 1 - 2/(2+x)$, qui est dérivable en $x = 2\theta/(1-\theta) > -2$ de dérivée
$$g'(x) = \frac{2}{(2+x)^2} = \frac{2}{(2+2\theta/(1-\theta))^2} = \frac{(1-\theta)^2}{2},$$
donne
$$\sqrt{n}(\widehat{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, 2\theta/(1-\theta)^2 \times (1-\theta)^4/4) = \mathcal{N}(0, \theta(1-\theta)^2/2).$$

Exercice 3

1. Ici, le modèle statistique considéré est $\mathcal{P} = \{f_\theta \cdot \lambda_{\mathbb{R}}, \theta \in \mathbb{R}\}$, qui est dominé par la mesure de Lebesgue sur $\mathbb{R}$, notée $\lambda_{\mathbb{R}}$, et chaque loi du modèle indicée par θ a pour densité f_θ. Il s'agit d'un modèle de translation, puisque X_1 est de même loi que $Y + \theta$, où Y est de loi exponentielle de paramètre 1 (de densité f_0). La vraisemblance de l'échantillon $X_1, \ldots, X_n$ s'écrit
$$L_n(\theta) = \exp\left(-\sum_{i=1}^n X_i + n\theta\right)\mathbf{1}_{\theta \leq X_{(1)}}.$$

La fonction de vraisemblance $\theta \mapsto L_\theta(X_1, \dots, X_n)$ est nulle si $\theta > X_{(1)}$ et croissante pour $\theta \leq X_{(1)}$, donc elle atteint son maximum pour $\widehat{\theta}_n = X_{(1)}$, qui est par conséquent l'estimateur du maximum de vraisemblance de θ.

On pourrait aussi proposer un estimateur par la méthode des moments : comme $\mathbb{E}X_1 = \theta + \mathbb{E}Y = \theta + 1$, ceci conduit à considérer l'estimateur $\overline{X}_n - 1$. Il est cependant moins bon que l'EMV car il converge à vitesse $1/\sqrt{n}$ par le TCL, alors que $\widehat{\theta}_n$ converge à vitesse $1/n$, comme vont le montrer les questions suivantes.

2. Nous avons, pour tout $x \geq \theta$:

$$\mathbb{P}_\theta[\widehat{\theta}_n \leq x] = 1 - \mathbb{P}_\theta[X_{(1)} > x] = 1 - \mathbb{P}_\theta[X_1 - \theta > x - \theta]^n = 1 - e^{-n(x-\theta)},$$

et 0 sinon. Ainsi, pour tout $x \geq 0$, $\frac{x}{n} + \theta \geq \theta$ et nous obtenons :

$$\mathbb{P}_\theta\left[\widehat{\theta}_n \leq \frac{x}{n} + \theta\right] = 1 - e^{-x},$$

et 0 sinon. Puisque $\mathbb{P}_\theta\left[\widehat{\theta}_n \leq \frac{x}{n} + \theta\right] = \mathbb{P}_\theta\left[n(\widehat{\theta}_n - \theta) \leq x\right]$, on constate que $n(\widehat{\theta}_n - \theta)$ suit une loi exponentielle de paramètre 1.

On part de ce constat pour construire un intervalle de confiance. On commence par remarquer que $X_i \geq \theta$, donc $\widehat{\theta}_n \geq \theta$. Si $\alpha \in]0,1[$ est un niveau fixé, on cherche donc $c > 0$ tel que

$$\mathbb{P}_\theta[\widehat{\theta}_n - c \leq \theta \leq \widehat{\theta}_n] = 1 - \alpha.$$

On a, d'après ce qui précède :

$$\mathbb{P}_\theta[\widehat{\theta}_n - c \leq \theta \leq \widehat{\theta}_n] = \mathbb{P}_\theta[\widehat{\theta}_n - c \leq \theta] = 1 - \mathbb{P}_\theta[n(\widehat{\theta}_n - \theta) \geq nc] = 1 - e^{-nc} = 1 - \alpha$$

si on choisit $c = c_\alpha = -(\ln \alpha)/n$. Donc

$$I_{1-\alpha} = \left[\widehat{\theta}_n + \frac{\ln \alpha}{n}, \widehat{\theta}_n\right]$$

est un intervalle de confiance de niveau $1 - \alpha$.

Remarque : On aurait pu partir d'une forme a priori d'intervalle de confiance, par exemple $\left[\widehat{\theta}_n - c_\alpha, \widehat{\theta}_n + d_\alpha\right]$, avec $c_\alpha \geq 0$ et $d_\alpha \geq 0$. Un simple calcul montre alors que $\left[\widehat{\theta}_n + \frac{\ln \alpha}{n}, \widehat{\theta}_n + d_\alpha\right]$ est intervalle de confiance de niveau $1-\alpha$ pour θ, et ce quel que soit $d_\alpha \geq 0$. La raison conduit à prendre l'intervalle de confiance le plus court, c'est-à-dire celui correspondant à $d_\alpha = 0$. Nous retombons ainsi sur le même résultat que celui exhibé précédemment.

3. (a) On rejette H_0 si $I_{1-\alpha} \cap \mathbb{R}_+ = \emptyset$: d'après le cours, ceci fournit un test de **niveau** α. La puissance du test est donnée pour tout θ par

$$\begin{aligned}
\pi(\theta) &= \mathbb{P}_\theta(\text{on rejette } H_0) \\
&= \mathbb{P}_\theta(\widehat{\theta}_n < 0) \\
&= \mathbb{P}_\theta(n(\widehat{\theta}_n - \theta) < -n\theta) \\
&= \mathbb{P}(\mathcal{E}(1) < -n\theta) \\
&= \begin{cases} 1 - e^{n\theta} & \text{si } \theta \leq 0 \\ 0 & \text{si } \theta \geq 0. \end{cases}
\end{aligned}$$

En particulier, on remarque que le test est de taille 0, il est donc bien de niveau α, mais cette situation ne semble pas idéale : on s'attend plutôt à $\pi(0) = 0.05$ si le test était de taille 0.05. C'est l'objet de la suite.

(b) A priori, un test logique consiste à rejeter H_0 si $\widehat{\theta}_n < c_\alpha$ tel que

$$\sup_{\theta\geq 0}\{\mathbb{P}_\theta(\widehat{\theta}_n < c_\alpha)\} = \alpha.$$

On cherche donc à déterminer c_α. Or, en se souvenant que la variable aléatoire $E = n(\widehat{\theta}_n - \theta)$ suit une loi exponentielle de paramètre 1, on a

$$\begin{aligned}
\sup_{\theta\geq 0}\mathbb{P}_\theta(\widehat{\theta}_n < c_\alpha) &= \sup_{\theta\geq 0}\mathbb{P}_\theta(n(\widehat{\theta}_n - \theta) < n(c_\alpha - \theta)) \\
&= \sup_{\theta\geq 0}\mathbb{P}(E < n(c_\alpha - \theta)) \\
&= \mathbb{P}(E < nc_\alpha)
\end{aligned}$$

car la fonction $\theta \mapsto \mathbb{P}_\theta(E < n(c_\alpha - \theta))$ est décroissante. Dès lors, on choisit c_α tel que

$$\mathbb{P}_\theta(E < nc_\alpha) = (1 - e^{-nc_\alpha})\mathbf{1}_{c_\alpha\geq 0} = \alpha$$

c'est-à-dire

$$c_\alpha = -\frac{\ln(1-\alpha)}{n}$$

et on vérifie que $c_\alpha \geq 0$. Donc la région de rejet est $\{\widehat{\theta}_n < -\ln(1-\alpha)/n\}$. Le test ainsi construit est bien de taille α, c'est-à-dire

$$\sup_{\theta\geq 0}\{\mathbb{P}_\theta(\widehat{\theta}_n < -\ln(1-\alpha)/n)\} = \alpha.$$

Notons que le test proposé précédemment (via le lien avec l'intervalle de confiance) était bien de niveau α (et même de niveau 0), mais pas de taille α. Ceci est illustré Figure 1.

Remarque : On aurait pu chercher un intervalle de confiance de sens opposé à $\Theta_0 = [0, \infty[$, par exemple $\left]-\infty, \widehat{\theta}_n + d_\alpha\right]$, avec d_α quelconque. Un simple calcul montre alors que $\left]-\infty, \widehat{\theta}_n + \frac{\ln(1-\alpha)}{n}\right]$ est intervalle de confiance de niveau $1-\alpha$ pour θ. Le test qui en découle est exactement celui obtenu plus haut.

