Dans ce sujet, par convention un vecteur x est un vecteur colonne. On notera x^T la transposée du vecteur (ou de la matrice) x.

Exercice 1. Soit $1 \leq p \leq n-2$ deux entiers naturels, et A une matrice de taille $n \times p$ de rang p, connue. Soit $(Y_1, \ldots, Y_n)$ une observation du modèle statistique :

$$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{\mathbb{P}_{\boldsymbol{\theta}} = p_{\boldsymbol{\theta}} \cdot \mathrm{Leb}^{\otimes n} \ : \ \theta = (\boldsymbol{\beta}, \sigma^2) \in \Theta := \mathbb{R}^p \times \mathbb{R}_+^*\})$$

où $\mathrm{Leb}^{\otimes n}$ est la mesure de Lebesgue sur $\mathbb{R}^n$ et $p_{\boldsymbol{\theta}}$ est une densité par rapport à $\mathrm{Leb}^{\otimes n}$ donnée par :

$$p_{\boldsymbol{\theta}}(\mathbf{y}) := \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - A\boldsymbol{\beta}\|^2\right); \qquad \boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2) \in \Theta, \quad \mathbf{y} := \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n$$

Nous notons

$$\mathbf{Y} := \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \in \mathbb{R}^n, \quad C := A^T A \in \mathbb{R}^{p \times p}, \qquad \mathrm{H} := AC^{-1}A^T \in \mathbb{R}^{n \times n}, \qquad \hat{\boldsymbol{\beta}} := C^{-1}A^T\mathbf{Y} \in \mathbb{R}^p.$$

Nous rappelons que

- H est le projecteur orthogonal sur $\mathrm{Im}(A)$, le sous-espace vectoriel de $\mathbb{R}^n$ engendré par les colonnes de la matrice A.
- $\hat{\boldsymbol{\beta}}$ est l'estimateur des moindre carrés, i.e. l'unique minimum de

$$\boldsymbol{\beta} \mapsto J(\boldsymbol{\beta}) := \|\mathbf{Y} - A\boldsymbol{\beta}\|^2.$$

Nous considérons maintenant un modèle dans lequel nous ajoutons un régresseur (i.e. "une variable explicative") $\mathbf{z} \in \mathbb{R}^{n \times 1}$ vérifiant $(\mathrm{I} - \mathrm{H})\mathbf{z} \neq \mathbf{0}_{n \times 1}$. Nous supposons maintenant que $(Y_1, \ldots, Y_n)$ est une observation du modèle statistique

$$\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{\mathbb{Q}_{\bar{\boldsymbol{\theta}}} = q_{\overline{\theta}} \cdot \mathrm{Leb}^{\otimes n} \ : \ \overline{\theta} = (\bar{\boldsymbol{\beta}}, \gamma, \bar{\sigma}^2) \in \bar{\Theta} := \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}_+^*\}\right)$$

où $q_{\bar{\boldsymbol{\theta}}}$ est une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$ donnée par :

$$q_{\bar{\boldsymbol{\theta}}}(\mathbf{y}) := \frac{1}{(2\pi\bar{\sigma}^2)^{n/2}} \exp\left(-\frac{1}{2\bar{\sigma}^2}\|\mathbf{y} - A\bar{\boldsymbol{\beta}} - \mathbf{z}\gamma\|^2\right); \qquad \overline{\theta} = (\bar{\boldsymbol{\beta}}, \gamma, \bar{\sigma}^2) \in \bar{\Theta}.$$

Nous posons

$$\mathbf{z}_\perp := (\mathrm{I} - \mathrm{H})\mathbf{z}, \quad V := [A\ ,\ \mathbf{z}_\perp]\ , \quad \mathbf{k}_z := C^{-1}A^T\mathbf{z}\ .$$

Remarquons que

$$A^T\mathbf{z}_\perp = \mathbf{0}_{p \times 1} \quad \text{et} \quad \mathbf{z} = A\mathbf{k}_z + \mathbf{z}_\perp.$$

1. Montrer que, pour tout $\bar{\boldsymbol{\theta}} = (\bar{\boldsymbol{\beta}}, \gamma, \bar{\sigma}^2) \in \bar{\Theta}$,

$$\mathbb{E}_{\bar{\boldsymbol{\theta}}}[\mathbf{Y}] = V \begin{bmatrix} \bar{\boldsymbol{\beta}} + \mathbf{k}_z\gamma \\ \gamma \end{bmatrix}\ .$$

2. Montrer que l'estimateur $\hat{\boldsymbol{\lambda}}$ des moindres carrés de $\boldsymbol{\lambda}$ défini par

$$\hat{\boldsymbol{\lambda}} = \underset{\boldsymbol{\lambda} \in \mathbb{R}^{p+1}}{\arg\min} \|\mathbf{Y} - V\boldsymbol{\lambda}\|^2$$

   est donné par

$$\hat{\boldsymbol{\lambda}} := \begin{bmatrix} C^{-1}A^T\mathbf{Y} \\ \|\mathbf{z}_\perp\|^{-2}\mathbf{z}_\perp^T\mathbf{Y} \end{bmatrix} .$$

3. Déterminer l'estimateur $(\hat{\bar{\boldsymbol{\beta}}}, \hat{\gamma})$ des moindres carrés de $(\bar{\boldsymbol{\beta}}, \gamma)$, défini par

$$(\hat{\bar{\boldsymbol{\beta}}}, \hat{\gamma}) = \underset{(\bar{\boldsymbol{\beta}}, \gamma) \in \mathbb{R}^p \times \mathbb{R}}{\arg\min} \|\mathbf{Y} - A\bar{\boldsymbol{\beta}} - \mathbf{z}\gamma\|^2$$

   en fonction de $\hat{\boldsymbol{\lambda}}$.

