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

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

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

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

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 $H_0 : \gamma = 0, \quad$ 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 $\mathrm{p} = 0,05$?