(c) La puissance du test est donnée pour tout θ par

$$\begin{aligned}
\pi(\theta) &= \mathbb{P}_\theta(\text{on rejette } H_0) \\
&= \mathbb{P}_\theta(\widehat{\theta}_n < -\ln(1-\alpha)/n) \\
&= \mathbb{P}_\theta(n(\widehat{\theta}_n - \theta) < -\ln(1-\alpha) - n\theta) \\
&= \begin{cases} 1-(1-\alpha)e^{n\theta} & \text{si } \theta \leq -\ln(1-\alpha)/n, \\ 0 & \text{sinon} \end{cases}
\end{aligned}$$

(d) La puissance du test est croissante en α pour tout θ. Ceci est le cas pour tout test correctement calibré. Elle est décroissante en n pour $\theta \in [0, -\ln(1-\alpha)/n]$, mais croissante en n pour $\theta < 0$. Ainsi, pour tout θ, $\pi(\theta) \underset{n\to\infty}{\to} \mathbb{1}_{\theta<0}$, ce qui est attendu.

4. La v.a. X_1 suit une loi exponentielle si et seulement si $\theta = 0$, donc on construit un test pour H_0 : "$\theta = 0$" contre H_1 : "$\theta \neq 0$". Il s'agit d'un test d'adéquation, donc on obtient un test de niveau α en considérant simplement le test suivant, basé sur l'intervalle de confiance $I_{1-\alpha}$:

$$\begin{cases} \text{rejette } H_0 \text{ si } 0 \notin I_{1-\alpha} \\ \text{accepte } H_0 \text{ sinon.} \end{cases}$$

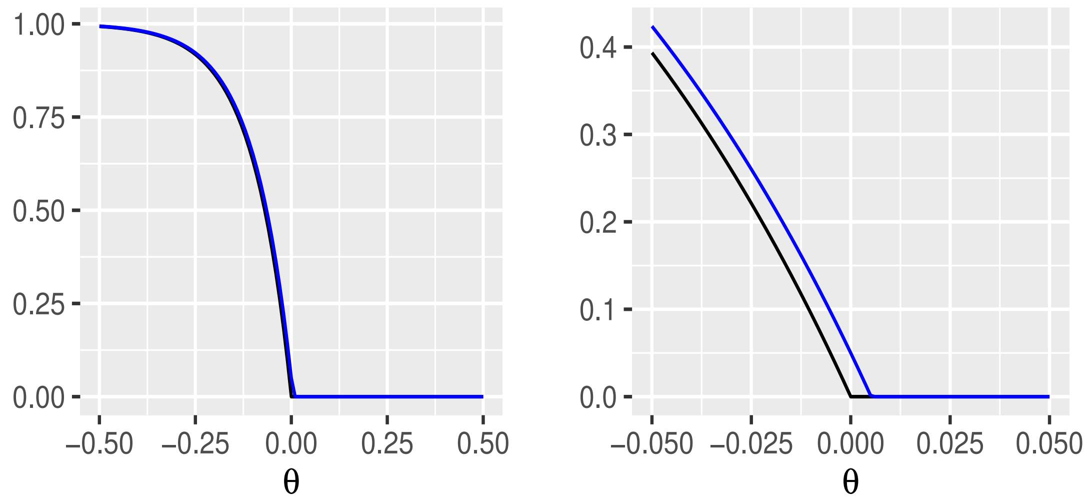

FIGURE 1 – Puissances de tests pour $n = 10$ et $\alpha = 0.05$: via les intervalles de confiance (noir) ou par méthode directe (bleu).

Pour tout $\theta \in \mathbb{R}$, la puissance du test est donnée par

$$\begin{aligned}\pi(\theta) = \mathbb{P}_\theta[0 \notin I_{1-\alpha}] &= \mathbb{P}_\theta[\widehat{\theta}_n + \frac{\ln \alpha}{n} > 0 \text{ ou } 0 > \widehat{\theta}_n] \\ &= \mathbb{P}_\theta[\widehat{\theta}_n > -\frac{\ln \alpha}{n}] + \mathbb{P}_\theta[\widehat{\theta}_n < 0] \\ &= \alpha e^{n\theta} \mathbf{1}_{n\theta \leq -\ln \alpha} + \mathbf{1}_{n\theta > -\ln \alpha} + \mathbf{1}_{\theta \leq 0}(1 - e^{n\theta}),\end{aligned}$$

la dernière étape étant obtenue après quelques lignes de calculs.

EXERCICE 4

1. Comme $\int f_\theta(x)dx = (1-\theta)/2 + (1+\theta)/2 = 1$, il suffit de vérifier que f_θ est une fonction positive. C'est le cas si et seulement si $\theta \in [-1, 1]$.
2. Ici, le modèle statistique considéré est $\mathcal{P} = \{f_\theta \cdot \lambda_\mathbb{R}, \theta \in [-1, 1]\}$, qui est dominé par la mesure de Lebesgue sur $\mathbb{R}$, notée $\lambda_\mathbb{R}$, et chaque loi du modèle indicée par θ a pour densité f_θ. On cherche à maximiser la vraisemblance sur $[-1, 1]$. En posant

$$Y_n = \sum_{i=1}^n \mathbf{1}_{X_i \in]0,1/2]},$$

nous avons pour tout $\theta \in]-1, 1[$,

$$\begin{aligned}\ell_n(\theta) &= \sum_{i=1}^n \ln f_\theta(X_i) \\ &= \sum_{i=1}^n \left(\ln(1-\theta)\mathbf{1}_{X_i \in]-1/2,0]} + \ln(1+\theta)\mathbf{1}_{X_i \in]0,1/2]}\right) \\ &= (n - Y_n)\ln(1-\theta) + Y_n \ln(1+\theta).\end{aligned}$$

Ainsi, la dérivée vaut

$$-\frac{n-Y_n}{1-\theta}+\frac{Y_n}{1+\theta}$$

donc est positive si et seulement si $\theta \leq 2Y_n/n-1$. Ainsi, si $Y_n/n \notin \{0,1\}$ alors la fonction $\theta \mapsto L_n(\theta)$ atteint son maximum sur $]-1,1[$ en $2Y_n/n-1$. Comme on a dans ce cas

$$L_n(-1)=L_n(1)=0,$$

la fonction $\theta \mapsto L_n(\theta)$ atteint son maximum sur $[-1,1]$ (fermé) en $2Y_n/n-1$, ce qui indique que lorsque $Y_n/n \notin \{0,1\}$, l'EMV existe sur $[-1,1]$ et vaut

$$\widehat{\theta}_n = 2Y_n/n-1.$$

Dans le cas où $Y_n/n=0$ (resp. $Y_n/n=1$), ce qui peut arriver avec probabilité strictement positive, on vérifie aisément que la fonction $\theta \mapsto L_n(\theta)$ (monotone) est maximum en $\widehat{\theta}_n=-1=2Y_n/n-1$ (resp. $\widehat{\theta}_n=1=2Y_n/n-1$). Finalement, lorsque l'espace des paramètres est $[-1,1]$, l'EMV existe dans tous les cas et vaut

$$\widehat{\theta}_n = 2Y_n/n-1,$$

qui peut éventuellement valoir -1 ou 1 (ce qui est autorisé car, encore une fois, l'espace des paramètres est $[-1,1]$).

3. Comme

$$\mathbb{E}_\theta(\widehat{\theta}_n)=\frac{2}{n}\mathbb{E}_\theta(Y_n)-1=2\mathbb{P}(X_1\in]0,1/2])-1=2(1+\theta)/2-1=\theta,$$

l'estimateur $\widehat{\theta}_n$ est sans biais. Notons que $Y_n/n=n^{-1}\sum_{i=1}^n \mathbf{1}_{X_i\in]0,1/2]}$ donc s'écrit comme une moyenne empirique, ce qui nous permet d'utiliser les théorèmes usuels (X_i sont i.i.d. et admettent des moments d'ordre 1 et d'ordre 2). Par la LFGN, Y_n/n converge presque sûrement vers $\mathbb{P}(X_1\in]0,1/2])=(1+\theta)/2$ donc $\widehat{\theta}_n$ converge p.s. vers θ. Il est donc consistant. Le TCL donne

$$\sqrt{n}(Y_n/n-(1+\theta)/2) \rightsquigarrow \mathcal{N}(0,(1-\theta)(1+\theta)/4)=\mathcal{N}(0,(1-\theta^2)/4),$$

d'où en multipliant par 2 (pas besoin de méthode Delta ici),

$$\sqrt{n}(\widehat{\theta}_n-\theta) \rightsquigarrow \mathcal{N}(0,1-\theta^2),$$

donc $\widehat{\theta}_n$ est asymptotiquement normal. La variance asymptotique est strictement positive si et seulement si $|\theta|<1$. Lorque $\theta=1$ ou -1, Y_n est p.s. constante et $\widehat{\theta}_n=\theta$ p.s., ce qui explique que la variance asymptotique est nulle.