4. Déterminer la distribution de l'estimateur $\hat{\gamma}$ sous $\mathbb{Q}_{\bar{\boldsymbol{\theta}}}$ pour $\bar{\boldsymbol{\theta}} = (\bar{\boldsymbol{\beta}}, \gamma, \bar{\sigma}^2) \in \bar{\Theta}$.
5. Montrer que le projecteur orthogonal $\bar{\mathrm{H}}$ sur $\mathrm{Im}(V)$ est donné par

$$\bar{\mathrm{H}} := \mathrm{H} + \|\mathbf{z}_\perp\|^{-2}\mathbf{z}_\perp\mathbf{z}_\perp^T .$$

6. En déduire un estimateur $\hat{\bar{\sigma}}^2$ sans biais de $\bar{\sigma}^2$ [exprimer $\hat{\bar{\sigma}}^2$ en fonction de $\hat{\sigma}^2$ et $z_\perp^T\mathbf{Y}$].
7. Déterminer la distribution de $(n-p-1)\hat{\bar{\sigma}}^2/\bar{\sigma}^2$ sous $\mathbb{Q}_{\bar{\boldsymbol{\theta}}}$ pour $\bar{\boldsymbol{\theta}} = (\bar{\boldsymbol{\beta}}, \gamma, \bar{\sigma}^2) \in \bar{\Theta}$.
8. Construire un test $H_0 : \ \gamma = 0, \quad \text{contre} \quad H_1 : \ \gamma \neq 0$ de niveau $a \in]0,1[$.

Exercice 2. Soit $n \geq 3$ et $Z = (X_1, \ldots, X_n)$ un n-échantillon du modèle

$$(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), \{p_\theta \cdot \mathrm{Leb}, \theta \in \mathbb{R}_+^*\}) ,$$

où Leb est la mesure de Lebesgue sur $\mathbb{R}$ et p_θ est définie par

$$x \mapsto p_\theta(x) := \theta x^{\theta-1}\mathbb{1}_{]0,1[}(x) = \theta \mathrm{e}^{(\theta-1)\log(x)}\mathbb{1}_{]0,1[}(x) .$$

Nous admettrons que le modèle statistique est régulier.

1. Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ du paramètre θ.
2. Déterminer la distribution asymptotique de la suite d'estimateurs $\sqrt{n}(\hat{\theta}_n - \theta)$.
3. Montrer que sous p_θ.Leb, la loi de la variable $Y_1 := -\log(X_1)$ est Gamma$(1, \theta)$ (voir Définition IV-5.12 dans le polycopié). En déduire la loi, sous $p_\theta \cdot \mu$, de $\sum_{i=1}^n Y_i$ où $Y_i := -\log(X_i)$.
4. Calculer pour tout $\theta \in \mathbb{R}_+^\star$, $\mathbb{E}_\theta[\hat{\theta}_n]$ et $\mathrm{Var}_\theta(\hat{\theta}_n)$.
5. Montrer qu'il existe une suite réelle (déterministe) $\{a_n, \ n \in \mathbb{N}\}$ telle que $\tilde{\theta}_n := a_n\hat{\theta}_n$ soit un estimateur sans biais de θ.
6. L'estimateur $\tilde{\theta}_n$ est-il efficace ?
7. Montrer que $n\theta/\hat{\theta}_n$ est une fonction pivotale pour le paramètre θ dont on déterminera la distribution. En déduire un intervalle de confiance exact de niveau $1-\alpha$ pour $\alpha \in]0,1[$.

Exercice 3. Soit $Z = (X_1, \dots, X_n)$ un n-échantillon du modèle

$$(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), \{p_\theta \cdot \mathrm{Leb}, \theta = (\mu, \rho) \in \Theta := \mathbb{R} \times \mathbb{R}_+^*\}),$$

où Leb est la mesure de Lebesgue sur $\mathbb{R}$ et p_θ est définie par

$$x \mapsto p_\theta(x) := \frac{1}{\rho} \mathbb{1}_{[\mu-\rho/2, \mu+\rho/2]}(x) , \qquad \theta = (\mu, \rho) \in \Theta.$$

Dans la suite nous supposons que $n \geq 3$. Nous notons $X_{n:1} := \min(X_1, \dots, X_n)$, $X_{n:n} := \max(X_1, \dots, X_n)$ et $R_n := X_{n:n} - X_{n:1}$ et pour $\theta = (\mu, \rho) \in \Theta$,

$$A_\theta := \left\{(u, v) \in \mathbb{R}^2 \, : \, \mu - \rho/2 \leq u \leq v \leq \mu + \rho/2\right\} .$$

1. Déterminer un estimateur de maximum de vraisemblance de (μ, ρ). Cet estimateur est-il unique ?
2. Soit $\theta = (\mu, \rho) \in \Theta$. Montrer que sous $\mathbb{P}_{n,\theta}$, la loi jointe de $(X_{n:1}, X_{n:n})$ a une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$ donnée par
$$(u, v) \mapsto f_{n,\theta}(u, v) := \frac{n(n-1)}{\rho^n}(v - u)^{n-2} \mathbb{1}_{A_\theta}(u, v) .$$
3. Soit $\theta = (\mu, \rho) \in \Theta$. Montrer que sous $\mathbb{P}_{n,\theta}$, la loi de R_n a une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$ donnée par
$$r \mapsto g_{n,\theta}(r) := \frac{n(n-1)}{\rho^n} r^{n-2}(\rho - r) \mathbb{1}_{]0,\rho]}(r) .$$
4. Soit $\theta = (\mu, \rho) \in \Theta$. Montrer que la suite $n(\rho - R_n)$ converge en loi sous $\mathbb{P}_{n,\theta}$ et identifier la loi limite.

