Par convention, les vecteurs sont des vecteurs colonne. Pour un vecteur ou une matrice x, $x^\top$ désigne la transposée de x. I_n est la matrice identité de taille $n \times n$.

L'espérance et la variance d'une loi géométrique sur $\mathbb{N}$ de paramètre θ, $\theta \in]0,1[$, sont : $(1-\theta)/\theta$ et $(1-\theta)/\theta^2$.

Sujet

Exercice 1. On cherche à expliquer une variable quantitative Y_i (réponse) par un vecteur de régresseurs $\mathbf{z}_i$, $i \in \{1,\ldots,n\}$. On pose $\theta := (\boldsymbol{\beta}, \sigma^2) \in \Theta := \mathbb{R}^p \times \mathbb{R}_+^*$. Le *modèle linéaire* consiste à supposer que, pour tout $\theta \in \Theta$, les variables

$$\varepsilon_i(\theta) := \sigma^{-1}\left\{Y_i - \mathbf{z}_i^\top \boldsymbol{\beta}\right\}, \quad i = 1,2,\ldots,n,$$

sont des variables indépendantes et identiquement distribuées. Nous utilisons les notations matricielles / vectorielles :

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \sigma\boldsymbol{\varepsilon}(\theta), \tag{1}$$

où

$$\mathbf{Y} := \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \qquad \boldsymbol{\varepsilon}(\theta) := \begin{bmatrix} \varepsilon_1(\theta) \\ \vdots \\ \varepsilon_n(\theta) \end{bmatrix}, \qquad \mathbf{Z} := \begin{bmatrix} \mathbf{z}_1^\top \\ \mathbf{z}_2^\top \\ \vdots \\ \mathbf{z}_n^\top \end{bmatrix} \in \mathbb{R}^{n\times p};$$

$\mathbf{Y}$ est le vecteur des observations, $\boldsymbol{\beta}$ est le vecteur des paramètres de régression et $\mathbf{Z}$ est la matrice de régression de taille $n \times p$. Nous supposons que $n > p+1$.

Pour $i \in \{1,\ldots,n\}$, nous notons $\mathbf{Z}_{[i]}$ et $\mathbf{Y}_{[i]}$ la matrice $\mathbf{Z}$ et le vecteur $\mathbf{Y}$ dont la i-ème ligne a été retirée. Nous supposons que

GM1 pour tout $i \in \{1,\ldots,n\}$, la matrice $\mathbf{Z}_{[i]}$ est de rang p.

GM2 les erreurs de régression sont *homoscédastiques* : pour tout $\theta \in \Theta$, $\mathbb{E}_\theta[\boldsymbol{\varepsilon}(\theta)] = 0$ et $\mathrm{Cov}_\theta(\boldsymbol{\varepsilon}(\theta)) = \mathrm{I}_n$.

L'hypothèse GM1 implique que la matrice de Gram $\mathbf{Z}^\top\mathbf{Z}$ est inversible et l'hypothèse GM2 entraine que pour tout $\theta \in \Theta$

$$\mathbb{E}_\theta\left[\mathbf{Y}\right] = \mathbf{Z}\boldsymbol{\beta}\,, \qquad \mathrm{Cov}_\theta\left(\mathbf{Y}\right) = \sigma^2 \mathrm{I}_n\,.$$

Nous notons par $\widehat{\boldsymbol{\beta}}$ l'estimateur des moindres carrés basé sur $(\mathbf{Z},\mathbf{Y})$ et pour $i \in \{1,\ldots,n\}$, par $\widehat{\boldsymbol{\beta}}_{[i]}$ l'estimateur des moindres carrés basé sur $(\mathbf{Z}_{[i]}, \mathbf{Y}_{[i]})$:

$$\widehat{\boldsymbol{\beta}} := (\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{Y} = \mathbf{Z}^{\#}\mathbf{Y}\,, \qquad \widehat{\boldsymbol{\beta}}_{[i]} := (\mathbf{Z}_{[i]}^\top\mathbf{Z}_{[i]})^{-1}\mathbf{Z}_{[i]}^\top\mathbf{Y}_{[i]} = \mathbf{Z}_{[i]}^{\#}\mathbf{Y}_{[i]}\,. \tag{2}$$

Nous introduisons le résidu et le résidu prédictif, donnés respectivement par

$$\hat{e}_i := Y_i - \mathbf{z}_i^\top\widehat{\boldsymbol{\beta}} \qquad \hat{e}_{[i]} := Y_i - \mathbf{z}_i^\top\widehat{\boldsymbol{\beta}}_{[i]}.$$

1. Montrer que pour tout $\theta \in \Theta$, $\mathbb{E}_\theta[\widehat{\boldsymbol{\beta}}_{[i]}] = \boldsymbol{\beta}$ et $\mathrm{Cov}_\theta(\widehat{\boldsymbol{\beta}}_{[i]}) = \sigma^2\,(\mathbf{Z}_{[i]}^\top\mathbf{Z}_{[i]})^{-1}$.
2. Montrer que pour tout $\theta \in \Theta$, $\mathbb{E}_\theta[\hat{e}_{[i]}] = 0$ et $\mathrm{Var}_\theta(\hat{e}_{[i]}) = \sigma^2 + \sigma^2\mathbf{z}_i^\top(\mathbf{Z}_{[i]}^\top\mathbf{Z}_{[i]})^{-1}\mathbf{z}_i$.

On admettra que si A est une matrice $q \times q$ inversible, et $\mathbf{u} \in \mathbb{R}^q$,

$$(A + \mathbf{u}\mathbf{u}^\top)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{u}\mathbf{u}^\top A^{-1}}{1 + \mathbf{u}^\top A^{-1}\mathbf{u}} \,,$$
$$(A - \mathbf{u}\mathbf{u}^\top)^{-1} = A^{-1} + \frac{A^{-1}\mathbf{u}\mathbf{u}^\top A^{-1}}{1 - \mathbf{u}^\top A^{-1}\mathbf{u}} \,, \quad \text{si } \mathbf{u}^\top A^{-1}\mathbf{u} \neq 1.$$

Nous posons, pour $i \in \{1, \ldots, n\}$, $m_{ii} := \mathbf{z}_i^\top \left(\mathbf{Z}^\top \mathbf{Z}\right)^{-1} \mathbf{z}_i$.

3. Montrer que $0 \leq m_{ii} < 1$ pour tout $i = 1, \cdots, n$.
4. Montrer que pour tout $i = 1, \cdots, n$,
$$\left(\mathbf{Z}_{[i]}^\top \mathbf{Z}_{[i]}\right)^{-1} = \left(\mathbf{Z}^\top \mathbf{Z}\right)^{-1} + \frac{1}{1 - m_{ii}} \left(\mathbf{Z}^\top \mathbf{Z}\right)^{-1} \mathbf{z}_i \mathbf{z}_i^\top \left(\mathbf{Z}^\top \mathbf{Z}\right)^{-1}.$$
5. En déduire que pour tout $i = 1, \cdots, n$,
$$\widehat{\boldsymbol{\beta}}_{[i]} = \widehat{\boldsymbol{\beta}} - \frac{\hat{e}_i}{1 - m_{ii}} \left(\mathbf{Z}^\top \mathbf{Z}\right)^{-1} \mathbf{z}_i.$$
6. Montrer que pour tout $i = 1, \cdots, n$: $\hat{e}_{[i]} = \hat{e}_i / (1 - m_{ii})$ et $\mathrm{Var}_\theta(\hat{e}_{[i]}) = \sigma^2 / (1 - m_{ii})$.