Exercice 5

1. Pour déterminer la loi de Y_1, on calcule sa fonction de répartition. Dans un premier temps, nous avons :

$$\forall y<0, \quad \mathbb{P}(Y_1\leq y)=\mathbb{P}(X_1\leq y)=0,$$

et

$$\forall y\geq 1, \quad \mathbb{P}(Y_1\leq y)=1.$$

Enfin

$$\forall y\in[0,1[, \quad \mathbb{P}(Y_1\leq y)=\mathbb{P}(X_1\leq y)=1-e^{-\theta^\star y}.$$

Pour conclure, la fonction de répartition $F_{\theta^\star}$ de Y_1 est

$$F_{\theta^\star}(y) = \begin{cases} 0 & \text{si } y < 0 \\ 1 - \mathrm{e}^{-\theta^\star y} & \text{si } y \in [0, 1[\\ 1 & \text{si } y \geq 1. \end{cases}$$

Notons que :

$$\mathbb{P}(Y_1 = 1) = \mathbb{P}(X_1 > 1) = 1 - \mathbb{P}(X_1 \leq 1) = \mathrm{e}^{-\theta^\star}.$$

Par définition, la fonction de répartition d'une loi discrète possède une quantité au plus dénombrable de discontinuités, ainsi, les valeurs prises par une telle fonction de répartition constituent un ensemble au plus dénombrable. Or l'image de la fonction de répartition considérée ici, $[0, 1 - \mathrm{e}^{-\theta^\star}[\cup \{1\}$, n'est pas dénombrable. Ceci assure que Y_1 n'est donc pas discrète. On aurait aussi pu remarquer que sur $[0, \frac{1}{2}]$, la fonction de répartition est continue et strictement croissante. En conséquence, elle n'est pas purement discontinue et Y_1 n'est donc pas discrète. Par ailleurs, Y_1 n'est pas non plus à densité puisque la fonction de répartition n'est pas continue en 1.

2. Soient $\lambda_{[0,1[}$ la mesure de Lebesgue sur $[0, 1[$ et δ_1 la mesure de Dirac en 1. Alors $\mu = \lambda_{[0,1[} + \delta_1$ est une mesure dominante du modèle $(P_\theta)_{\theta>0}$. En effet, pour tout $\theta > 0$ et tous $0 \leq a \leq b \leq 1$, si $\mu(]a, b]) = \lambda_{[0,1[}(]a, b]) + \delta_1(]a, b]) = 0$, alors $\lambda_{[0,1[}(]a, b]) = 0$ et $\delta_1(]a, b]) = 0$ (car les mesures sont positives). Cela signifie que $a = b$ et que $1 \notin]a, b]$, autrement dit que $b < 1$. Ainsi, puisque

$$P_\theta(]a, b]) = F_\theta(b) - F_\theta(a) = \begin{cases} 1 - (1 - \mathrm{e}^{-\theta a}) = \mathrm{e}^{-\theta a} & \text{si } b = 1 \\ \mathrm{e}^{-\theta a} - \mathrm{e}^{-\theta b} & \text{si } b < 1, \end{cases}$$

il vient $P_\theta(]a, b]) = \mathrm{e}^{-\theta a} - \mathrm{e}^{-\theta a} = 0$.
Soit $\theta > 0$. On cherche à calculer une densité $f_\theta = dP_\theta/d\mu$. Sur $[0, 1[$, la mesure dominante de P_θ est $\lambda_{[0,1[}$ et F_θ est continue et dérivable. La densité f_θ est donc égale à F'_θ :

$$\forall y \in [0, 1[, \quad f_\theta(y) = \theta e^{-\theta y}.$$

De plus,

$$f_\theta(1) = \mathbb{P}(Y_1 = 1) = e^{-\theta}$$

et $f_\theta(y) = 0$ pour tout $y \notin [0, 1]$.
On vérifie alors aisément que pour tous $0 \leq a \leq b \leq 1$, $\int_{]a,b]} f_\theta(u)\mu(\,\mathrm{d}u) = \int_{]a,b]} f_\theta(u)\lambda_{[0,1[}(\,\mathrm{d}u) + \int_{]a,b]} f_\theta(u)\delta_1(\,\mathrm{d}u) = \int_{]a,b]} f_\theta(u)\,\mathrm{d}u + \mathrm{e}^{-\theta}\mathbb{1}_{1\in]a,b]} = F_\theta(b) - F_\theta(a) = P_\theta(]a, b])$.
<u>Remarque :</u> $P_{\theta^\star}$ n'est pas absolument continue par rapport à $\lambda_{[0,1[}$. En effet, si cela était le cas, nous aurions :

$$e^{-\theta^\star} = P_{\theta^\star}(\{1\}) = f_{\theta^\star}(1)\lambda_{[0,1[}(\{1\}) = 0.$$

3. La log-vraisemblance ℓ_n de l'échantillon $(Y_1, \ldots, Y_n)$ est donnée, pour tout $\theta > 0$, par :

$$\ell_n(\theta) = \sum_{i=1}^{n} \ln(f_\theta(Y_i)) = \sum_{\substack{1\leq i\leq n \\ Y_i<1}} (\ln(\theta) - \theta Y_i) + \sum_{\substack{1\leq i\leq n \\ Y_i=1}} (-\theta) = (n - Z_n)\ln(\theta) - \theta \sum_{i=1}^{n} Y_i,$$

où $Z_n = \sum_{i=1}^n \mathbb{1}_{Y_i=1}$ est le nombre d'observations égales à 1. ℓ_n est dérivable et nous avons :

$$\ell'_n(\theta) = \frac{n - Z_n}{\theta} - \sum_{i=1}^{n} Y_i > 0 \iff \theta < \frac{n - Z_n}{\sum_{i=1}^n Y_i}.$$

Ceci assure que ℓ_n atteint son maximum en $\widehat{\theta}_n = \frac{n-Z_n}{\sum_{i=1}^n Y_i}$. Puisque pour tout $i \in \{1, \ldots, n\}$, $Y_i > 0$ p.s., $\widehat{\theta}_n$ est bien défini. Attention : nous avons $\mathbb{P}(\widehat{\theta}_n = 0) = \mathbb{P}(Y_1 = \cdots = Y_n = 1) = e^{-n\theta^\star} > 0$. Dans ce cas, l'EMV n'existe pas. Dans le cas contraire, $\widehat{\theta}_n$ est l'EMV de $\theta^\star$.

4. On remarque tout d'abord que
$$\widehat{\theta}_n = \frac{n}{\sum_{i=1}^n Y_i}\left(1 - \frac{Z_n}{n}\right),$$
ce qui nous incite à appliquer la LGN à $\frac{\sum_{i=1}^n Y_i}{n}$. Pour ce faire, calculons l'espérance de Y_1 :
$$\begin{aligned}\mathbb{E}[Y_1] &= \int_0^1 y f_{\theta^\star}(y)\,dy + \mathbb{P}(Y_1 = 1)\\ &= \int_0^1 y\theta^\star e^{-\theta^\star y}\,dy + e^{-\theta^\star}\\ &= [-ye^{-\theta^\star y}]_0^1 + \int_0^1 e^{-\theta^\star y}\,dy + e^{-\theta^\star}\\ &= -e^{-\theta^\star} + \left[-\frac{1}{\theta^\star}e^{-\theta^\star y}\right]_0^1 + e^{-\theta^\star}\\ &= \frac{1-e^{-\theta^\star}}{\theta^\star}.\end{aligned}$$
Par la LFGN ($Y_1, \ldots, Y_n$ i.i.d. et $\mathbb{E}[Y_1] < \infty$) et le théorème de continuité ($x \in \mathbb{R}_+^* \mapsto 1/x$ continue en $\mathbb{E}[Y_1] \neq 0$),
$$\frac{n}{\sum_{i=1}^n Y_i} \overset{p.s.}{\to} \frac{1}{\mathbb{E}[Y_1]} = \frac{\theta^\star}{1-e^{-\theta^\star}}.$$
De plus, $\frac{Z_n}{n} = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{Y_i=1}$, où $(\mathbb{1}_{Y_1=1}, \ldots, \mathbb{1}_{Y_n=1})$ sont des variables i.i.d. de Bernoulli de paramètre $p = \mathbb{P}(Y_1 = 1) = e^{-\theta^\star}$. Ainsi, par la LFGN
$$\frac{Z_n}{n} \overset{p.s.}{\to} \mathbb{E}[\mathbb{1}_{Y_1=1}] = p = e^{-\theta^\star}.$$
On en conclue que $\widehat{\theta}_n \overset{p.s.}{\to} \theta^\star$ et a fortiori que $\widehat{\theta}_n \overset{\mathbb{P}}{\to} \theta^\star$ (i.e. $\widehat{\theta}_n$ est un estimateur consistant de $\theta^\star$).