Exercice 4. Dans la suite, toutes les variables sont définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

On considère un problème de classification binaire. Nous étudions la situation où les données d'apprentissage sont affectées d'erreurs sur les labels. Pour formaliser cette situation, nous supposons disposer d'un ensemble d'apprentissage $\{(X_i, Z_i)\}_{i=1}^n$, où n est le nombre d'exemples, $\{(X_i, Z_i)\}_{i=1}^n$ est une suite de variables aléatoires i.i.d. de même loi que (X, Z). Nous notons (X, Y) le couple "attribut" "label", avec $Y \in \{0, 1\}$ et notons Z, le label bruité

$$Z := \begin{cases} Y & \text{si } B = 0 , \\ 1 - Y & \text{si } B = 1 , \end{cases}$$

où B est une variable de Bernoulli de paramètre $\mathrm{p} \in [0, 1/2[$ indépendante de (X, Y).

Nous notons $\mathcal{C}$ un ensemble de règles de classification que nous supposons de cardinal $|\mathcal{C}|$ fini. Pour tout $g \in \mathcal{C}$, nous définissons

- le risque de classification sur les observations sans bruit : $R(g) := \mathbb{P}(g(X) \neq Y)$,
- le risque de classification sur les observations bruitées : $R^b(g) := \mathbb{P}(g(X) \neq Z)$,
- le risque empirique sur les observations bruitées : $R_n^b(g) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{g(X_i) \neq Z_i\}}$.
- g^* une règle de classification minimisant le risque de classification sans bruit : $g^* \in \arg\min_{g \in \mathcal{C}} R(g)$

— $\hat{g}_n^*$ une règle de classification minimisant le risque empirique sur les observations bruitées : $\hat{g}_n^* \in \arg\min_{g\in\mathcal{C}} R_n^b(g)$

— R^* le minimum du risque de classification : $R^* := \inf_{g\in\mathcal{C}} R(g) = R(g^*)$.

Soit $\varepsilon > 0$ et $\delta \in]0,1[$, fixés dans toute la suite. Nous allons démontrer que la suite de règles de classification $\hat{g}_n^*$ est (ε,δ)-PAC, i.e. il existe un entier $n(\varepsilon,\delta)$ (que nous déterminerons) tel que

$$\forall n \geq n(\varepsilon,\delta), \qquad \mathbb{P}(R(\hat{g}_n^*) - R^* > \varepsilon) \leq \delta\ . \tag{1}$$

Posons $\bar{\varepsilon} := \varepsilon(1-2\mathrm{p})$.

1. Montrer que pour tout $g \in \mathcal{C}$,

$$R^b(g) = \mathrm{p} + R(g)(1-2\mathrm{p})\ .$$

   En déduire une relation entre $R^b(g) - R^b(g^*)$ et $R(g) - R(g^*)$ pour tout classifieur $g \in \mathcal{C}$.

2. Montrer qu'il existe $n_0(\varepsilon,\delta)$ (que l'on déterminera) tel que pour tout $n \geq n_0(\varepsilon,\delta)$,

$$\mathbb{P}\left(R_n^b(g^*) - R^b(g^*) \leq \bar{\varepsilon}/2\right) \geq 1 - \delta/2\ .$$

3. Montrer qu'il existe $n_1(\varepsilon,\delta)$ (que l'on déterminera) tel que pour tout $n \geq n_1(\varepsilon,\delta)$,

$$\mathbb{P}\left(\sup_{g\in\mathcal{C}}\{R^b(g) - R_n^b(g)\} \leq \bar{\varepsilon}/2\right) \geq 1 - \delta/2\ .$$

4. Etablir (1) [déterminer $n(\varepsilon,\delta)$].
5. De combien doit augmenter le nombre de données pour avoir la même garantie entre $\mathrm{p} = 0$ et $\mathrm{p} = 0,05$?

1 Solutions

Solution 1. 1. Nous avons

$$\mathbb{E}_{\bar{\boldsymbol{\beta}},\gamma,\bar{\sigma}^2}[\mathbf{Y}] = A\bar{\boldsymbol{\beta}} + \mathbf{z}\gamma = A(\bar{\boldsymbol{\beta}} + \mathbf{k}_z\gamma) + \mathbf{z}_\perp\gamma.$$

2. L'estimateur des moindres carrés de λ est donné par

$$\hat{\boldsymbol{\lambda}} = (V^TV)^{-1}V^T\mathbf{Y} = \begin{bmatrix} A^TA & 0 \\ 0 & \|\mathbf{z}_\perp\|^2 \end{bmatrix}^{-1} \begin{bmatrix} A^T\mathbf{Y} \\ \mathbf{z}_\perp^T\mathbf{Y} \end{bmatrix} = \begin{bmatrix} C^{-1}A^T\mathbf{Y} \\ \|\mathbf{z}^T\|^{-2}z_\perp^T\mathbf{Y} \end{bmatrix}$$

3. Nous posons $\hat{\boldsymbol{\alpha}} = C^{-1}A^T\mathbf{Y}$, $\hat{\gamma} = \|\mathbf{z}^T\|^{-2}z_\perp^T\mathbf{Y}$ et $\hat{\bar{\boldsymbol{\beta}}} = \hat{\boldsymbol{\alpha}} - \mathbf{k}_z\hat{\gamma}$. Nous allons montrer que $(\hat{\bar{\boldsymbol{\beta}}}, \hat{\gamma})$ est l'estimateur des moindres carrés de $(\bar{\boldsymbol{\beta}}, \gamma)$. En effet : pour tout $\boldsymbol{\lambda} \in \mathbb{R}^{p+1}$, nous avons

$$\|\mathbf{Y} - A\hat{\bar{\boldsymbol{\beta}}} - \mathbf{z}\hat{\gamma}\|^2 = \|\mathbf{Y} - A\hat{\boldsymbol{\alpha}} + A\mathbf{k}_z\hat{\gamma} - \mathbf{z}\hat{\gamma}\|^2 = \|\mathbf{Y} - A\hat{\boldsymbol{\alpha}} - \mathbf{z}_\perp\hat{\gamma}\|^2 \leq \|\mathbf{Y} - V\boldsymbol{\lambda}\|^2$$