On suppose maintenant que $\mathbf{Y}$ est l'observation canonique d'un modèle gaussien

$$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \left\{\mathrm{N}_n(\mathbf{Z}\boldsymbol{\beta}, \sigma^2 \mathrm{I}_n) \;:\; \theta = (\boldsymbol{\beta}, \sigma^2) \in \mathbb{R}^p \times \mathbb{R}_+^*\right\}).$$

Nous notons $\mathbf{H}_{[i]} := \mathbf{Z}_{[i]} \mathbf{Z}_{[i]}^\#$ et $\hat{\sigma}_{[i]}^2 := (n - p - 1)^{-1} \|(\mathrm{I}_{n-1} - \mathbf{H}_{[i]})\mathbf{Y}_{[i]}\|^2$.

7. Sous $\mathbb{P}_\theta$, montrer que $\hat{e}_{[i]} = Y_i - \mathbf{z}_i^\top \widehat{\boldsymbol{\beta}}_{[i]}$ est distribué suivant une loi Gaussienne dont on précisera les paramètres en fonction de σ^2 et m_{ii}.
8. Sous $\mathbb{P}_\theta$, montrer que $(n - p - 1)\hat{\sigma}_{[i]}^2 / \sigma^2$ est distribué suivant une loi de χ^2 à $(n - p - 1)$ degrés de liberté, et est indépendant de $\widehat{\boldsymbol{\beta}}_{[i]}$ et Y_i.
9. En déduire la distribution sous $\mathbb{P}_\theta$ du résidu standardisé
$$t_i := \frac{\hat{e}_{[i]}}{\sqrt{\hat{\sigma}_{[i]}^2 / (1 - m_{ii})}}.$$
10. Proposer une méthode pour détecter une valeur aberrante dans les observations.

Exercice 2. Soit $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ une suite de n-échantillons du modèle statistique

$$\left(\mathbb{R}_+^2, \mathcal{B}(\mathbb{R}_+^2), \{p_\theta \cdot \mathrm{Leb}^{\otimes 2}, \theta = (\theta_1, \theta_2) \in \Theta := \mathbb{R}_+^\star \times \mathbb{R}_+^\star\}\right)$$

où

$$p_\theta : (x, y) \in \mathbb{R}_+^2 \mapsto \theta_1^{-1} \exp(-x/\theta_1) \theta_2^{-1} \exp(-y/\theta_2).$$

On s'intéresse dans cet exercice à l'estimation de $g(\theta)$, où la fonction g est définie par $g : \Theta \to \mathbb{R}_+,\ \theta \mapsto \mathbb{P}_\theta(X_1 \leq Y_1)$. On admettra que le modèle statistique est régulier. On rappelle que $\mathbb{E}_\theta[X_1] = \theta_1$, $\mathrm{Var}_\theta(X_1) = \theta_1^2$.

On considère d'abord l'estimateur

$$T_n = n^{-1} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq Y_i\}} \,.$$

1. Montrer que, pour tout $\theta \in \Theta$, $g(\theta) = \theta_2/(\theta_1 + \theta_2)$.
2. Montrer que (T_n) est une suite d'estimateurs de $g(\theta)$ asymptotiquement normale. Déterminer sa variance asymptotique.

On note $\hat{\theta}_n^{\mathrm{MV}}$ l'estimateur du maximum de vraisemblance de θ et on considère $\tilde{T}_n = g(\hat{\theta}_n^{\mathrm{MV}})$ comme second estimateur de $g(\theta)$.

3. Montrer que $\hat{\theta}_n^{\mathrm{MV}}$ est une suite d'estimateurs de θ asymptotiquement normale et déterminer sa covariance asymptotique.
4. Montrer que $(\tilde{T}_n)$ est une suite d'estimateurs de $g(\theta)$ asymptotiquement normale et déterminer sa variance asymptotique.
5. Quel estimateur de $g(\theta)$ doit-on choisir ?
6. Construire un intervalle de confiance asymptotique de $g(\theta)$ de probabilité de couverture $1 - \alpha$, pour $\alpha \in]0, 1[$.
7. Construire une suite de tests de niveau asymptotique α de l'hypothèse $H_0 : \ g(\theta) \leq 1/2, \quad \text{contre} \quad H_1 : \ g(\theta) < 1/2$.

Exercice 3. La probabilité $\theta \in]0, 1[$ d'un événement est généralement estimée en examinant la fréquence à laquelle il se produit (on parle de "succès") au cours de n essais indépendants. Une autre façon d'estimer θ est de déterminer le nombre d'échecs Y_r pour obtenir $r \geq 2$ succès.

Soit $\{X_k\}_{k \in \mathbb{N}}$ une suite de variables aléatoires indépendantes distribuées suivant une loi de Bernoulli de paramètre $\theta \in \Theta :=]0, 1[$. Nous notons pour $r \in \mathbb{N}^*$, $Z_r := Y_r - Y_{r-1}$, le nombre d'échecs entre le $(r-1)$-ème et le r-ème succès, où par convention nous posons $Y_0 := 0$.

1. Montrer que sous $\mathbb{P}_\theta$, les statistiques $\{Z_1, \ldots, Z_r\}$ sont indépendantes et identiquement distribuées de loi géométrique sur $\mathbb{N}$ de paramètre θ.
2. Déterminer le modèle statistique induit par les statistiques $Z_1, \ldots, Z_r$.
3. Justifier soigneusement que, sous $\mathbb{P}_\theta$, la loi de Y_r est donnée par
$$\mathbb{P}_\theta(Y_r = y) = \binom{y+r-1}{r-1} \ \theta^r (1-\theta)^y, \qquad y \in \mathbb{N}.$$
4. Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}_r$ du paramètre θ sous le modèle induit par $Z_1, \ldots, Z_r$.
5. Montrer que $\hat{\theta}_r$ coïncide avec l'estimateur du maximum de vraisemblance sous le modèle induit par Y_r.
6. Montrer que la suite d'estimateurs $\{\hat{\theta}_r\}_{r=1}^{\infty}$ est asymptotiquement normale. Préciser la variance asymptotique.
7. Montrer que l'estimateur
$$\tilde{\theta}_r := \frac{r-1}{Y_r + r - 1}$$
est un estimateur sans biais du paramètre θ.
8. Montrer $\tilde{\theta}_r$ est l'unique estimateur sans biais sous le modèle induit par Y_r.

Soit $T(Z_1, \ldots, Z_r)$ un estimateur sans biais de θ basé sur $Z_1, \ldots, Z_r$.

9. Montrer qu'il existe une fonction $\widetilde{T}(Y_r)$ telle que, pour tout $\theta \in \Theta$, $\mathbb{E}_\theta\left[T(Z_1, \ldots, Z_r) \,|\, Y_r\right] = \widetilde{T}(Y_r)$.
10. Montrer que $\widetilde{T}(Y_r)$ est un estimateur sans biais de θ.
11. Montrer que pour tout $\theta \in \Theta$,
$$\mathbb{E}_\theta\left[(T(Z_1, \ldots, Z_r) - \theta)^2\right] = \mathbb{E}_\theta\left[(T(Z_1, \ldots, Z_r) - \widetilde{T}(Y_r))^2\right] + \mathbb{E}_\theta\left[(\widetilde{T}(Y_r) - \theta)^2\right]$$
12. En déduire que $\tilde{\theta}_r$ est aussi l'estimateur sans biais de variance minimale dans le modèle induit par $Z_1, \ldots, Z_r$.

Exercice 4. Nous considérons un problème de classification binaire. Nous supposons que l'observation (X, Y) (où $X \in \mathsf{X} \subseteq \mathbb{R}^d$ est le vecteur des attributs, et $Y \in \{-1, 1\}$ est l'étiquette) est un vecteur aléatoire défini sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d \times \{-1, 1\}$. Nous notons
$$\eta(X) := \mathbb{E}\left[\, \mathbb{1}_{\{Y=1\}} \,\middle|\, X\right] \ ,$$
et nous supposons que $\eta(X)$ est à valeurs dans $]0, 1[$ avec probabilité 1.