5. L'idée est évidemment d'appliquer le TCL et la méthode delta multivariés à $U_i = \begin{pmatrix}\mathbb{1}_{Y_i=1}\\ Y_i\end{pmatrix}$, $i \in [\![1, n]\!]$ (qui sont bien des vecteurs aléatoires i.i.d) et de remarquer que $\widehat{\theta}_n = \varphi(\overline{U}_n)$, avec
$$\varphi : (x, y) \in \mathbb{R} \times \mathbb{R}_+^* \mapsto \frac{1-x}{y}.$$
Montrons d'abord que U_1 est de carré intégrable. Pour ce faire, il suffit de s'assurer que les deux lois marginales le sont. Or, une double intégration par partie, nous assure que
$$\mathbb{E}[Y_1^2] = 2\left(\frac{1}{\theta^{\star 2}} - \frac{e^{-\theta^\star}}{\theta^\star} - \frac{e^{-\theta^\star}}{\theta^{\star 2}}\right),$$
d'où
$$\mathrm{Var}(Y_1) = \frac{1}{\theta^{\star 2}} - \frac{e^{-\theta^\star}}{\theta^\star}\left(\frac{e^{-\theta^\star}}{\theta^\star} + 2\right).$$
Nous avons de plus
$$\mathrm{Var}(\mathbb{1}_{Y_1=1}) = e^{-\theta^\star}(1-e^{-\theta^\star}).$$
Nous sommes donc en mesure d'appliquer le TCL multivarié. Il reste cependant à déterminer la covariance entre les deux composantes :
$$\mathrm{Cov}(\mathbb{1}_{Y_1=1}Y_1) = \mathbb{E}[\mathbb{1}_{Y_1=1}Y_1] - \mathbb{E}[\mathbb{1}_{Y_1=1}]\mathbb{E}[Y_1] = e^{-\theta^\star}\left(1 - \frac{1-e^{-\theta^\star}}{\theta^\star}\right).$$

Afin d'appliquer la méthode delta, nous remarquons que φ admet bien des dérivées partielles sur $\mathbb{R} \times \mathbb{R}_+^*$:

$$\frac{\partial \varphi}{\partial x}(x, y) = -\frac{1}{y},$$

et

$$\frac{\partial \varphi}{\partial y}(x, y) = -\frac{1-x}{y},$$

qui sont toutes deux continues. φ est donc bien différentiable (au sens de Fréchet). Par ailleurs,

$$\frac{\partial \varphi}{\partial x}\left(e^{-\theta^\star}, \frac{1-e^{-\theta^\star}}{\theta^\star}\right) = -\frac{\theta^\star}{1-e^{-\theta^\star}} \neq 0,$$

et

$$\frac{\partial \varphi}{\partial y}\left(e^{-\theta^\star}, \frac{1-e^{-\theta^\star}}{\theta^\star}\right) = -\frac{\theta^{\star 2}}{1-e^{-\theta^\star}} \neq 0,$$

donc $\nabla\varphi(\mathbb{E}[U_1]) \neq 0$. Ainsi, en notant σ^2 la variance asymptotique cherchée, il vient

$$\begin{aligned}
\sigma^2 &= \nabla\varphi(\mathbb{E}[U_1])^\top \mathrm{Var}(U_1) \nabla\varphi(\mathbb{E}[U_1]) \\
&= \mathrm{Var}(U_{(1)}) \frac{\partial \varphi}{\partial x}(\mathbb{E}[U_1])^2 + \mathrm{Var}(U_{(2)}) \frac{\partial \varphi}{\partial y}(\mathbb{E}[U_1])^2 + 2\mathrm{Cov}(U_{(1)}, U_{(2)}) \frac{\partial \varphi}{\partial x}(\mathbb{E}[U_1]) \frac{\partial \varphi}{\partial y}(\mathbb{E}[U_1]).
\end{aligned}$$

Calculons donc les trois termes :

$$\mathrm{Var}(\mathbb{1}_{Y_1=1}) \frac{\partial \varphi}{\partial x}\left(e^{-\theta^\star}, \frac{1-e^{-\theta^\star}}{\theta^\star}\right)^2 = \frac{\theta^{\star 2} e^{-\theta^\star}}{1-e^{-\theta^\star}},$$

$$\mathrm{Var}(Y_1) \frac{\partial \varphi}{\partial y}\left(e^{-\theta^\star}, \frac{1-e^{-\theta^\star}}{\theta^\star}\right)^2 = \theta^{\star 2} \frac{1+e^{-\theta^\star}}{1-e^{-\theta^\star}} - 2\frac{\theta^{\star 3} e^{-\theta^\star}}{(1-e^{-\theta^\star})^2},$$

et

$$2\mathrm{Cov}(\mathbb{1}_{Y_1=1} Y_1) \frac{\partial \varphi}{\partial x}\left(e^{-\theta^\star}, \frac{1-e^{-\theta^\star}}{\theta^\star}\right) \frac{\partial \varphi}{\partial y}\left(e^{-\theta^\star}, \frac{1-e^{-\theta^\star}}{\theta^\star}\right) = 2\frac{\theta^{\star 3} e^{-\theta^\star}}{(1-e^{-\theta^\star})^2} - 2\frac{\theta^{\star 2} e^{-\theta^\star}}{1-e^{-\theta^\star}}.$$

En sommant l'ensemble, on obtient

$$\sigma^2 = -\frac{\theta^{\star 2} e^{-\theta^\star}}{1-e^{-\theta^\star}} + \theta^{\star 2}\frac{1+e^{-\theta^\star}}{1-e^{-\theta^\star}} = \frac{\theta^{\star 2}}{1-e^{-\theta^\star}}.$$

Par conséquent,

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

6. Pour tout $y \in]0,1[$, la fonction $\theta \longmapsto \theta e^{-\theta y}$ est de classe C^1 et pour $y = 1$ la fonction $\theta \longmapsto e^{-\theta}$ est elle aussi de classe C^1. Ainsi, pour μ presque tout y, la fonction $\theta \longmapsto f_\theta(y)$ est de classe C^1. De plus, on a, en prenant Y de densité f_θ,

$$\begin{aligned}
I(\theta) = \mathbb{E}\left[\left(\frac{\mathrm{d}(\log \circ f_\theta)}{\mathrm{d}\theta}(Y)\right)^2\right] &= \mathbb{E}\left[\left(\frac{\mathbf{1}_{Y\neq 1}}{\theta} - Y\right)^2\right] \\
&= \mathbb{E}\left[Y^2\right] - \frac{2}{\theta}\mathbb{E}\left[Y\mathbf{1}_{Y\neq 1}\right] + \frac{1}{\theta^2}\mathbb{E}\left[\mathbf{1}_{Y\neq 1}\right] \\
&= \int_0^1 y^2 \theta e^{-\theta y} dy + e^{-\theta} - 2\int_0^1 y e^{-\theta y} dy + \frac{1-e^{-\theta}}{\theta^2}.
\end{aligned}$$

A l'aide d'une intégration par partie, on obtient donc

$$I(\theta) = \left[y^2 e^{-\theta y}\right]_0^1 + 2\int_0^1 y e^{-\theta y} dy + e^{-\theta y} - 2\int_0^1 y e^{-\theta y} dy + \frac{1-e^{-\theta}}{\theta^2} = \frac{1-e^{-\theta}}{\theta^2}.$$

De plus, l'application $\theta \mapsto I(\theta)$ est continue sur $\mathbb{R}_+^*$ et le modèle est donc régulier. Puisque $I(\theta^\star) > 0$ et $I(\theta^\star)^{-1} = \frac{\theta^{\star 2}}{1-e^{-\theta^\star}}$, qui est la variance asymptotique de $\widehat{\theta}_n$, on en déduit que cet estimateur est asymptotiquement efficace.

Exercice 6

1. Ici, le modèle statistique considéré est $\mathcal{P} = \{f_\theta \cdot \lambda_{\mathbb{R}}, \theta \in \mathbb{R}\}$, qui est dominé par la mesure de Lebesgue sur $\mathbb{R}$, notée $\lambda_{\mathbb{R}}$, et chaque loi du modèle indicée par θ a pour densité f_θ. La vraisemblance s'écrit

$$L_n(\theta) = \prod_{i=1}^n \mathbf{1}_{\theta \le X_i \le \theta+1} = \mathbf{1}_{\forall i, \theta \le X_i \le \theta+1} = \mathbf{1}_{X_{(n)}-1 \le \theta \le X_{(1)}},$$

donc tout élément de l'intervalle $[X_{(n)} - 1, X_{(1)}]$ est un EMV.

2. **Elégant mais subtil.** Remarquons que le modèle est un modèle de translation issue d'une loi uniforme sur $[0,1]$, ainsi $X_i = \theta + Z_i$, avec $Z_1, \ldots, Z_n$ i.i.d. uniformes sur $[0,1]$. En notant $\mathcal{L}$ la loi d'une variable aléatoire, comme $Z_i \sim 1 - Z_i$, on a $Z_{(1)} \sim 1 - Z_{(n)}$ d'où

$$\mathcal{L}(n(X_{(1)} - \theta)) = \mathcal{L}(nZ_{(1)}) = \mathcal{L}(n(1 - Z_{(n)})) = \mathcal{L}(n(\theta + 1 - X_{(n)})),$$

et il suffit de calculer $\mathcal{L}(nZ_{(1)})$. On utilise pour cela la fonction de répartition, pour tout $t \in [0,n]$,

$$\mathbb{P}(nZ_{(1)} \le t) = 1 - \mathbb{P}(Z_{(1)} > t/n) = 1 - (1 - t/n)^n,$$

et la fonction de répartition vaut 0 pour $t < 0$ et 1 pour $t > n$. A la limite,

$$\mathbb{P}(nZ_{(1)} \le t) \longrightarrow (1 - e^{-t})\mathbb{1}_{[0,+\infty[}(t),$$

fonction de répartition d'une exponentielle $\mathcal{E}(1)$.