Et donc, pour tout $(\bar{\boldsymbol{\beta}}, \gamma) \in \mathbb{R}^p \times \mathbb{R}$, en posant

$$\boldsymbol{\lambda} = \begin{bmatrix} \bar{\boldsymbol{\beta}} + A\mathbf{k}_z\gamma \\ \gamma \end{bmatrix}$$

nous avons

$$\|\mathbf{Y} - A\hat{\bar{\boldsymbol{\beta}}} - \mathbf{z}\hat{\gamma}\|^2 \leq \|\mathbf{Y} - A\bar{\boldsymbol{\beta}} - \mathbf{z}\gamma\|^2$$

4. $\mathbf{Y}$ est un vecteur gaussien $n \times 1$ de moyenne $A\bar{\boldsymbol{\beta}} + \mathbf{z}\gamma$ et de covariance $\sigma^2\mathrm{I}$. Donc, $\mathbf{z}_\perp^T\mathbf{Y}/\|\mathbf{z}_\perp\|^2$ est une variable gaussienne de moyenne γ (car $\mathbf{z}_\perp^TA = 0$ et $\mathbf{z}_\perp^T\mathbf{z} = \|\mathbf{z}_\perp\|^2$) et de variance $\sigma^2\|\mathbf{z}_\perp\|^{-2}$.

5. $\mathrm{Im}(V) = \mathrm{Im}(A) \perp \mathrm{Vect}(\mathbf{z}_\perp)$ où $\perp$ est la somme directe orthogonale et $\mathrm{Vect}(\mathbf{z}_\perp)$ est l'espace vectoriel engendré par $\mathbf{z}_\perp$. Par conséquent, $\bar{\mathrm{H}}$ est la somme du projecteur orthogonal sur $\mathrm{Im}(A)$ et du projecteur orthogonal sur $\mathbf{z}_\perp$.

6. $\hat{\hat{\sigma}}^2 = (n-p-1)^{-1}\{\|\mathbf{Y}\|^2 - \|\bar{\mathrm{H}}\mathbf{Y}\|^2\}$ est un estimateur sans biais de σ^2. Comme

$$\|\bar{\mathrm{H}}\mathbf{Y}\|^2 = \|\,\mathrm{H}\,\mathbf{Y}\|^2 + \frac{(\mathbf{z}_\perp^T\mathbf{Y})^2}{\|\mathbf{z}_\perp\|^2}$$

nous en déduisons que

$$\hat{\hat{\sigma}}^2 = \frac{n-p}{n-p-1}\hat{\sigma}^2 - \frac{1}{n-p-1}\frac{(\mathbf{z}_\perp^T\mathbf{Y})^2}{\|\mathbf{z}_\perp\|^2}\ .$$

7. Comme $\sigma^{-1}\{\mathbf{Y} - (A\boldsymbol{\beta} - \mathbf{z}\gamma)\}$ est sous $\mathbb{Q}_{\bar{\boldsymbol{\beta}},\gamma,\bar{\sigma}^2}$ un vecteur gaussien centré de covariance identité,

$$(n-p-1)\hat{\sigma}^2/\sigma^2$$

est distribué suivant une loi du *chi*^{2} à $(n-p-1)$ degrés de liberté. Par le théorème de Cochran, les vecteurs $\bar{\mathrm{H}}\mathbf{Y}$ et $(\mathrm{I} - \bar{\mathrm{H}})\mathbf{Y}$ sont indépendants sous $\mathbb{Q}_{\bar{\boldsymbol{\beta}},\gamma,\bar{\sigma}^2}$. L'indépendance de $\hat{\hat{\sigma}}^2$ et $\hat{\gamma}$ en découle car $\hat{\hat{\sigma}}^2$ est une fonction de $(\mathrm{I} - \bar{\mathrm{H}})\mathbf{Y}$ et $\hat{\gamma}$ de $\bar{\mathrm{H}}\mathbf{Y}$.

8. Il découle de la question précédente que $\mathbb{Q}_{\bar{\boldsymbol{\beta}},\gamma,\bar{\sigma}^2}$,

$$T_n(\mathbf{Y}) = \frac{\sqrt{n}\|\mathbf{z}_\perp\|(\hat{\gamma}-\gamma)}{\hat{\hat{\sigma}}}$$

est une loi de Student à $n-p-1$-degrés de liberté. En notant pour $b \in [0,1[$, t_b^{n-p-1} le quantile d'ordre b de la loi de Student, le test de

$$\delta_n(\mathbf{Y}) = \mathbb{1}_{\left[t^{n-p-1}_{1-a/2},\infty\right[}(|T_n(\mathbf{Y})|)$$

est un test de niveau a.

Solution . 1. Sous $\mathbb{P}_\theta$, les v.a. $(X_i)_i$ sont indépendantes et de densité p_θ donc la vraisemblance est donnée par

$$\theta \mapsto \theta^n \exp\left((\theta-1)\sum_{i=1}^n \log X_i\right)\prod_{i=1}^n \mathbb{1}_{X_i\in]0,1[}.$$

Pour $Z \in]0,1[^n$, la log-vraisemblance normalisée est

$$\theta \mapsto \log\theta + (\theta-1)\frac{1}{n}\sum_{i=1}^n \log X_i.$$

Il s'agit d'une fonction de classe C^1 sur $\mathbb{R}_+^*$ dont la dérivée vaut $\theta^{-1}+\frac{1}{n}\sum_{i=1}^n \log X_i$ et s'annule en

$$\theta_{n,\star} := -\frac{1}{\frac{1}{n}\sum_{i=1}^n \log X_i}.$$

De plus, on peut vérifier que la dérivée est positive sur $]0,\theta_{n,\star}]$ et négative sur $[\theta_{n,\star},+\infty[$. Ce point est donc bien un maximum de la fonction vraisemblance. Ainsi $\hat{\theta}_n = \theta_{n,\star}$.