Pour une règle de décision $g : \mathsf{X} \to \mathbb{R}$, nous nous intéressons à la fonction de risque exponentiel
$$R_e(g) := \mathbb{E}[\ \exp(-Y g(X))\ ] \ . \tag{3}$$
Rappelons que g est une règle de décision "douce". Nous déciderons 1 si signe$(g(x)) \geq 0$ et -1 sinon. Nous posons pour $(\eta, p) \in\,]0, 1[\, \times \mathbb{R}$,
$$C_e(\eta, p) := \eta \exp(-p) + (1 - \eta) \exp(p) \ .$$

1. Montrer que
$$R_e(g) = \mathbb{E}[\ C_e(\eta(X), g(X))\ ] \ .$$
2. Montrer que pour tout $\eta \in\,]0, 1[$, $p \mapsto C_e(\eta, p)$ admet un minimum unique ; on le notera $p_e(\eta)$.

On pose pour tout $x \in \mathsf{X}$,
$$g_e(x) := p_e\left(\eta(x)\right) .$$

3. En déduire que g_e est le classifieur bayésien pour le risque exponentiel (3).

On pose $R_e^* := R_e(g_e)$. Pour tout $\eta \in\,]0, 1[$, on définit
$$H_e(\eta) := C_e(\eta, p_e(\eta)) = \inf_{p \in \mathbb{R}} C_e(\eta, p).$$

4. Montrer que
$$H_e(\eta) = 2\sqrt{\eta(1 - \eta)} \quad \text{pour } \eta \in\,]0, 1[.$$

Nous notons R_{0-1} le risque 0-1 : pour toute fonction $g : \mathsf{X} \to \mathbb{R}$,
$$R_{0-1}(g) := \mathbb{E}[\mathbb{1}_{\{\text{signe}(g(X)) \neq Y\}}] \ .$$
On a toujours, pour toute fonction $g : \mathsf{X} \to \mathbb{R}$, $R_{0-1}(g) = R_{0-1}(\text{signe}(g))$. Soit
$$g^*_{0-1}(x) := \text{signe}(2\eta(x) - 1) \ , \qquad x \in \mathsf{X} \ .$$
On rappelle que g^*_{0-1} minimise le risque R_{0-1}. On pose
$$R^*_{0-1} := R_{0-1}(g^*_{0-1}) \ .$$

5. Montrer que signe(g_e) est une règle de décision Bayésienne pour le risque 0-1.
6. Montrer que pour tout classifieur g

$$R_{0-1}(g) - R^*_{0-1} = \mathbb{E}[\ |2\eta(X) - 1|\ \mathbb{1}_{\{g(X)\,(2\eta(X)-1)\leq 0\}}\ ].$$

7. En déduire que si ψ est convexe sur $[0,1[$ et $\psi(0) = 0$:

$$\psi(R_{0-1}(g) - R^*_{0-1}) \leq \mathbb{E}\left[\ \psi(|2\eta(X) - 1|)\ \mathbb{1}_{\{g(X)\,(2\eta(X)-1)\leq 0\}}\ \right]$$

8. Montrer que, pour tout $\eta \in]0,1[$,

$$\inf_{p\in\mathbb{R}\ :\, p\,(2\eta-1)\leq 0} C_e(\eta, p) = 1\ \ .$$

On définit la fonction H_e^- sur $]0,1[$ et la fonction ψ_e sur $[0,1[$, par

$$\begin{aligned} H_e^-(\eta) &:= \inf_{p\in\mathbb{R}\ :\, p\,(2\eta-1)\leq 0} C_e(\eta, p)\ , \quad \eta \in]0,1[\ , \\ \psi_e(\theta) &:= H_e^-\left(\frac{1+\theta}{2}\right) - H_e\left(\frac{1+\theta}{2}\right) = 1 - H_e\left(\frac{1+\theta}{2}\right), \quad \theta \in [0,1[\ . \end{aligned}$$

9. Montrer que pour tout $\eta \in]0,1[$, $H_e(\eta) = H_e(1-\eta)$ et

$$\psi_e(|2\eta - 1|) = 1 - H_e(\eta) = H_e^-(\eta) - H_e(\eta)\ .$$

10. Montrer que $\psi_e(0) = 0$ et que ψ_e est convexe sur $[0,1[$.
11. En déduire que, pour tout classifieur g, nous avons

$$\psi_e\left(R_{0-1}(g) - R^*_{0-1}\right) \leq R_e(g) - R^*_e\ .$$

Solution . 1. Soit $\theta \in \Theta$. Par linéarité de l'espérance,

$$\mathbb{E}_\theta[\widehat{\boldsymbol{\beta}}_{[i]}] = (\mathbf{Z}_{[i]}^\top \mathbf{Z}_{[i]})^{-1} \mathbf{Z}_{[i]}^\top \, \mathbb{E}_\theta[\mathbf{Y}_{[i]}];$$

nous savons que $\mathbb{E}_\theta[\mathbf{Y}] = \mathbf{Z}\boldsymbol{\beta}$ dont nous déduisons que $\mathbb{E}_\theta[\mathbf{Y}_{[i]}] = \mathbf{Z}_{[i]}\boldsymbol{\beta}$. Ainsi

$$\mathbb{E}_\theta[\widehat{\boldsymbol{\beta}}_{[i]}] = \boldsymbol{\beta}.$$

De même

$$\mathrm{Cov}_\theta(\widehat{\boldsymbol{\beta}}_{[i]}) = \mathrm{Cov}_\theta((\mathbf{Z}_{[i]}^\top \mathbf{Z}_{[i]})^{-1} \mathbf{Z}_{[i]}^\top \, \mathbf{Y}_{[i]}) = (\mathbf{Z}_{[i]}^\top \mathbf{Z}_{[i]})^{-1} \mathbf{Z}_{[i]}^\top \, \mathrm{Cov}_\theta(\mathbf{Y}_{[i]}) \, \mathbf{Z}_{[i]} (\mathbf{Z}_{[i]}^\top \mathbf{Z}_{[i]})^{-1}.$$

Nous savons que $\mathrm{Cov}_\theta(\mathbf{Y}) = \sigma^2 \mathrm{I}_n$ dont nous déduisons que $\mathrm{Cov}_\theta(\mathbf{Y}_{[i]}) = \sigma^2 \mathrm{I}_{n-1}$. Ainsi

$$\mathrm{Cov}_\theta(\widehat{\boldsymbol{\beta}}_{[i]}) = \sigma^2 \, (\mathbf{Z}_{[i]}^\top \mathbf{Z}_{[i]})^{-1}.$$

2. Soit $\theta \in \Theta$. Nous savons que $\mathbb{E}_\theta[Y_i] = \mathbf{z}_i^\top \boldsymbol{\beta}$ dont nous déduisons, en utilisant la question 1,

$$\mathbb{E}_\theta \left[\hat{e}_{[i]}\right] = \mathbb{E}_\theta \left[Y_i\right] - \mathbf{z}_i^\top \mathbb{E}_\theta \left[\widehat{\boldsymbol{\beta}}_{[i]}\right] = \mathbf{z}_i^\top \boldsymbol{\beta} - \mathbf{z}_i^\top \boldsymbol{\beta} = 0.$$

De plus,

$$\mathrm{Var}_\theta(\hat{e}_{[i]}) = \mathrm{Var}_\theta(Y_i - \mathbf{z}_i^\top \widehat{\boldsymbol{\beta}}_{[i]}).$$