Version plus soft. On calcule la fonction de répartition de $n(X_{(1)} - \theta)$ et on étudie sa convergence simple. On trouve les mêmes résultats pour la fonction de répartition $n(\theta + 1 - X_{(n)})$.

3. La formule résulte du Théorème de Fubini-Tonelli (tout est positif) :

Version double intégrale :

$$\begin{aligned}\int_0^\infty t^{k-1}\mathbb{P}(Y > t)dt &= \int_0^\infty t^{k-1}\mathbb{E}[\mathbf{1}_{Y>t}]dt = \int_0^\infty t^{k-1}\int_\Omega \mathbb{1}_{Y>t}(\omega)d\mathbb{P}(\omega)dt \\ &= \int_\Omega \left(\int_0^{Y(\omega)} t^{k-1}dt\right) d\mathbb{P}(\omega) = \mathbb{E}\left[\int_0^Y t^{k-1}dt\right] \\ &= \mathbb{E}[Y^k/k].\end{aligned}$$

Version Fubini entre $\int$ et $\mathbb{E}$:

$$\begin{aligned}k\int_0^\infty t^{k-1}\mathbb{P}(Y > t)dt &= k\int_0^\infty t^{k-1}\mathbb{E}[\mathbf{1}_{Y>t}]dt = k\mathbb{E}\left[\int_0^\infty t^{k-1}\mathbf{1}_{Y>t}dt\right] \\ &= k\mathbb{E}\left[\int_0^Y t^{k-1}dt\right] = \mathbb{E}[Y^k].\end{aligned}$$

Ainsi, nous avons pour $k = 1$,

$$\mathbb{E}(n(X_{(1)} - \theta)) = \int_0^\infty \mathbb{P}(n(X_{(1)} - \theta) > t)dt = \int_0^n (1 - t/n)^n dt = \frac{n}{n+1},$$

et pour $k = 2$, en utilisant une intégration par parties,

$$\begin{aligned}\mathbb{E}(n^2(X_{(1)} - \theta)^2) &= 2\int_0^n t(1 - t/n)^n dt \\ &= 2\Big[- \frac{n}{n+1}(1 - t/n)^{n+1} t\Big]_0^n + 2\frac{n}{n+1}\int_0^n (1 - t/n)^{n+1} dt \\ &= \frac{2n^2}{(n+2)(n+1)}.\end{aligned}$$

4. On a vu que tout EMV $\widehat{\theta}_n$ vérifie

$$X_{(n)} - 1 \leq \widehat{\theta}_n \leq X_{(1)},$$

d'où

$$X_{(n)} - 1 - \theta \leq \widehat{\theta}_n - \theta \leq X_{(1)} - \theta.$$

Par conséquent,

$$(\widehat{\theta}_n - \theta)^2 \leq (X_{(1)} - \theta)^2 + (X_{(n)} - 1 - \theta)^2,$$

donc les questions précédentes donnent

$$\mathbb{E}((\widehat{\theta}_n - \theta)^2) \leq 2\mathbb{E}((X_{(1)} - \theta)^2) = \frac{4}{(n+2)(n+1)},$$

qui tend vers 0. Donc $\widehat{\theta}_n$ converge en moyenne quadratique vers θ, donc aussi en probabilité. Finalement, tout EMV est consistant.

5. La question précédente donne directement que, pour tout n,

$$n^2 R(\widehat{\theta}_n, \theta) \leq \frac{4n^2}{(n+2)(n+1)} \leq 4.$$

6. Choisissons $\widehat{\theta} = X_{(1)}$ comme EMV pour estimer θ. Nous avons vu que, pour tout $t \in [0, n]$,

$$\begin{aligned}&\mathbb{P}(\theta \leq \widehat{\theta} \leq \theta + t/n) = 1 - (1 - t/n)^n \\ &\iff \mathbb{P}(-\widehat{\theta} \leq -\theta \leq -\widehat{\theta} + t/n) = 1 - (1 - t/n)^n \\ &\iff \mathbb{P}(\widehat{\theta} - t/n \leq \theta \leq \widehat{\theta}) = 1 - (1 - t/n)^n\end{aligned}$$

Donc en choisissant $(1 - t/n)^n = \alpha$, on en déduit $t = n(1 - \alpha^{1/n})$, on obtient

$$\mathbb{P}(\widehat{\theta} - (1 - \alpha^{1/n}) \leq \theta \leq \widehat{\theta}) = 1 - \alpha,$$

ce qui fournit l'intervalle de confiance $I = [\widehat{\theta} - (1 - \alpha^{1/n}), \widehat{\theta}]$ pour θ de niveau $1 - \alpha$. Par exemple, pour $\alpha = 0.05$, $\widehat{\theta} = 3$, on obtient pour $n = 10$, $I \simeq [2.74, 3]$ et pour $n = 100$, $I \simeq [2.97, 3]$.

Exercice 7

1. Notons $P_{(\lambda,\theta)}$ la loi de X_1. Le modèle est identifiable si et seulement si l'application
$$(\lambda,\theta) \mapsto P_{(\lambda,\theta)}$$
est injective. Pour $\theta = 1$, on voit que $P_{(\lambda,\theta)}$ est une loi uniforme, indépendamment de λ, donc le modèle n'est pas identifiable. Si on exclut $\theta = 1$, alors montrons que le modèle est identifiable : soit (λ,θ) et (λ',θ') dans $[0,1[\times]1,\infty[$ tels que $P_{(\lambda,\theta)} = P_{(\lambda',\theta')}$. Alors $f_{\lambda,\theta}(x) = f_{\lambda',\theta'}(x)$ pour μ-presque tout x (μ étant la mesure de Lebesgue), ce qui implique que $\lambda + (1-\lambda)/\theta = \lambda' + (1-\lambda')/\theta'$ et $(1-\lambda)/\theta = (1-\lambda')/\theta'$ (car $\mu([0,1]) > 0$ et $\mu(]1,\theta]) > 0$). Ceci implique $(\lambda,\theta) = (\lambda',\theta')$. Donc le modèle est identifiable si on prend comme espace de paramètres $[0,1[\times]1,\infty[$ (c'est-à-dire en excluant $\theta = 1$).
2. Soit $x \in \mathbb{R}$, alors
$$\begin{aligned} f_{\theta,\lambda}(x) &= \left(\lambda + \frac{1-\lambda}{\theta}\right) \mathbf{1}_{0\le x\le 1} + \left(\frac{1-\lambda}{\theta}\right) \mathbf{1}_{1<x\le\theta} \\ &= \left(\lambda + \frac{1-\lambda}{\theta}\right)^{\mathbf{1}_{0\le x\le 1}} \left(\frac{1-\lambda}{\theta}\right)^{\mathbf{1}_{1<x\le\theta}} \mathbf{1}_{0\le x\le\theta} \end{aligned}$$
3. Nous pouvons alors écrire la vraisemblance comme suit :
$$\begin{aligned} L_n(\theta,\lambda) &= \prod_{i=1}^n \left(\lambda + \frac{1-\lambda}{\theta}\right)^{\mathbf{1}_{0\le X_i\le 1}} \left(\frac{1-\lambda}{\theta}\right)^{\mathbf{1}_{1<X_i\le\theta}} \mathbf{1}_{0\le X_i\le\theta} \\ &= \left(\lambda + \frac{1-\lambda}{\theta}\right)^{\sum_{i=1}^n \mathbf{1}_{0\le X_i\le 1}} \left(\frac{1-\lambda}{\theta}\right)^{\sum_{i=1}^n \mathbf{1}_{1<X_i\le\theta}} \mathbf{1}_{0\le X_{(n)}\le\theta} \\ &= \left(\lambda + \frac{1-\lambda}{\theta}\right)^{N} \left(\frac{1-\lambda}{\theta}\right)^{n-N} \mathbf{1}_{X_{(n)}\le\theta}, \end{aligned}$$
car $X_i \ge 0$ (p.s.) et en posant
$$N = \sum_{i=1}^n \mathbf{1}_{0\le X_i\le 1}.$$
Considérons maintenant la function g définie par
$$\begin{aligned} g(\theta,\lambda) &= \left(\lambda + \frac{1-\lambda}{\theta}\right)^{N} \left(\frac{1-\lambda}{\theta}\right)^{n-N} \\ &= \left(1 + \frac{\lambda\theta}{1-\lambda}\right)^{N} \left(\frac{1-\lambda}{\theta}\right)^{n}. \end{aligned}$$
Lorsque θ est connu et λ est inconnu, on cherche à maximiser $g(\lambda,\theta)$ en λ. La dérivée de $\ln g(\lambda,\theta) = n\ln(1-\lambda)/\theta) + N\ln\left(1 + \frac{\lambda\theta}{1-\lambda}\right)$ en λ est :
$$\frac{1}{1-\lambda}\left(\frac{N\theta}{1+\lambda(\theta-1)} - n\right),$$
qui est positive si et seulement si $\lambda \le \frac{\theta N/n - 1}{\theta - 1}$. Ainsi, l'EMV en λ (et à θ connu) est
$$\widehat{\lambda}(\theta) = \frac{\theta N/n - 1}{\theta - 1}.$$
Comme par la LFGN (les v.a. X_i sont i.i.d. donc les $\mathbf{1}_{0\le X_i\le 1}$ aussi, et elles sont intégrables car bornées) N/n tend vers $\lambda + (1-\lambda)/\theta$, on a $\widehat{\lambda}(\theta)$ qui tend p.s. vers λ donc $\widehat{\lambda}(\theta) \in [0,1[$ à partir

d'un certain rang, ce qui indique que l'EMV est défini p.s. à partir d'un certain rang et est bien consistant.