2. **Réponse 1.** Utiliser le cours pour donner la loi asymptotique du maximum de vraisemblance dans un modèle régulier.
Réponse 2. L'estimateur est de la forme $\left(\frac{1}{n}\sum_{i=1}^n U_i\right)^{-1}$ où $U_i := -\log X_i$. Le TCL pour des v.a. i.i.d. possédant un moment d'ordre 2 donne

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n U_i - \mathbb{E}_{n,\theta}[U_i]\right) \overset{\mathbb{P}_{n,\theta}}{\Longrightarrow} \mathrm{N}(0,\mathrm{Var}_{n,\theta}(U_1))\ .$$

Puis nous appliquons la méthode δ avec la fonction $g : u \mapsto 1/u$, qui est de classe C^1 sur $\mathbb{R}^+$ et inversible. Il vient

$$\sqrt{n}\left(g\left(\frac{1}{n}\sum_{i=1}^n U_i\right) - g\left(\mathbb{E}_{n,\theta}[U_i]\right)\right) \overset{\mathbb{P}_{n,\theta}}{\Longrightarrow} g'\left(\mathbb{E}_{n,\theta}[U_i]\right)\mathrm{N}(0,\mathrm{Var}_{n,\theta}(U_1)) \equiv \mathrm{N}(0,\left(\mathbb{E}_{n,\theta}[U_i]\right)^{-4}\mathrm{Var}_{n,\theta}(U_1))\ .$$

Ici, sous $\mathbb{P}_{n,\theta}$, U_i suit une Gamma de paramètres $(1,\theta)$ (voir question suivante). Son espérance est $1/\theta$ et sa variance est $1/\theta^2$. On en déduit que

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \overset{\mathbb{P}_{n,\theta}}{\Longrightarrow} \mathrm{N}(0,\theta^2)\ .$$

3. Soit h une fonction mesurable positive. Nous avons, en utilisant que le changement de variable $y = -\log x$ est bijectif de $]0,1[$ dans $\mathbb{R}_+^*$,

$$\mathbb{E}_\theta\left[h(Y_1)\right] = \mathbb{E}_\theta\left[h(-\log X_1)\right] = \theta \int_0^1 h(-\log x) x^{\theta-1} \mathrm{d}x$$
$$= \theta \int_0^{+\infty} h(y)\exp(-(\theta-1)y)\exp(-y)\mathrm{d}y = \theta \int_0^{+\infty} h(y)\exp(-\theta y)\mathrm{d}y.$$

On reconnaît la densité d'une loi Gamma de paramètres $(1,\theta)$. Sous $\mathbb{P}_\theta$, les v.a. $(X_i)_i$ sont i.i.d. donc il en est de même pour les v.a. $(Y_i)_i$. En utilisant un résultat sur la somme de lois Gamma indépendantes de premier paramètre identique (voir Lemme IV-5.13), il vient $\sum_{i=1}^n Y_i$ suit une loi Gamma de paramètres (n,θ).

4. Soit $\theta > 0$. Nous avons, en utilisant le résultat de la question précédente,

$$\mathbb{E}_\theta\left[\hat{\theta}_n\right] = \mathbb{E}_\theta\left[\frac{1}{\frac{1}{n}\sum_{i=1}^n(-\log X_i)}\right] = \frac{\theta^n}{\Gamma(n)}\int_0^\infty \frac{n}{s} s^{n-1}\exp(-\theta s)\mathrm{d}s$$
$$= \frac{n\theta^n}{\Gamma(n)}\int_0^\infty s^{n-2}\exp(-\theta s)\mathrm{d}s = \frac{n\theta^n\Gamma(n-1)}{\Gamma(n)\theta^{n-1}}\frac{\theta^{n-1}}{\Gamma(n-1)}\int_0^\infty s^{n-2}\exp(-\theta s)\mathrm{d}s$$
$$= \frac{n\Gamma(n-1)}{\Gamma(n)}\theta = \frac{n}{n-1}\theta.$$

Dans la dernière ligne, on a observé que l'on intégrait la densité d'une loi Gamma de paramètres $(n-1,\theta)$; et utilisé que pour un entier $n > 0$, $\Gamma(n) = (n-1)!$.

Calculons le moment d'ordre 2. Il vient

$$\mathbb{E}_\theta\left[\hat{\theta}_n^2\right] = \mathbb{E}_\theta\left[\frac{1}{\frac{1}{n^2}\left(\sum_{i=1}^n(-\log X_i)\right)^2}\right] = \frac{\theta^n}{\Gamma(n)}\int_0^\infty \frac{n^2}{s^2} s^{n-1}\exp(-\theta s)\mathrm{d}s$$
$$= \frac{n^2\theta^n}{\Gamma(n)}\int_0^\infty s^{n-3}\exp(-\theta s)\mathrm{d}s = \frac{n^2\theta^n\Gamma(n-2)}{\Gamma(n)\theta^{n-2}}\frac{\theta^{n-2}}{\Gamma(n-2)}\int_0^\infty s^{n-3}\exp(-\theta s)\mathrm{d}s$$
$$= \frac{n^2\Gamma(n-2)}{\Gamma(n)}\theta^2 = \frac{n^2}{(n-1)(n-2)}\theta^2.$$

On en déduit que la variance vaut

$$\mathrm{Var}_\theta(\hat{\theta}_n) = \frac{n^2}{(n-1)(n-2)}\theta^2 - \frac{n^2}{(n-1)^2}\theta^2 = \frac{n^2}{(n-1)}\theta^2\left(\frac{1}{n-2} - \frac{1}{n-1}\right) = \frac{n^2}{(n-1)^2(n-2)}\theta^2.$$