$\widehat{\boldsymbol{\beta}}_{[i]}$ est une fonction de $Y_{[i]}$ donc sous $\mathbb{P}_\theta$, Y_i et $\widehat{\boldsymbol{\beta}}_{[i]}$ sont indépendants. La variance de la somme est donc égale à la somme des variances. Il vient, en utilisant la question 1,

$$\mathrm{Var}_\theta(\hat{e}_{[i]}) = \mathrm{Var}_\theta(Y_i) + \mathrm{Var}_\theta(\mathbf{z}_i^\top \widehat{\boldsymbol{\beta}}_{[i]}) = \sigma^2 + \mathbf{z}_i^\top \, \mathrm{Var}_\theta(\widehat{\boldsymbol{\beta}}_{[i]}) \mathbf{z}_i = \sigma^2 + \sigma^2 \mathbf{z}_i^\top (\mathbf{Z}_{[i]}^\top \mathbf{Z}_{[i]})^{-1} \mathbf{z}_i.$$

3. Soit $i \in \{1, \cdots, n\}$. Nous avons

$$\mathbf{Z}^\top \mathbf{Z} = \mathbf{Z}_{[i]}^\top \mathbf{Z}_{[i]} + \mathbf{z}_i \mathbf{z}_i^\top.$$

Appliquons la première relation avec $A \leftarrow \mathbf{Z}_{[i]}^\top \mathbf{Z}_{[i]}$ et $\mathbf{u} \leftarrow \mathbf{z}_i$; puis multiplions par $\mathbf{z}_i^\top$ à gauche et par $\mathbf{z}_i$ à droite. Cela donne

$$m_{ii} = \mathbf{z}_i^\top \left(\mathbf{Z}^\top \mathbf{Z}\right)^{-1} \mathbf{z}_i = \alpha_i - \frac{\alpha_i^2}{1+\alpha_i} = \frac{\alpha_i}{1+\alpha_i}, \qquad \alpha_i := \mathbf{z}_i^\top \left(\mathbf{Z}_{[i]}^\top \mathbf{Z}_{[i]}\right)^{-1} \mathbf{z}_i.$$

Puisque $\alpha_i \geq 0$, nous avons $m_{ii} \in [0, 1[$.

4. Soit $i \in \{1, \cdots, n\}$. Nous appliqons la seconde relation avec $A \leftarrow \mathbf{Z}^\top \mathbf{Z}$ et $\mathbf{u} \leftarrow \mathbf{z}_i$. Cela donne la relation demandée.

5. Soit $i \in \{1, \cdots, n\}$. Nous remplaçons $(\mathbf{Z}_{[i]}^\top \mathbf{Z}_{[i]})^{-1}$ par l'expression trouvée à la question 4, et nous utilisons le fait que $\mathbf{Z}^\top \mathbf{Y} = \mathbf{Z}_{[i]}^\top \mathbf{Y}_{[i]} + Y_i \mathbf{z}_i$. Cela donne

$$\widehat{\boldsymbol{\beta}}_{[i]} = \left(\mathbf{Z}^\top \mathbf{Z}\right)^{-1} \left(\mathbf{Z}^\top \mathbf{Y} - Y_i \mathbf{z}_i\right) + \frac{1}{1-m_{ii}} \left(\mathbf{Z}^\top \mathbf{Z}\right)^{-1} \mathbf{z}_i \mathbf{z}_i^\top \left(\mathbf{Z}^\top \mathbf{Z}\right)^{-1} \left(\mathbf{Z}^\top \mathbf{Y} - Y_i \mathbf{z}_i\right).$$

Ainsi,

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{[i]} &= \widehat{\boldsymbol{\beta}} - Y_i \left(\mathbf{Z}^\top \mathbf{Z}\right)^{-1} \mathbf{z}_i + \frac{1}{1-m_{ii}} \left(\mathbf{Z}^\top \mathbf{Z}\right)^{-1} \mathbf{z}_i \mathbf{z}_i^\top \widehat{\boldsymbol{\beta}} - Y_i \frac{m_{ii}}{1-m_{ii}} \left(\mathbf{Z}^\top \mathbf{Z}\right)^{-1} \mathbf{z}_i \\
&= \widehat{\boldsymbol{\beta}} - \frac{Y_i}{1-m_{ii}} \left(\mathbf{Z}^\top \mathbf{Z}\right)^{-1} \mathbf{z}_i + \frac{1}{1-m_{ii}} \left(\mathbf{Z}^\top \mathbf{Z}\right)^{-1} \mathbf{z}_i \mathbf{z}_i^\top \widehat{\boldsymbol{\beta}} \\
&= \widehat{\boldsymbol{\beta}} - \frac{\hat{e}_i}{1-m_{ii}} \left(\mathbf{Z}^\top \mathbf{Z}\right)^{-1} \mathbf{z}_i.
\end{aligned}$$

6. Soit $i \in \{1, \cdots, n\}$. En utilisant la question 5, nous avons par définition de $\hat{e}_{[i]}$

$$\hat{e}_{[i]} = Y_i - \mathbf{z}_i^\top \widehat{\boldsymbol{\beta}} + \frac{\hat{e}_i}{1 - m_{ii}} m_{ii} = \hat{e}_i + \frac{m_{ii}}{1 - m_{ii}} \hat{e}_i = \frac{\hat{e}_i}{1 - m_{ii}}.$$

   De plus, d'après la question 2 puis la question 4

$$\begin{aligned} \mathrm{Var}_\theta(\hat{e}_{[i]}) &= \sigma^2 \left(1 + \mathbf{z}_i^\top (\mathbf{Z}_{[i]}^\top \mathbf{Z}_{[i]})^{-1} \mathbf{z}_i\right) \\ &= \sigma^2 \left(1 + m_{ii} + \frac{m_{ii}^2}{1 - m_{ii}}\right) = \frac{\sigma^2}{1 - m_{ii}}. \end{aligned}$$

7. Sous $\mathbb{P}_\theta$, les v.a. Y_i et $\mathbf{Y}_{[i]}$ sont indépendantes et

$$Y_i \sim \mathrm{N}(\mathbf{z}_i^\top \boldsymbol{\beta}, \sigma^2), \qquad \mathbf{Y}_{[i]} \sim \mathrm{N}(\mathbf{Z}_{[i]}\boldsymbol{\beta}, \sigma^2 \mathrm{I}_{n-1}).$$

   Puisque

$$Y_i - \mathbf{z}_i^\top \widehat{\boldsymbol{\beta}}_{[i]} = Y_i - \mathbf{z}_i^\top (\mathbf{Z}_{[i]}^\top \mathbf{Z}_{[i]})^{-1} \mathbf{Z}_{[i]}^\top \mathbf{Y}_{[i]},$$

   nous en déduisons que $Y_i - \mathbf{z}_i^\top \boldsymbol{\beta}_{[i]}$ suit une loi gaussienne. L'espérance et la variance sont données par (voir questions 2 et 6)

$$0, \qquad \frac{\sigma^2}{1 - m_{ii}}.$$

8. Sous $\mathbb{P}_\theta$, $\mathbf{Y}_{[i]} \sim \mathrm{N}(\mathbf{Z}_{[i]}^\top \boldsymbol{\beta}, \sigma^2 \mathrm{I}_{n-1})$. De plus, $\mathbf{H}_{[i]}$ est un projecteur sur un espace de dimension p (d'après GM1) et donc $\mathrm{I}_{n-1} - \mathbf{H}_{[i]}$ est un projecteur sur un espace de dimension $n-1-p$. D'après la Proposition IV-3.20, en observant que $(\mathrm{I}_{n-1} - \mathbf{H}_{[i]})\mathbf{Z}_{[i]}^\top \boldsymbol{\beta} = 0$,

$$(n - p - 1)\frac{\hat{\sigma}_{[i]}^2}{\sigma^2} = \sigma^{-2} \left\|(\mathrm{I}_{n-1} - \mathbf{H}_{[i]})\mathbf{Y}_{[i]}\right\|^2 \sim \chi^2(n - 1 - p).$$