A présent, si λ est connu et θ est inconnu, on remarque que la fonction $g(\lambda, \theta)$ est décroissante en θ, ce qui donne directement que, lorsque $X_{(n)} > 1$, la vraisemblance atteint son maximum sur $]1, +\infty[$ en

$$\widehat{\theta} = X_{(n)}.$$

Si $X_{(n)} \leq 1$, le maximum n'existe pas sur $]1, +\infty[$ (car 1 est exclu).

Pour déterminer maintenant l'EMV lorsque les deux paramètres λ et θ sont inconnus, on remarque que, lorsque $X_{(n)} > 1$, la vraisemblance est toujours plus petite que sa valeur en $\widehat{\theta}$, c'est-à-dire $g(\lambda, \widehat{\theta})$. Comme $\widehat{\theta}$ ne dépend pas de λ, le même calcul que plus haut nous donne que la vraisemblance est maximum en $(\widehat{\lambda}(\widehat{\theta}), \widehat{\theta})$. Dans ce cas l'EMV pour (λ, θ) est donc

$$(\widehat{\lambda}(\widehat{\theta}), \widehat{\theta}) = \left(\frac{X_{(n)} \times N/n - 1}{X_{(n)} - 1}, X_{(n)} \right),$$

qui est bien défini sur l'événement où $X_{(n)} > 1$ et $\frac{X_{(n)}N/n-1}{X_{(n)}-1} \in [0, 1[$. On voit que cet événement à une probabilité qui tend vers 1, ce qui permet de faire les études asymptotiques qui vont suivre.

4. Nous obtenons facilement que, pour tout $x \in \mathbb{R}$,

$$\mathbb{P}_{\lambda,\theta}(X_1 \leq x) = \begin{cases} 0 & \text{si } x \leq 0 \\ (\lambda + (1-\lambda)/\theta)x & \text{si } 0 < x < 1 \\ \lambda + x(1-\lambda)/\theta & \text{si } 1 \leq x \leq \theta \\ 1 & \text{si } x > \theta \end{cases}$$

D'où, avec la méthode habituelle, on obtient pour tout $t \geq 0$

$$\mathbb{P}_{\lambda,\theta}(n(\theta - \widehat{\theta}) \leq t) = 1 - \exp\left(n \ln \left(1 - \frac{t(1-\lambda)}{\theta n} \right) \right)$$
$$\xrightarrow[n\to\infty]{} 1 - \exp\left(-\frac{t(1-\lambda)}{\theta} \right),$$

ce qui donne

$$n(\theta - \widehat{\theta}) \rightsquigarrow \mathcal{E}((1-\lambda)/\theta).$$

En particulier, $\widehat{\theta}$ est consistant.

5. Si θ est connu, on a déjà montré en question 3 que $\widehat{\lambda}(\theta)$ est consistant (converge en probabilité vers λ). Lorsque θ est inconnu, l'EMV est

$$\frac{X_{(n)} \times N/n - 1}{X_{(n)} - 1} \xrightarrow{P} \frac{\theta(\lambda + (1-\lambda)/\theta) - 1}{\theta - 1} = \lambda$$

en utilisant le lemme de Slutsky, notamment en utilisant le fait que $\theta > 1$. Lorsque $\theta = 1$, les X_i suivent la loi uniforme sur $[0, 1]$ et on a

$$\widehat{\lambda}(\widehat{\theta}) = \frac{X_{(n)} \times N/n - 1}{X_{(n)} - 1} = \frac{X_{(n)} - 1}{X_{(n)} - 1} = 1$$

car $N = n$. Donc l'estimateur $\widehat{\lambda}(\widehat{\theta})$ ne converge plus vers λ. Ceci s'explique car le modèle n'est plus identifiable en λ lorsque $\theta = 1$.

6. On cherche $c = c_n(\alpha)$ tel que $\mathbb{P}_{\theta=1,\lambda}(\widehat{\theta} > c) = \alpha$. On remarque que c est nécessairement plus petit que 1. D'après la question 3, on sait que

$$\mathbb{P}_{\theta=1,\lambda}(\widehat{\theta} > c) = 1 - \mathbb{P}_{\theta=1,\lambda}(\widehat{\theta} \leq c) = 1 - \mathbb{P}^n_{\theta=1,\lambda}(X_1 \leq c) = 1 - c^n = \alpha.$$

On choisit donc $c = (1-\alpha)^{1/n}$. Alors la puissance de ce test est donnée par la formule suivante : pour tout $\theta > 1$, pour tout λ,

$$\begin{aligned}\mathbb{P}_{H_1}(\text{rejet de } H_0) = \mathbb{P}_{\theta,\lambda}(\widehat{\theta} > (1-\alpha)^{1/n}) &= 1 - \mathbb{P}^n_{\theta,\lambda}(X_1 \leq (1-\alpha)^{1/n}) \\ &= 1 - \left(\left(\lambda + \frac{1-\lambda}{\theta}\right)(1-\alpha)^{1/n}\right)^n \\ &= 1 - (1-\alpha)\left(\lambda + \frac{1-\lambda}{\theta}\right)^n.\end{aligned}$$

Cette dernière quantité tend vers 1 lorsque n tend vers l'infini, ce qui indique que le test est consistant.

Exercice 8

1. Dans cette expérience, on n'observe que x, le nombre de poissons marqués parmi les n pêchés. Une modélisation probabiliste vise donc à considérer x comme une réalisation d'une variable aléatoire X et à définir la loi de cette dernière.

   Ici, $x = \sum_{i=1}^n y_i$, où $y_i = 1$ lorsque le i^e poisson pêché est marqué et 0 sinon. Puisque les poissons sont pêchés au hasard et avec remise, on peut modéliser $y_1, \ldots, y_n$ comme étant des réalisations de n variables aléatoires $Y_1, \ldots, Y_n$ indépendantes et identiquement distribuées de loi de Bernoulli $\mathcal{B}(p)$, avec p la probabilité de pêcher un poisson marqué, qui vaut ici $p = \frac{k}{N}$.

   Ainsi, $X \stackrel{\mathcal{L}}{=} \sum_{i=1}^n Y_i \sim \mathcal{B}\left(n, \frac{k}{N}\right)$. Soit donc le modèle statistique $\mathcal{P} = \left\{\mathcal{B}(n, \frac{k}{M}), M \in [k, +\infty[\right\}$, qui est bien indicé par un intervalle. Il découle de ce choix qu'il sera aisé d'obtenir un estimateur par maximum de vraisemblance de N mais en contrepartie que celui-ci ne sera pas entier (ce qui compliquerait grandement l'analyse).

2. Le modèle $\mathcal{P}$ est dominé par la mesure de comptage sur $[\![0, n]\!]$ et chaque loi candidate de paramètre $M \in [k, +\infty[$ possède une densite $x \in [\![0, n]\!] \mapsto \binom{n}{x}\left(\frac{k}{M}\right)^x\left(1 - \frac{k}{M}\right)^{n-x}$. Ainsi, pour tout $M \in [k, +\infty[$, la vraisemblance de M vis-à-vis de X est :

$$L(M) = \binom{n}{X}\left(\frac{k}{M}\right)^X\left(1 - \frac{k}{M}\right)^{n-X}.$$

Sur l'événement $\{X = 0\}$ Pour tout $M \in [k, +\infty[$,

$$L(M) = \left(1 - \frac{k}{M}\right)^n.$$

L est une fonction croissante de M et n'atteint donc pas son maximum.

Sur l'événement $\{X = n\}$ Pour tout $M \in [k, +\infty[$,

$$L(M) = \left(\frac{k}{M}\right)^n.$$

L est une fonction décroissante de M et atteint donc son maximum en $M = k$.