5. Nous cherchons a_n tel que pour tout $\theta > 0$, $\mathbb{E}_\theta\left[\tilde{\theta}_n\right] = \theta$. D'après la question précédente, il faut prendre

$$a_n := \frac{n-1}{n}.$$

6. Il faut comparer l'erreur quadratique de l'estimateur à la borne de Cramer-Rao. Comme l'estimateur est sans biais, sa variance et son erreur quadratique coïncident. Nous avons donc

$$\mathbb{E}_\theta\left[(\tilde{\theta}_n - \theta)^2\right] = a_n^2\,\mathrm{Var}_\theta(\hat{\theta}_n) = \frac{\theta^2}{n-2}.$$

Par ailleurs, la borne de Cramer-Rao est donnée par $(nI(\theta))^{-1}$ où $I(\theta)$ désigne l'information de Fisher associée à une seule observation :

$$I(\theta) := \mathbb{E}_\theta \left[(\partial_\theta \log p_\theta(X_1))^2 \right] ;$$

cette quantité est indépendante de n. Nous avons donc une dépendance en la taille n de l'échantillon qui ne coïncide pas ; l'estimateur sans biais n'est pas efficace.

7. Montrons que pour tout $\theta > 0$, la loi de $n\theta/\hat{\theta}_n$ sous $\mathbb{P}_\theta$ ne dépend pas de θ, et identifions cette loi. Soit h une fonction mesurable positive ; il vient, en utilisant l'expression de $\hat{\theta}_n$ et la question 3,

$$\begin{aligned}\mathbb{E}_\theta \left[h\left(\frac{n\theta}{\hat{\theta}_n} \right) \right] &= \mathbb{E}_\theta \left[h\left(\theta \sum_{i=1}^n (-\log X_i) \right) \right] \\ &= \frac{\theta^n}{\Gamma(n)} \int_0^\infty h(\theta\, s) s^{n-1} \exp(-\theta s) \mathrm{d}s = \frac{1}{\Gamma(n)} \int_0^\infty h(w) w^{n-1} \exp(-w) \mathrm{d}w.\end{aligned}$$

On reconnaît la loi Gamma de paramètres $(n, 1)$.
Notons $q_{n,\tau}$ les quantiles d'une loi Gamma de paramètres $(n, 1)$. Nous avons donc pour tout $\theta > 0$,

$$\mathbb{P}_\theta \left(q_{n,\alpha/2} \le \frac{n\theta}{\hat{\theta}_n} \le q_{n,1-\alpha/2} \right) = 1 - \alpha.$$

Ce qui permet de proposer un intervalle de confiance *exact* de la forme

$$\left[\hat{\theta}_n \, \frac{q_{n,\alpha/2}}{n}, \hat{\theta}_n \, \frac{q_{n,1-\alpha/2}}{n} \right].$$

Solution . 1. La vraisemblance du n-échantillon Z est

$$\theta \mapsto \frac{1}{\rho^n} \prod_{i=1}^n \mathbb{1}_{[\mu-\rho/2, \mu+\rho/2]}(X_i) = \frac{1}{\rho^n} \mathbb{1}_{A_\theta}(X_{n:1}, X_{n:n}).$$

En particulier, $\mathbb{1}_{A_\theta}(X_{n:1}, X_{n:n}) = 1$ pour tout $\theta = (\mu, \rho) \in \mathbb{R} \times \mathbb{R}_+^*$ vérifiant

$$\rho = \mu + \rho/2 - (\mu - \rho/2) \ge X_{n:n} - X_{n:1}$$

et la vraisemblance est donc maximisé par la plus petite valeur de ρ satisfaisant cette contrainte. Nous avons donc $\hat{\rho}_n = X_{n:n} - X_{n:1}$. Cette valeur de $\hat{\rho}_n$ étant choisie, $\mathbb{1}_{A_{\mu,\hat{\rho}_n}(X_{n:1},X_{n:n})} = 1$ pour tout $\mu \in \mathbb{R}$ vérifiant

$$\mu - \hat{\rho}_n/2 \le X_{n:1} \le X_{n:n} \le \mu + \hat{\rho}_n/2 \implies \mu \le (1/2)(X_{n:1} + X_{n:n}) \le \mu.$$

L'estimateur du maximum de vraisemblance est unique et donné par

$$\hat{\mu}_n = \frac{X_{n:1} + X_{n:n}}{2}, \qquad \hat{\rho}_n = X_{n:n} - X_{n:1}.$$

2. Soient $s \le t$. Nous avons, par indépendance des v.a. puis leur loi identique,

$$\mathbb{P}_{n,\theta} \left(s \le \min_i X_i \le \max_i X_i \le t \right) = \prod_{i=1}^n \mathbb{P}_{n,\theta} \left(s \le X_i \le t \right) = \left(\mathbb{P}_{n,\theta} \left(s \le X_1 \le t \right) \right)^n.$$

En remarquant que sous $\mathbb{P}_\theta$, X_1 suit une loi uniforme sur $[\mu-\rho/2, \mu+\rho/2]$, il vient que cette probabilité est nulle lorsque $s \geq \mu+\rho/2$ ou $t \leq \mu-\rho/2$; et sinon :

$$\mathbb{P}_{n,\theta}\left(s \leq \min_i X_i \leq \max_i X_i \leq t\right) = \rho^{-n}\ \{(t \wedge (\mu+\rho/2)) - (s \vee (\mu-\rho/2))\}^n \tag{2}$$

D'autre part, calculons pour tout $s \leq t$

$$\begin{aligned}\pi(s,t) &:= \int_{\{(x,y): s \leq x \leq y \leq t\}} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_s^t \left(\int_x^t f(x,y)\mathrm{d}y\right)\,\mathrm{d}x \\ &= \frac{n(n-1)}{\rho^n}\int_s^t \left(\int_x^t (y-x)^{n-2}\mathbb{1}_{\mu-\rho/2 \leq x \leq y \leq \mu+\rho/2}(x,y)\,\mathrm{d}y\right)\,\mathrm{d}x.\end{aligned}$$