   De plus, $\|(\mathrm{I}_{n-1} - \mathbf{H}_{[i]})\mathbf{Y}_{[i]}\|^2$ est indépendant de $\mathbf{H}_{[i]} \mathbf{Y}_{[i]}$ donc de $\mathbf{Z}_{[i]} \widehat{\boldsymbol{\beta}}_{[i]}$ et par suite de $\mathbf{z}_i^\top \widehat{\boldsymbol{\beta}}_{[i]}$ en observant que

$$\mathbf{z}_i^\top \widehat{\boldsymbol{\beta}}_{[i]} = \mathbf{z}_i^\top (\mathbf{Z}_{[i]}^\top \mathbf{Z}_{[i]})^{-1} (\mathbf{Z}_{[i]}^\top \mathbf{Z}_{[i]}) \widehat{\boldsymbol{\beta}}_{[i]} = \mathbf{z}_i^\top (\mathbf{Z}_{[i]}^\top \mathbf{Z}_{[i]})^{-1} \mathbf{Z}_{[i]}^\top \left(\mathbf{Z}_{[i]} \widehat{\boldsymbol{\beta}}_{[i]}\right).$$

   Il est aussi indépendant de Y_i puisqu'il ne dépend que de $Y_j, j \neq i$. En conclusion, $\|(\mathrm{I}_{n-1} - \mathbf{H}_{[i]})\mathbf{Y}_{[i]}\|^2$ est indépendant de $\hat{e}_{[i]}$.

9. La question 8 et la définition d'une loi de Student (voir définition IV-3.21 dans le polycopié) montrent que

$$\frac{\hat{e}_{[i]}}{\sqrt{\hat{\sigma}_{[i]}^2 / (1 - m_{ii})}} = \frac{\sqrt{1 - m_{ii}}\, \hat{e}_{[i]}}{\sigma} \, \frac{1}{\sqrt{\hat{\sigma}_{[i]}^2 / \sigma^2}} \sim t_{n-1-p},$$

   où t_q désigne une loi de Student à q degrés de liberté.

10. Soit $q_{1-\alpha/2}$ le quantile d'ordre $1-\alpha/2$ d'une loi de Student à $n-1-p$ degrés de liberté. Nous avons

$$\mathbb{P}_\theta\left(\frac{|\hat{e}_{[i]}|}{\sqrt{\hat{\sigma}^2_{[i]}/(1-m_{ii})}} \geq q_{1-\alpha/2}\right) = \alpha.$$

Nous pouvons mettre en oeuvre le test de zone de rejet

$$\frac{|\hat{e}_{[i]}|}{\sqrt{\hat{\sigma}^2_{[i]}/(1-m_{ii})}} \geq q_{1-\alpha/2},$$

qui rejettera si l'observation Y_i est vue comme une donnée aberrante.

Solution . 1. Sous $\mathbb{P}_{n,\theta}$ les variables aléatoires $\mathbb{1}_{\{X_i \leq Y_i\}}$ sont i.i.d., de loi de Bernoulli de paramètre $g(\theta)$, donc, d'après le théorème de la limite centrale, on a

$$\sqrt{n}(T_n - g(\theta)) \xrightarrow{\mathbb{P}_\theta} \mathrm{N}(0, g(\theta)(1-g(\theta))) \ .$$

2. On donne ici 2 preuves du résultat.
Preuve 1 : avec le calcul de $\hat{\theta}_n^{\mathrm{MV}}$. On calcule la vraisemblance

$$L_n(\theta) = \frac{1}{\theta_1^n}\exp\left(-\frac{\sum_{i=1}^n X_i}{\theta_1}\right)\frac{1}{\theta_2^n}\exp\left(-\frac{\sum_{i=1}^n Y_i}{n}\right) \ .$$

On en déduit que

$$\nabla \log L_n(\theta) = \begin{bmatrix} -\frac{n}{\theta_1} + \frac{\sum_{i=1}^n X_i}{\theta_1^2} \\ -\frac{n}{\theta_2} + \frac{\sum_{i=1}^n Y_i}{\theta_2^2} \end{bmatrix} \ . \tag{4}$$

Il s'en suit que, pour toute valeur de θ_2, la fonction $\theta_1 \mapsto L_n(\theta_1, \theta_2)$ atteint son maximum en $\hat{\theta}_1 = n^{-1}\sum_{i=1}^n X_i$ et que la fonction $L_n(\hat{\theta}_1, \theta_2)$ atteint son maximum en $\hat{\theta}_2 = n^{-1}\sum_{i=1}^n Y_i$. On a donc, pour tout θ,

$$L_n(\theta) \leqslant L_n(\hat{\theta}_1, \theta_2) \leqslant L_n(\hat{\theta}_1, \hat{\theta}_2) \ .$$

et donc

$$\hat{\theta}_n^{\mathrm{MV}} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = \frac{1}{n}\sum_{i=1}^n \begin{bmatrix} X_i \\ Y_i \end{bmatrix} \ .$$

Sous $\mathbb{P}_{n,\theta}$, les vecteurs $(X_i, Y_i)^T$ sont i.i.d. et chaque couple (X_i, Y_i) est constitué de variables indépendantes, X_i suit une loi exponentielle de paramètre $1/\theta_1$ et Y_i une loi exponentielle de paramètre $1/\theta_2$. Ces deux variables sont donc de carré intégrable et chaque couple $(X_i, Y_i)^T$ a pour espérance θ. On a donc, par le théorème de la limite centrale multivarié,

$$\sqrt{n}(\hat{\theta}_n^{\mathrm{MV}} - \theta) \xrightarrow{\mathbb{P}_\theta} \mathrm{N}(0, \Sigma_\theta) \ ,$$

où la matrice Σ_θ est la matrice de variance-covariance du couple (X_1, Y_1) et vaut donc

$$\Sigma_\theta = \begin{bmatrix} \theta_1^2 & 0 \\ 0 & \theta_2^2 \end{bmatrix} \ .$$

Preuve 2 : On utilise le fait que le modèle est régulier et on a donc la normalité asymptotique de la suite $\hat{\theta}_n^{\mathrm{MV}}$ par le théorème II-3.4 p134 du cours avec

$$\sqrt{n}(\hat{\theta}_n^{\mathrm{MV}} - \theta) \xrightarrow{\mathbb{P}_\theta} \mathrm{N}(0, \mathbb{I}(\theta)^{-1}) ,$$

où $\mathbb{I}(\theta)$ est la matrice d'information de Fisher du modèle. Calculons cette matrice, on a par (4) (avec $n = 1$)

$$\nabla \log p_\theta(X, Y) = \begin{bmatrix} -\frac{1}{\theta_1} + \frac{X}{\theta_1^2} \\ -\frac{1}{\theta_2} + \frac{Y}{\theta_2^2} \end{bmatrix} .$$

On a donc par définition de l'information de Fisher

$$\begin{aligned}
\mathbb{I}(\theta) &= \mathbb{E}_\theta[\nabla \log p_\theta(X, Y) \nabla \log p_\theta(X, Y)^T] \\
&= \begin{bmatrix} \frac{1}{\theta_1^4}\mathrm{Var}_\theta(X) & \frac{1}{\theta_1^2\theta_2^2}\mathrm{Cov}_\theta(X, Y) \\ \frac{1}{\theta_1^2\theta_2^2}\mathrm{Cov}_\theta(X, Y) & \frac{1}{\theta_2^4}\mathrm{Var}_\theta(Y) \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\theta_1^2} & 0 \\ 0 & \frac{1}{\theta_2^2} \end{bmatrix} .
\end{aligned}$$

On conclut que

$$\sqrt{n}(\hat{\theta}_n^{\mathrm{MV}} - \theta) \xrightarrow{\mathbb{P}_\theta} \mathrm{N}\left(0, \begin{bmatrix} \theta_1^2 & 0 \\ 0 & \theta_2^2 \end{bmatrix}\right) ,$$