Sur l'événement $\{0 < X < n\}$: La vraisemblance est nulle en $M = k$ et pour tout $M \in]k, +\infty[$, on peut définir la log-vraisemblance par :

$$\begin{aligned}\ell(M) &= \ln\left(\binom{n}{X}\left(\frac{k}{M}\right)^X\left(1-\frac{k}{M}\right)^{n-X}\right)\\ &= C - X\ln M + (n-X)\ln(M-k) - (n-X)\ln(M)\\ &= C + (n-X)\ln(M-k) - n\ln(M),\end{aligned}$$

où C est une constante. Puisque $\lim_{M\to k^+}\ell(M) = \lim_{M\to+\infty}\ell(M) = -\infty$ (car $0 < X < n$) et ℓ est deux fois dérivable, on vérifie aisément qu'elle atteint son maximum sur $[k, +\infty[$ pour $M = \frac{kn}{X} \in]k, +\infty[$.
Ainsi, l'EMV de N n'est pas défini si $X = 0$ mais on choisit comme estimateur $\widehat{N} = \frac{kn}{X+\mathbb{1}_{X=0}}$, qui regroupe les deux derniers cas. $\widehat{N}$ est une variable aléatoire bien définie, y compris sur l'événement $X = 0$.
Remarque : on pourrait tout aussi bien choisir comme estimateur $\frac{kn}{X+1} = \frac{k}{\frac{X}{n}+\frac{1}{n}}$, l'important étant que le terme correctif, ici $\frac{1}{n}$, tende vers 0 en probabilité. De manière un peu plus légère et par simplicité, on aurait aussi pu prendre $\frac{kn}{X}$, variable aléatoire mal définie avec probabilité $\mathbb{P}(X=0) = \left(1-\frac{k}{N}\right)^n$, quantité qui est asymptotiquement nulle. Cela nous aurait suffit pour continuer l'analyse et éviter l'utilisation du lemme de Slutsky.

3. En notant $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$, nous avons $\widehat{N} \overset{\mathcal{L}}{=} \frac{k}{\overline{Y}_n+\frac{\mathbb{1}_{X=0}}{n}}$. Par la loi faible des grands nombres, $\overline{Y}_n \xrightarrow[n\to\infty]{\mathbb{P}} \frac{k}{N}$ et puisque $\frac{\mathbb{1}_{X=0}}{n} \xrightarrow[n\to\infty]{\mathbb{P}} 0$, il vient par continuité $\overline{Y}_n + \frac{\mathbb{1}_{X=0}}{n} \xrightarrow[n\to\infty]{\mathbb{P}} \frac{k}{N}$. Enfin, par continuité de la fonction $\varphi : x \mapsto \frac{k}{x}$ en $\frac{k}{N} > 0$, $\frac{k}{\overline{Y}_n+\frac{\mathbb{1}_{X=0}}{n}} \xrightarrow[n\to\infty]{\mathbb{P}} N$, résultat qui reste vrai pour $\widehat{N}$.
De même, en supposant $k < N$, le TCL indique que
$$\sqrt{n}\left(\overline{Y}_n - \frac{k}{N}\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{k}{N}\left(1-\frac{k}{N}\right)\right).$$
Puisque $\frac{\mathbb{1}_{X=0}}{\sqrt{n}} \xrightarrow[n\to\infty]{\mathbb{P}} 0$, il vient alors par le lemme de Slutsky
$$\sqrt{n}\left(\overline{Y}_n + \frac{\mathbb{1}_{X=0}}{n} - \frac{k}{N}\right) = \sqrt{n}\left(\overline{Y}_n - \frac{k}{N}\right) + \frac{\mathbb{1}_{X=0}}{\sqrt{n}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{k}{N}\left(1-\frac{k}{N}\right)\right).$$
Enfin, par la méthode delta appliquée avec φ, dérivable en $\frac{k}{N}$ et de dérivée non-nulle en ce point,
$$\sqrt{n}\left(\frac{k}{\overline{Y}_n + \frac{\mathbb{1}_{X=0}}{n}} - N\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, N^2\left(\frac{N}{k}-1\right)\right),$$
résultat qui reste vrai pour $\widehat{N}$.
4. La variance asymptotique étant d'autant plus petite que k est grand, il faudrait prendre k aussi grand que possible (il faut alors un grand chalut), c'est-à-dire $k = N$, ce qui nous assurerait un estimateur avec une variance (asymptotique) nulle !

Exercice 9

1. Par définition, l'estimateur des moindres carrés est la valeur de t qui minimise la quantité $\sum_{i=1}^n (X_i - t)^2$. La minimisation de ce trinôme en t donne bien la moyenne empirique $\overline{X}_n$.

2. Si $\overline{X}_n$ est l'EMV alors il maximise la log-vraisemblance

$$\ell_n(\theta) = \sum_{i=1}^{n} g(X_i - \theta).$$

Puisque celle-ci est C^1 sur $\mathbb{R}$, c'est en particulier un point critique donc nécessairement

$$\sum_{i=1}^{n} g'(X_i - \overline{X}_n) = 0.$$

En passant aux réalisations de l'échantillon et puisque $f > 0$ sur $\mathbb{R}$, ceci signifie que pour tout n-uplet $(x_1, \ldots, x_n)$ de réels, la fonction g doit vérifier

$$\sum_{i=1}^{n} g'(x_i - \overline{x}_n) = 0.$$

3. En prenant $x_2 = \cdots = x_n = x_1 - nt$, cette relation devient

$$g'((n-1)t) + (n-1)g'(-t) = 0.$$

Or f est paire donc $g'(-t) = -g(t)$ donc $g'((n-1)t) = (n-1)g'(t)$. Le réel $t \neq 0$ et l'entier $n > 1$ étant arbitraires, on a bien $\forall t \in \mathbb{R}^\star$, $\forall k \in \mathbb{N}^\star$, $\frac{g'(kt)}{kt} = \frac{g'(t)}{t}$.

4. Posons $c = g'(1)$. La relation précédente implique que $\forall k \in \mathbb{N}^\star$, $\frac{g'(k)}{k} = c$. Soit $x = p/q$ un rationnel strictement positif, alors

$$\frac{g'(x)}{x} = \frac{g'(qx)}{qx} = \frac{g'(p)}{p} = c.$$

La fonction $\varphi(x) = \frac{g'(x)}{x}$ étant continue sur $]0, \infty[$, la densité des rationnels dans $]0, \infty[$ assure que $\forall x > 0$, $\frac{g'(x)}{x} = c$. Par imparité de g', ceci est aussi vrai pour tout $x < 0$ et on peut donc prolonger par continuité la fonction φ en 0 par la même valeur. On a ainsi établi qu'il existe une constante c telle que, pour tout $t \in \mathbb{R}^\star$, $\frac{g'(t)}{t} = c$.

5. L'intégration de l'équation différentielle $g'(t) = ct$ nous dit qu'il existe $a \in \mathbb{R}$ telle que $g(t) = a + ct^2/2$ donc une constante $b = e^a > 0$ telle que $f(t) = be^{ct^2/2}$. Puisque f est une densité, elle est intégrable donc il est clair que $c < 0$ et on peut ainsi écrire c sous la forme $c = -1/\sigma^2$ de sorte que $f(t) = be^{-t^2/2\sigma^2}$. Comme de plus l'intégrale vaut 1, il vient $b = 1/\sqrt{2\pi\sigma^2}$ et f est la densité d'une loi normale centrée. La réciproque est claire : si les X_i sont i.i.d. de loi $\mathcal{N}(\theta, \sigma^2)$, alors un calcul facile montre que l'EMV est $\overline{X}_n$, qui est comme on l'a vu l'estimateur des moindres carrés.

Exercice 10

1. On sait que $\overline{X}_{100} \sim \mathcal{N}(\theta, 10^{-2})$, ce qui équivaut à dire que $10(\overline{X}_{100} - \theta) \sim \mathcal{N}(0, 1)$ donc

$$\mathbb{P}\left(-\Phi^{-1}(0.975) \leq 10(\overline{X}_{100} - \theta) \leq \Phi^{-1}(0.975)\right) = 0.95$$

ou encore, avec l'approximation $\Phi^{-1}(0.975) \approx 2$ et en notant $\widehat{I} = [\overline{X}_{100} - 0.2 \ ; \ \overline{X}_{100} + 0.2]$, que $\mathbb{P}(\theta \in \widehat{I}) \approx 0.95$.

2. Si $\theta = 0$, sachant $X_{100} = 50$, on peut écrire

$$\overline{X}_{100} = \frac{1}{100}\left(50 + \sum_{i=1}^{99} X_i\right) = \frac{1}{2} + \frac{1}{100}\sum_{i=1}^{99} X_i \sim \mathcal{N}\left(\frac{1}{2}, \frac{99}{10^4}\right).$$

Via l'approximation $99/10^4 \approx 10^{-2}$, on a donc $\mathcal{L}(\overline{X}_{100} \mid X_{100} = 50) = \mathcal{N}\left(\frac{1}{2}, \frac{1}{100}\right)$.