Nous avons, du fait de l'indicatrice, que $\pi(s,t) = 0$ lorsque $s \geq \mu+\rho/2$ ou $t \leq \mu-\rho/2$; sinon,

$$\begin{aligned}\pi(s,t) &= \frac{n(n-1)}{\rho^n}\int_s^t \mathbb{1}_{[\mu-\rho/2,\mu+\rho/2]}(x)\left(\int_x^{t\wedge(\mu+\rho/2)} (y-x)^{n-2}\,\mathrm{d}y\right)\,\mathrm{d}x \\ &= \frac{n}{\rho^n}\int_{s\vee(\mu-\rho/2)}^{t\wedge(\mu+\rho/2)} (\{t \wedge (\mu+\rho/2)\} - x)^{n-1}\,\mathrm{d}x \\ &= \rho^{-n}\ \{(t \wedge (\mu+\rho/2)) - (s \vee (\mu-\rho/2))\}^n\,.\end{aligned}$$

La comparaison de cette valeur de $\pi(s,t)$ et de la quantité (2) permet de démontrer l'expression de la densité proposée.

3. Soit $\theta \in \Theta$. Pour toute fonction mesurable positive h, nous avons

$$\mathbb{E}_\theta\left[h\left(R_n\right)\right] = \frac{n(n-1)}{\rho^n}\int_{A_\theta} h(y-x)\ (y-x)^{n-2}\ \mathrm{d}x\,\mathrm{d}y.$$

Nous faisons un changement de variable $(x,z) \leftarrow (x, y-x)$ qui donne, combiné au théorème de Fubini,

$$\begin{aligned}\mathbb{E}_\theta\left[h\left(R_n\right)\right] &= \frac{n(n-1)}{\rho^n}\int_{\{(x,z): \mu-\rho/2 \leq x \leq x+z \leq \mu+\rho/2\}} h(z)\ (z)^{n-2}\ \mathrm{d}x\,\mathrm{d}z \\ &= \frac{n(n-1)}{\rho^n}\int_0^\rho h(z)\ z^{n-2}\left(\int_{\mu-\rho/2}^{\mu+\rho/2-z}\ \mathrm{d}x\right)\,\mathrm{d}z \\ &= \frac{n(n-1)}{\rho^n}\int_0^\rho h(z)\ z^{n-2}(\rho-z)\,\mathrm{d}z.\end{aligned}$$

On reconnaît l'expression de la densité donnée.

4. Remarquons que nous avons $\mathbb{P}_{n,\theta}(\rho - R_n \geq 0) = 1$ pour tout $\theta \in \Theta$; donc la limite en loi sera une v.a. positive ou nulle. Etablissons la convergence en loi par un argument de convergence des fonctions de répartition. Soit $t > 0$. Nous écrivons, en utilisant la question 3,

$$\mathbb{P}_{n,\theta}\left(n(\rho - R_n) \leq t\right) = \mathbb{P}_{n,\theta}\left(\rho - R_n \leq t/n\right) = \mathbb{P}_{n,\theta}\left(\rho - t/n \leq R_n\right).$$

A t fixé, nous allons considérer la limite $n \to +\infty$; nous pouvons donc considérer que n est assez grand pour que $\rho - t/n \geq 0$. Il vient, pour une toute valeur de n assez grande ($t > 0$ fixé)

$$\begin{aligned}
\mathbb{P}_{n,\theta}\left(n(\rho - R_n) \leq t\right) &= \frac{n(n-1)}{\rho^n} \int_{\rho - t/n}^{\rho} r^{n-2}(\rho - r)\mathrm{d}r \\
&= \frac{n(n-1)}{\rho^n} \left\{ \frac{\rho}{n-1}\left(\rho^{n-1} - (\rho - t/n)^{n-1}\right) - \frac{1}{n}\left(\rho^n - (\rho - t/n)^n\right) \right\} \\
&= n\left\{1 - \left(1 - \frac{t}{\rho n}\right)^{n-1}\right\} - (n-1)\left\{1 - \left(1 - \frac{t}{\rho n}\right)^{n}\right\} \\
&= 1 - \left(1 - \frac{t}{\rho n}\right)^{n} \left\{ n\left(1 - \frac{t}{\rho n}\right)^{-1} - (n-1) \right\}
\end{aligned}$$

Par conséquent, nous avons

$$\lim_{n\to\infty} \mathbb{P}_{n,\theta}\left(n(\rho - R_n) \leq t\right) = 1 - \left(1 + \frac{t}{\rho}\right)\mathrm{e}^{-t/\rho} .$$

En dérivant par rapport à t, nous obtenons que la densité de la loi limite est donnée par

$$g(t) = \frac{t}{\rho^2}\mathrm{e}^{-t/\rho}\mathbb{1}_{\mathbb{R}^+}(t)$$

Nous reconnaissons la densité d'une loi Gamma$(2, 1/\rho)$.