3. On calcule directement

$$\begin{aligned}
g(\theta) &= \mathbb{P}_\theta(X_1 \leqslant Y_1) \\
&= \int_0^{+\infty} \left(\int_x^{+\infty} p_\theta(x, y) \mathrm{d}y \right) \mathrm{d}x \\
&= \int_0^{+\infty} \theta_1^{-1} \exp\left(-\frac{x}{\theta_1}\right) \left(\int_x^{+\infty} \theta_2^{-1} \exp\left(-\frac{y}{\theta_2}\right) \mathrm{d}y \right) \mathrm{d}x \\
&= \int_0^{+\infty} \theta_1^{-1} \exp\left(-\frac{x}{\theta_1}\right) \exp\left(-\frac{x}{\theta_2}\right) \mathrm{d}x \\
&= \theta_1^{-1} \int_0^{+\infty} \exp\left(-\left(\frac{1}{\theta_1} + \frac{1}{\theta_2}\right)x\right) \mathrm{d}x \\
&= \theta_1^{-1} \left(\frac{1}{\theta_1} + \frac{1}{\theta_2}\right)^{-1} = \frac{\theta_2}{\theta_1 + \theta_2} .
\end{aligned}$$

4. On utilise la méthode Delta, ce qui est possible car, sous $\mathbb{P}_{n,\theta}$,
   (i) $\hat{\theta}_n^{\mathrm{MV}}$ est une suite asymptotiquement normale d'après la question 2.
   (ii) La fonction g est dérivable en θ d'après la question 3, de gradient

$$\nabla g(\theta) = \begin{bmatrix} -\frac{\theta_2}{(\theta_1+\theta_2)^2} \\ \frac{\theta_1+\theta_2-\theta_2}{(\theta_1+\theta_2)^2} \end{bmatrix} = \frac{1}{(\theta_1 + \theta_2)^2} \begin{bmatrix} -\theta_2 \\ \theta_1 \end{bmatrix} .$$

Il vient donc de la méthode Delta que

$$\sqrt{n}(g(\hat{\theta}_n^{\mathrm{MV}}) - g(\theta)) \xrightarrow{\mathbb{P}_\theta} \mathrm{N}\left(0, \nabla g(\theta)^T \Sigma_\theta \nabla g(\theta)\right) .$$

On calcule

$$\nabla g(\theta)^T \Sigma_\theta \nabla g(\theta) = \frac{1}{(\theta_1+\theta_2)^4}(\theta_1^2(-\theta_2)^2 + 2*0*(-\theta_2)\theta_1 + \theta_2^2\theta_1^2) = \frac{2\theta_1^2\theta_2^2}{(\theta_1+\theta_2)^4} \ .$$

5. Il reste à comparer la variance asymptotique précédente $v_2(\theta) = 2\theta_1^2\theta_2^2/(\theta_1+\theta_2)^4$ de l'estimateur $g(\hat\theta_n^{\mathrm{MV}})$ à la variance asymptotique $v_1(\theta) = g(\theta)(1-g(\theta))$ de l'estimateur T_n. Or, en utilisant la valeur de $g(\theta)$ calculée à la question 3,

$$v_1(\theta) = \frac{\theta_1\theta_2}{(\theta_1+\theta_2)^2} \ .$$

On a donc

$$v_2(\theta) = \frac{2\theta_1\theta_2}{(\theta_1+\theta_2)^2} v_1(\theta) \ ,$$

et comme $2\theta_1\theta_2 < (\theta_1+\theta_2)^2$, on a $v_2(\theta) < v_1(\theta)$. Ainsi, l'estimateur $g(\hat\theta_n^{\mathrm{MV}})$ est donc asymptotiquement préférable à T_n comme estimateur de $g(\theta)$.

Solution . 1. Les v.a. Z_k sont à valeur dans $\mathbb{N}$. Soit $a_1, \cdots, a_r$ des entiers naturels non nuls. Nous avons

$$\begin{aligned}\{Z_1 = a_1, \cdots, Z_r = a_r\} = \{&X_1 = \cdots = X_{a_1} = 0, X_{a_1+1} = 1,\} \\ &\cap \{X_{a_1+2} = \cdots = X_{a_1+1+a_2} = 0, X_{a_1+a_2+2} = 1\} \\ &\cdots \\ &\cap \{X_{\sum_{j=1}^{r-1} a_j + (r-1)+1} = \cdots = X_{\sum_{j=1}^{r-1} a_j + (r-1)+a_r} = 0, X_{\sum_{j=1}^{r} a_j + r} = 1\}\end{aligned}$$

Sous $\mathbb{P}_\theta$, les v.a. $(X_k)_k$ sont i.i.d. Bernoulli de paramètre θ ; nous en déduisons que

$$\mathbb{P}_\theta(Z_1 = a_1, \cdots, Z_r = a_r) = \prod_{j=1}^{r} (\theta(1-\theta)^{a_j}) = \theta^r (1-\theta)^{\sum_{j=1}^r a_j}.$$

Ainsi, sous $\mathbb{P}_\theta$, les v.a. $(Z_j)_j$ sont i.i.d. de loi géométrique à valeur dans $\mathbb{N}$, et de paramètre θ.

2. La question 1 entraine que $Z_1, \cdots, Z_r$ est un n-échantillon de taille r du modèle

$$(\mathbb{N}, \mathcal{P}(\mathbb{N}), \{\mathcal{G}(\theta), \theta \in]0,1[\})$$

3. Nous avons $Y_r = Y_r - Y_0 = \sum_{j=1}^r (Y_j - Y_{j-1}) = \sum_{j=1}^r Z_j$. Y_r est à valeur dans $\mathbb{N}$. Pour $y \in \mathbb{N}$, exprimons l'événement $\{Y_r = y\}$ à l'aide des v.a. X_i.

$$\{Y_r = y\} = \{\text{il y a } (r-1) \text{ v.a. } X_i \text{ valant 1 parmi } X_1, \cdots, X_{y+r-1}, \text{ et } X_{y+r} = 1\}.$$

La probabilité d'une suite binaire $(x_1, \cdots, x_{y+r})$ possédant r valeurs à 1 et y à zero, est

$$\mathbb{P}_\theta((X_1, \cdots, X_{y+r}) = (x_1, \cdots, x_{y+r})) = \theta^r (1-\theta)^y;$$

cette probabilité est indépendante de l'ordre des 0 et 1 dans cette suite. Il y a $\binom{y+r-1}{r-1}$ suites binaires $(x_1, \cdots, x_{y+r})$ ayant cette propriété. Ainsi