3. Par conséquent

$$P\left(0 \in \widehat{I} \mid X_{100} = 50\right) = \mathbb{P}\left(\overline{X}_{100} - 0.2 \leq 0 \leq \overline{X}_{100} + 0.2 \mid X_{100} = 50\right),$$

c'est-à-dire

$$P\left(0 \in \widehat{I} \mid X_{100} = 50\right) = \mathbb{P}\left(-0.2 \leq \overline{X}_{100} \leq 0.2 \mid X_{100} = 50\right),$$

et par ci-dessus

$$P\left(0 \in \widehat{I} \mid X_{100} = 50\right) = \mathbb{P}\left(-0.2 \leq \mathcal{N}\left(\frac{1}{2}, \frac{1}{100}\right) \leq 0.2\right),$$

ou encore, en centrant et normalisant,

$$P\left(0 \in \widehat{I} \mid X_{100} = 50\right) = \mathbb{P}\left(-7 \leq \mathcal{N}\left(0,1\right) \leq -3\right) = \Phi(-3) - \Phi(-7) \approx 10^{-3}.$$

Conclusion : à cause d'une seule donnée aberrante, l'intervalle de confiance à 95% est devenu un intervalle de confiance à 0,1% ! Ceci illustre la non-robustesse de l'EMV aux données aberrantes. En revanche, il est clair que l'estimateur de la médiane empirique n'aurait pas été sensible à cette donnée : il est plus robuste.

Exercice 11

1. Puisque

$$h(f,g) = \frac{1}{\sqrt{2}}\|\sqrt{f} - \sqrt{g}\|_2,$$

il est clair que h est une distance. La formule

$$h^2(f,g) = 1 - \int_{\mathbb{R}} \sqrt{f(x)g(x)}\, dx$$

découle du fait que f et g intègrent à 1. Une distance étant positive et le dernier terme étant positif, on en déduit aussi que $0 \leq h(f,g) \leq 1$.

2. Soit $\theta > 0$ et f_θ la densité de la loi uniforme sur $[0,\theta]$. Le fait que $h^2(f_\theta, f_{\theta'}) = 1 - \sqrt{\frac{\theta}{\theta'}}$ si $\theta \leq \theta'$ est une conséquence directe de la formule précédente.

3. Soit $X_1, \ldots, X_n$ i.i.d. de loi uniforme sur $[0,\theta]$ pour un $\theta > 0$ inconnu. La vraisemblance associée à cet échantillon s'écrit

$$L_n(\theta) = \frac{1}{\theta^n}\mathbf{1}_{\theta \geq X_{(n)}},$$

qui est maximale pour $\theta = \widehat{\theta}_n = X_{(n)}$. La variable $\sqrt{\widehat{\theta}_n}$ étant positive (et inférieure à $\sqrt{\theta}$), on peut écrire que

$$\mathbb{E}_\theta\left[\sqrt{\widehat{\theta}_n}\right] = \int_0^\infty \mathbb{P}\left(\sqrt{\widehat{\theta}_n} > x\right) dx = \int_0^{\sqrt{\theta}} \mathbb{P}\left(\sqrt{\widehat{\theta}_n} > x\right) dx,$$

d'où

$$\mathbb{E}_\theta\left[\sqrt{\widehat{\theta}_n}\right] = \int_0^{\sqrt{\theta}} \left(1 - \mathbb{P}\left(\sqrt{\widehat{\theta}_n} \leq x\right)\right) dx = \int_0^{\sqrt{\theta}} \left(1 - \left(\frac{x^2}{\theta}\right)^n\right) dx,$$

ce qui donne finalement

$$\mathbb{E}_\theta\left[\sqrt{\widehat{\theta}_n}\right] = \sqrt{\theta}\left(1 - \frac{1}{2n+1}\right).$$

Puisque presque sûrement $0 < \widehat{\theta}_n \leq \theta$, le calcul d'espérance par conditionnement et les résultats précédents donnent

$$\mathbb{E}_\theta\left[h^2(f_\theta, f_{\widehat{\theta}_n})\right] = \mathbb{E}_\theta\left[\mathbb{E}_\theta\left[h^2(f_\theta, f_{\widehat{\theta}_n})\middle|\widehat{\theta}_n\right]\right] = \mathbb{E}_\theta\left[1 - \sqrt{\frac{\widehat{\theta}_n}{\theta}}\right] = \frac{1}{2n+1}.$$

4. Soit la densité

$$f_n^*(x) = 10\left(1 - \frac{1}{n}\right)\mathbf{1}_{0\leq x\leq 1/10} + \frac{10}{n}\mathbf{1}_{9/10\leq x\leq 1}.$$

   (a) Pour toute fonction continue bornée φ, il est clair que

$$\mathbb{E}[\varphi(Y_n)] = 10\left(1 - \frac{1}{n}\right)\int_0^{1/10}\varphi(x)dx + \frac{10}{n}\int_{9/10}^1\varphi(x)dx \xrightarrow[n\to\infty]{} 10\int_0^{1/10}\varphi(x)dx,$$

   ce qui montre que (Y_n) converge en loi vers une loi uniforme sur $[0, 1/10]$.

   (b) Un calcul immédiat donne

$$h^2(f_n^*, f_{1/10}) = 1 - \int_{\mathbb{R}}\sqrt{f_n^*(x)f_{1/10}(x)}\,dx = 1 - \sqrt{1 - \frac{1}{n}},$$

   quantité qui tend bien vers 0 lorsque n tend vers l'infini.

5. Soit $n > 1$. Les variables $X_1, \ldots, X_n$ sont désormais i.i.d. de densité f_n^*, mais on les croit toujours i.i.d. suivant une densité uniforme $(f_\theta)_{\theta>0}$, avec θ inconnu. En particulier, l'estimateur $\widehat{\theta}_n$ est le même que celui défini ci-dessus.

   (a) Chaque X_i tombe indépendamment dans les intervalles $[0, 1/10]$ et $[9/10, 1]$ avec les probabilités respectives $1 - 1/n$ et $1/n$. Puisque $\widehat{\theta}_n = X_{(n)}$, la probabilité que $\widehat{\theta}_n$ soit entre $9/10$ et 1 est donc la probabilité que l'un au moins des X_i y soit, i.e.

$$\mathbb{P}\left(9/10 \leq \widehat{\theta}_n \leq 1\right) = 1 - \mathbb{P}\left(0 \leq X_1 \leq 1/10\right)^n = 1 - (1 - 1/n)^n.$$

   (b) Sur l'événement $\{9/10 \leq \widehat{\theta}_n \leq 1\}$, puisque $f_{\widehat{\theta}_n}(x) = \mathbf{1}_{[0,\widehat{\theta}_n]}(x)/\widehat{\theta}_n$, on obtient :

$$h^2(f_n^*, f_{\widehat{\theta}_n}) = 1 - \int_{\mathbb{R}}\sqrt{f_n^*(x)f_{\widehat{\theta}_n}(x)}\,dx = 1 - \sqrt{\frac{10}{\widehat{\theta}_n}}\left(\int_0^{1/10}\sqrt{1 - \frac{1}{n}}\,dx + \int_{9/10}^{\widehat{\theta}_n}\frac{1}{\sqrt{n}}\,dx\right),$$

   ce qui donne

$$h^2(f_n^*, f_{\widehat{\theta}_n}) = 1 - \frac{\sqrt{1 - \frac{1}{n}}}{\sqrt{10\widehat{\theta}_n}} - \sqrt{\frac{10}{n\widehat{\theta}_n}}\,(\widehat{\theta}_n - 9/10) \geq 1 - \frac{1}{3}\sqrt{1 - \frac{1}{n}} - \frac{1}{3\sqrt{n}},$$

   la dernière inégalité découlant du fait que $9/10 \leq \widehat{\theta}_n \leq 1$.

   (c) En conditionnant par rapport à la valeur de $\widehat{\theta}_n$ et en tenant compte du fait que la distance de Hellinger est positive, il s'ensuit que

$$\mathbb{E}_n^*[h^2(f_n^*, f_{\widehat{\theta}_n})] \geq \mathbb{E}_n^*[h^2(f_n^*, f_{\widehat{\theta}_n})|9/10 \leq \widehat{\theta}_n \leq 1]\mathbb{P}(9/10 \leq \widehat{\theta}_n \leq 1),$$

   d'où, par les calculs précédents,

$$\mathbb{E}_n^*[h^2(f_n^*, f_{\widehat{\theta}_n})] \geq u_n = \left(1 - \frac{1}{3}\sqrt{1 - \frac{1}{n}} - \frac{1}{3\sqrt{n}}\right) \times \left(1 - (1 - 1/n)^n\right),$$

   avec $\lim_{n\to\infty} u_n = 2/3 \times (1 - 1/e)$.

(d) Pour n grand, on a vu que la distance $h(f_n^*, f_{1/10})$ est arbitrairement petite. La vraie densité f_n^* n'appartient pas au modèle $(f_\theta)_{\theta>0}$ des densités uniformes : c'est un mélange de lois où, en moyenne sur n données, une seule est aberrante. Un "bon" estimateur de θ devrait être arbitrairement proche de $1/10$. La question précédente montre que ça n'est pas le cas pour l'estimateur du maximum de vraisemblance, qui n'est donc pas robuste. A contrario, sur cet exemple, il est facile de vérifier qu'un estimateur basé sur la médiane est robuste.