Solution 4. 1. Notons que

$$\begin{aligned}
R^b(g) = \mathbb{P}(g(X) \neq Z) &= \mathbb{P}(g(X) \neq Z, Y = Z) + \mathbb{P}(g(X) \neq Z, Y \neq Z) \\
&= \mathbb{P}(g(X) \neq Y, B = 0) + \mathbb{P}(g(X) = Y, B = 1).
\end{aligned}$$

Par indépendance du couple (X, Y) et de B, nous avons

$$R^b(g) = (1 - \mathrm{p})R(g) + \{1 - R(g)\}\mathrm{p} = \mathrm{p} + (1 - 2\mathrm{p})R(g) .$$

On en déduit que

$$R^b(g) - R^b(g^*) = (1 - 2\mathrm{p})\ \left(R(g) - R(g^*)\right).$$

2. Soit $g \in \mathcal{C}$. Nous allons démontrer que

$$\mathbb{P}\left(R_n^b(g) - R^b(g) \geq \bar{\varepsilon}/2\right) \leq \delta/2 , \qquad \mathbb{P}\left(R_n^b(g) - R^b(g) \leq -\bar{\varepsilon}/2\right) \leq \delta/2 . \tag{3}$$

Nous écrivons

$$R_n^b(g) - R^b(g) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{g(X_i) \neq Z_i} - \mathbb{P}(g(X) \neq Z) = \frac{1}{n}\sum_{i=1}^{n} U_i$$

avec

$$U_i := \mathbb{1}_{g(X_i) \neq Z_i} - \mathbb{P}(g(X) \neq Z) .$$

Par hypothèses, les v.a. $(U_i)_i$ sont indépendantes ; de plus, elles sont centrées, et on a

$$\mathbb{P}\left(-m \leq U_i \leq 1-m\right) = 1 \qquad m := \mathbb{P}(g(X) \neq Z)\ .$$

En utilisant l'inégalité de Hoeffding, nous avons puisque $b_i - a_i = 1$,

$$\mathbb{P}\left(R_n^b(g) - R^b(g) \geq \bar{\varepsilon}/2\right) \leq \mathrm{e}^{-n\bar{\varepsilon}^2/2}\ ,$$

Il suffit de choisir n tel que

$$\mathrm{e}^{-n\bar{\varepsilon}^2/2} \leq \delta/2$$

soit

$$n \geq n_0(\mathrm{p}, \varepsilon, \delta) := \frac{2}{\bar{\varepsilon}^2} \log\left(\frac{2}{\delta}\right)\ .$$

On raisonne de même pour établir la seconde inéqualité (en utilisant encore Hoeffding) ; et on obtient la même valeur de n_0, qui est donc celle que l'on retient comme garantissant les deux contrôles (3).

3. Nous allons démontrer que

$$\mathbb{P}\left(\sup_{g \in \mathcal{C}} \{R^b(g) - R_n^b(g)\} \geq \bar{\varepsilon}/2\right) \leq \delta/2\ .$$

Puisque $\mathcal{C}$ est de cardinal fini, et en utilisant la question 2, nous avons

$$\begin{aligned}\mathbb{P}\left(\sup_{g \in \mathcal{C}} \{R^b(g) - R_n^b(g)\} \geq \bar{\varepsilon}/2\right) &\leq \sum_{g \in \mathcal{C}} \mathbb{P}\left(R^b(g) - R_n^b(g) \geq \bar{\varepsilon}/2\right) \\ &\leq \sum_{g \in \mathcal{C}} \mathbb{P}\left(R_n^b(g) - R^b(g) \leq -\bar{\varepsilon}/2\right) \\ &\leq |\mathcal{C}|\, \mathrm{e}^{-n\bar{\varepsilon}^2/2}\end{aligned}$$

(voir aussi le Théorème III-2.7). Il vient

$$n \geq n_1(\mathrm{p}, \varepsilon, \delta) := \frac{2}{\bar{\varepsilon}^2} \left\{\log(|\mathcal{C}|) + \log(2/\delta)\right\}\ .$$

4. En utilisant la question 1, nous avons

$$R(\hat{g}_n^*) - R(g^*) = \frac{1}{1-2\mathrm{p}} \left(R^b(\hat{g}_n^*) - R^b(g^*)\right)\ .$$

Nous écrivons

$$R^b(\hat{g}_n^*) - R^b(g^*) \leq R^b(\hat{g}_n^*) - R_n^b(\hat{g}_n^*) + R_n^b(\hat{g}_n^*) - R_n^b(g^*) + R_n^b(g^*) - R^b(g^*)$$

Par définition de $\hat{g}_n^*$, le terme $R_n^b(\hat{g}_n^*) - R_n^b(g^*)$ est négatif ; il vient

$$R(\hat{g}_n^*) - R(g^*) \leq \frac{1}{1-2p} \left(\sup_{g \in \mathcal{C}} \{R^b(g) - R_n^b(g)\} + R_n^b(g^*) - R^b(g^*)\right)\ .$$

Par suite

$$\begin{aligned}\mathbb{P}\left(R(\hat{g}_n^*) - R(g^*) > \varepsilon\right) &\leq \mathbb{P}\left(\sup_{g \in \mathcal{C}}\{R^b(g) - R_n^b(g)\} + R_n^b(g^*) - R^b(g^*) > \bar{\varepsilon}\right) \\ &\leq \mathbb{P}\left(\sup_{g \in \mathcal{C}}\{R^b(g) - R_n^b(g)\} > \bar{\varepsilon}/2\right) + \mathbb{P}\left(R_n^b(g^*) - R^b(g^*) > \bar{\varepsilon}/2\right)\end{aligned}$$

En utilisant les questions 2 (voir Eq. (3)) et 3, chacun des deux termes de droite est inférieur à $\delta/2$ dès lors que $n \geq n_0 \vee n_1$ i.e.

$$n \geq n(\mathrm{p}, \varepsilon, \delta) := \frac{2}{\varepsilon^2(1-2\mathrm{p})^2}\left\{\log(|\mathcal{C}|) + \log(2/\delta)\right\}$$

5. Observons que

$$n(p, \varepsilon, \delta) = \frac{1}{(1-2\mathrm{p})^2} n(0, \varepsilon, \delta) \ .$$

Par suite, pour obtenir la précision (ε, δ), nous avons

$$n(0.05, \varepsilon, \delta) = \frac{100}{81} n(0, \varepsilon, \delta) \approx 1.23\, n(0, \varepsilon, \delta) \ .$$