$$\mathbb{P}_\theta(Y_r = y) = \binom{y+r-1}{r-1} \ \theta^r (1-\theta)^y.$$

4. La vraisemblance est donnée par
$$\theta \mapsto \theta^r \, (1-\theta)^{\sum_{j=1}^r Z_j}$$
dont on déduit la log-vraisemblance
$$\theta \mapsto r \ln \theta + \left(\sum_{j=1}^r Z_j \right) \ln(1-\theta).$$
Cette fonction possède un unique maximum, donné par
$$\hat{\theta}_r := \frac{r}{\sum_{j=1}^r Z_j + r} = \frac{r}{Y_r + r}.$$
5. Ecrivons la vraisemblance dans le modèle induit par Y_r :
$$\theta \mapsto \binom{y+r-1}{r-1} \; \theta^r (1-\theta)^{Y_r} \propto \theta^r (1-\theta)^{Y_r}.$$
Un calcul analogue à celui de la question 4 montre que l'estimateur MV est donné par $r/(Y_r + r)$, qui est bien $\hat{\theta}_r$.
6. Nous avons $\hat{\theta}_r = \phi(r^{-1} \sum_{j=1}^r Z_j)$, avec $\phi(u) := 1/(1+u)$ de classe C^1 sur $\mathbb{R}_+$; la normalité asymptotique va être établie par application de la méthode delta.
Par le TCL pour des v.a. i.i.d. possédant un moment d'ordre deux, nous savons que
$$\sqrt{r} \left(r^{-1} \sum_{j=1}^r Z_j - \mathbb{E}_\theta \left[Z_1 \right] \right) \overset{\mathbb{P}_\theta}{\Longrightarrow} \mathrm{N}\left(0, \mathrm{Var}_\theta(Z_1)\right).$$
Nous savons que $\mathbb{E}_\theta \left[Z_1 \right] = (1-\theta)/\theta$, et $\mathrm{Var}_\theta(Z_1) = (1-\theta)/\theta^2$ (voir le rappel). Par application de la méthode delta, nous obtenons
$$\sqrt{r} \left(\phi \left(r^{-1} \sum_{j=1}^r Z_j \right) - \phi\left(\frac{1-\theta}{\theta} \right) \right) \overset{\mathbb{P}_\theta}{\Longrightarrow} \phi' \left(\frac{1-\theta}{\theta} \right) \; \mathrm{N} \left(0, \frac{1-\theta}{\theta^2} \right).$$
Puisque $\phi'(u) = -1/(1+u)^2$, nous avons
$$\phi' \left(\frac{1-\theta}{\theta} \right) = -\theta^2$$
dont nous déduisons que
$$\sqrt{r} \left(\hat{\theta}_r - \theta \right) \overset{\mathbb{P}_\theta}{\Longrightarrow} \mathrm{N}\left(0, (1-\theta)\theta^2\right).$$
7. Calculons le biais de l'estimateur à l'aide de la loi de Y_r sous $\mathbb{P}_\theta$ (voir question 3). Soit $\theta \in \Theta$.
$$\begin{aligned} \mathbb{E}_\theta \left[\frac{r-1}{Y_r + r - 1} \right] &= \sum_{y \geq 0} \frac{r-1}{y+r-1} \binom{y+r-1}{r-1} \; \theta^r (1-\theta)^y \\ &= \sum_{y \geq 0} \binom{y+r-2}{r-2} \; \theta^r (1-\theta)^y \end{aligned}$$

Méthode 1. Nous savons que pour tout $u \in \mathbb{N}^\star$,

$$\sum_{y \geq 0} \tbinom{y+u}{u} \ \theta^{u+1}(1-\theta)^y = 1$$

(constante de normalisation d'une loi de probabilité sur $\mathbb{N}$, voir question 3). Ainsi,

$$\mathbb{E}_\theta \left[\frac{r-1}{Y_r + r - 1} \right] = \theta \sum_{y \geq 0} \binom{y+r-2}{r-2} \ \theta^{r-1}(1-\theta)^y = \theta.$$

Ce qui établit l'absence de biais.
Méthode 2.

$$\mathbb{E}_\theta \left[\frac{r-1}{Y_r + r - 1} \right] = \frac{\theta^r}{(r-2)!} \sum_{y \geq 0} (y+1) \cdots (y+r-2) \ (1-\theta)^y$$

Nous savons que $\sum_{y \geq 0} (1-\theta)^y = \theta^{-1}$ et en dérivant ces deux expressions $r-2$ fois, nous obtenons

$$(-1)^{r-2} \sum_{y \geq 0} (y+1) \cdots (y+r-2) \ (1-\theta)^y = (-1)^{r-2}(r-2)! \, \theta^{-(r-1)}.$$

Ainsi

$$\mathbb{E}_\theta \left[\frac{r-1}{Y_r + r - 1} \right] = \theta.$$

8. Si il existe un autre estimateur $T(Y_r)$ sans biais, alors pour tout $\theta \in \Theta$

$$\theta = \mathbb{E}_\theta[T(Y_r)] = \sum_{y \geq 0} T(y) \binom{y+r-1}{r-1} \ \theta^r (1-\theta)^y.$$

Puisque $(r-1)/(Y_r + r - 1)$ est aussi sans biais, il vient que pour tout $\theta \in \Theta$

$$\Delta(\theta) = 0, \qquad \Delta(\theta) := \sum_{y \geq 0} \left(T(y) - \frac{r-1}{y+r-1} \right) \binom{y+r-1}{r-1} \ \theta^r (1-\theta)^y.$$

Δ est une série entière, qui est nulle sur Θ. Donc ses coefficients sont identiquement nuls. On en déduit que pour tout $y \in \mathbb{N}$, $T(y) = (r-1)/(y+r-1)$; ce qui établit l'unicité.

9. Soit f une fonction mesurable de carré intégrable ; par définition de l'espérance conditionnelle, nous avons

$$\mathbb{E}_\theta \left[\mathbb{E}_\theta \left[T(Z_1, \ldots, Z_r) \,|\, Y_r \right] \, f(Y_r) \right] = \mathbb{E}_\theta \left[T(Z_1, \cdots, Z_r) \, f(Y_r) \right]$$

Calculons le terme de droite et procédons par identification.

$$\begin{aligned}\mathbb{E}_\theta \left[T(Z_1, \cdots, Z_r) \, f(Y_r) \right] &= \mathbb{E}_\theta \left[T(Z_1, \cdots, Z_r) \, f(\sum_{j=1}^r Z_j) \right] \\ &= \sum_{a_1 \geq 0} \cdots \sum_{a_r \geq 0} T(a_1, \cdots, a_r) \, f(\sum_{j=1}^r a_j) \, \theta^r (1-\theta)^{\sum_{j=1}^r a_j}\end{aligned}$$

Nous faisons ensuite le changement de variable

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_{r-1} \\ y_r \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ \cdots & & & & & \\ 1 & 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \cdots \\ z_{r-1} \\ z_r \end{bmatrix}$$

dont le jacobien est 1, et nous obtenons

$$\begin{aligned}
&\mathbb{E}_\theta \left[T(Z_1, \cdots, Z_r)\, f(Y_r)\right] \\
&\quad = \sum_{y_1 \geq 0} \sum_{y_2 \geq y_1} \cdots \sum_{y_r \geq y_{r-1}} T(y_1, y_2 - y_1, \cdots, y_r - y_{r-1})\, f(y_r)\, \theta^r (1-\theta)^{y_r} \\
&\quad = \sum_{y_r \geq 0} \left(\sum_{y_{r-1}=0}^{y_r} \cdots \sum_{y_2=0}^{y_3} \sum_{y_1=0}^{y_2} \frac{1}{\binom{y_r+r-1}{r-1}} T(y_1, y_2 - y_1, \cdots, y_r - y_{r-1}) \right) \binom{y_r+r-1}{r-1} f(y_r)\, \theta^r (1-\theta)^{y_r} \\
&\quad = \sum_{y_r \geq 0} \widetilde{T}(y_r)\, f(y_r) \binom{y_r+r-1}{r-1} \theta^r (1-\theta)^{y_r} = \mathbb{E}_\theta \left[\widetilde{T}(Y_r)\, f(Y_r)\right]
\end{aligned}$$

Nous avons donné l'expression de la fonction $\widetilde{T}$ qui ne dépend pas de θ.

10. Nous savons que $T(Z_1, \cdots, Z_r)$ est un estimateur sans biais de θ, donc pour tout $\theta \in \Theta$

$$\mathbb{E}_\theta \left[T(Z_1, \cdots, Z_r)\right] = \theta.$$

Or, par définition de l'espérance conditionnelle,

$$\mathbb{E}_\theta \left[\mathbb{E}_\theta \left[T(Z_1, \ldots, Z_r) \,|\, Y_r\right]\right] = \mathbb{E}_\theta \left[T(Z_1, \cdots, Z_r)\right].$$

Nous en déduisons que pour tout $\theta \in \Theta$

$$\mathbb{E}_\theta \left[\mathbb{E}_\theta \left[T(Z_1, \ldots, Z_r) \,|\, Y_r\right]\right] = \theta,$$

ce qui signifie que $\widetilde{T}(Y_r) := \mathbb{E}_\theta \left[T(Z_1, \ldots, Z_r) \,|\, Y_r\right]$ est un estimateur sans biais de θ, basé sur Y_r.

11. Nous écrivons

$$\begin{aligned}
(T(Z_1, \ldots, Z_r) - \theta)^2 &= (T(Z_1, \ldots, Z_r) - \widetilde{T}(Y_r))^2 + (\widetilde{T}(Y_r) - \theta)^2 \\
&\quad + 2\,(T(Z_1, \ldots, Z_r) - \widetilde{T}(Y_r))\,(\widetilde{T}(Y_r) - \theta). \qquad (5)
\end{aligned}$$

Pour tout $\theta \in \Theta$, nous avons

$$\begin{aligned}
&\mathbb{E}_\theta \left[\left(T(Z_1, \ldots, Z_r) - \widetilde{T}(Y_r)\right) \left(\widetilde{T}(Y_r) - \theta\right) \Big| Y_r \right] \\
&\qquad = \left(\mathbb{E}_\theta \left[T(Z_1, \ldots, Z_r) \,|\, Y_r\right] - \widetilde{T}(Y_r)\right) \left(\widetilde{T}(Y_r) - \theta\right) = 0,
\end{aligned}$$

puisque $\widetilde{T}(Y_r)$ est dans la tribu engendrée par la v.a. Y_r. Ainsi, pour tout $\theta \in \Theta$,

$$\mathbb{E}\left[\left(T(Z_1, \ldots, Z_r) - \widetilde{T}(Y_r)\right) \left(\widetilde{T}(Y_r) - \theta\right) \right] = 0.$$

Nous déduisons de (5) que pour tout $\theta \in \Theta$,

$$\mathbb{E}_\theta \left[(T(Z_1, \ldots, Z_r) - \theta)^2\right] = \mathbb{E}_\theta \left[(T(Z_1, \ldots, Z_r) - \widetilde{T}(Y_r))^2\right] + \mathbb{E}_\theta \left[(\widetilde{T}(Y_r) - \theta)^2\right].$$

12. Nous savons que $\tilde{\theta}_r$ est un estimateur sans biais de θ, basé sur Y_r et donc aussi sur $(Z_1, \cdots, Z_r)$. Supposons qu'il ne soit pas de variance minimale : il existe un autre estimateur $T_\star(Z_1, \cdots, Z_r)$ de θ, sans biais, qui vérifie pour tout $\theta \in \Theta$,

$$\mathbb{E}_\theta\left[T_\star(Z_1, \cdots, Z_r)\right] = \theta \qquad \mathbb{E}_\theta\left[(T_\star(Z_1, \ldots, Z_r) - \theta)^2\right] \leq \mathbb{E}_\theta\left[(\tilde{\theta}_r - \theta)^2\right]. \tag{6}$$

En utilisant les questions 9 et 10, nous savons que $\mathbb{E}_\theta\left[T_\star(Z_1, \cdots, Z_r) \,|\, Y_r\right]$ est une fonction de Y_r indépendante de θ, qui est un estimateur sans biais de θ. D'après la question 8, cela entraine

$$\mathbb{E}_\theta\left[T_\star(Z_1, \cdots, Z_r) \,|\, Y_r\right] = \tilde{\theta}_r.$$

Par suite, d'après la question 11,

$$\mathbb{E}_\theta\left[(T_\star(Z_1, \ldots, Z_r) - \theta)^2\right] = \mathbb{E}_\theta\left[(T_\star(Z_1, \ldots, Z_r) - \tilde{\theta}_r)^2\right] + \mathbb{E}_\theta\left[(\tilde{\theta}_r - \theta)^2\right].$$

La comparaison de cette relation et de (6) donne

$$\mathbb{E}_\theta\left[(T_\star(Z_1, \ldots, Z_r) - \tilde{\theta}_r)^2\right] = 0.$$

Ce qui conclut la démonstration.

Solution . 1. On écrit la définition de $R_e(g)$ en opérant une disjonction des cas sur la valeur de $Y \in \{-1; 1\}$ puis on utilise l'espérance conditionnelle :

$$\begin{aligned}
R_{0-1}(g) &= \mathbb{E}\left[\exp(-Yg(X))(\mathbf{1}_{Y=1} + \mathbf{1}_{Y=-1})\right] \\
&= \mathbb{E}\left[\exp(-Yg(X))(\mathbf{1}_{Y=1} + \mathbf{1}_{Y=-1})\right] \\
&= \mathbb{E}\left[\exp(-g(X))\mathbb{E}[\mathbf{1}_{Y=1}|X] + \exp(g(X))\mathbb{E}[\mathbf{1}_{Y=-1}|X]\right] \\
&= \mathbb{E}\left[\exp(-g(X))\eta(X) + \exp(g(X))(1 - \eta(X))\right] \\
&= \mathbb{E}[C_e(\eta(X), g(X))].
\end{aligned}$$

2. On fixe $\eta \in]0, 1[$ et on considère la fonction $p \longmapsto C_e(\eta, p)$ qui est continue et dérivable. Un rapide calcul montre que le point qui annule la dérivée de $C_e(\eta, .)$ est :

$$p_e(\eta) = \frac{1}{2} \log\left(\frac{\eta}{1 - \eta}\right).$$

3. On déduit que pour x fixé, la valeur de $g(x)$ optimale qui minimise $C_e(\eta(x), \cdot)$ est donnée par :

$$g_e^*(x) = p_e\left(\eta(x)\right).$$

Ainsi, $p_e \circ \eta$ est la règle de décision Bayésienne pour le risque R_e.

4. Pour obtenir l'expression de H_e, on remplace simplement par l'expression de $p_e(\eta)$:

$$\begin{aligned}
H_e(\eta) &= C_e(\eta, p_e(\eta)) \\
&= \eta \exp\left(\frac{1}{2} \log \frac{1 - \eta}{\eta}\right) + (1 - \eta) \exp\left(\frac{1}{2} \log \frac{\eta}{1 - \eta}\right) \\
&= \eta \exp\left(\log \sqrt{\frac{1 - \eta}{\eta}}\right) + (1 - \eta) \exp\left(\log \sqrt{\frac{\eta}{1 - \eta}}\right) \\
&= 2\sqrt{\eta(1 - \eta)}
\end{aligned}$$

5. On observe que :

$$g_e(\eta) \geq 0 \Longleftrightarrow \log \frac{\eta}{1-\eta} \geq 0 \Longleftrightarrow \eta \geq \frac{1}{2}.$$

   Aussi, signe(g_e) est la règle de décision Bayésienne pour le risque R_{0-1}.
6. Adapter les calculs faits en Amphi 8, pour tenir compte du fait que $Y \in \{-1, 1\}$ et non pas $Y \in \{0, 1\}$.
7. On utilise l'inégalité de Jensen avec la convexité de la fonction ψ :

$$\begin{aligned}\psi(R_{0-1}(g) - R^*_{0-1}) &= \psi\left(\mathbb{E}[|2\eta(X) - 1| \mathbb{1}_{\{g(X)(2\eta(X)-1)\leq 0\}}]\right) \\ &\leq \mathbb{E}\left[\psi\left(|2\eta(X) - 1| \mathbb{1}_{\{g(X)(2\eta(X)-1)\leq 0\}}\right)\right] \\ &= \mathbb{E}\left[\psi\left(|2\eta(X) - 1|\right) \mathbb{1}_{\{g(X)(2\eta(X)-1)\leq 0\}}\right] \\ &\quad + \mathbb{E}\left[\psi\left(0\right) \mathbb{1}_{\{g(X)(2\eta(X)-1)> 0\}}\right] \\ &= \mathbb{E}\left[\psi\left(|2\eta(X) - 1|\right) \mathbb{1}_{\{g(X)(2\eta(X)-1)\leq 0\}}\right],\end{aligned}$$

   car $\psi(0) = 0$.
8. Si $\eta \geq \frac{1}{2}$, alors $p \longmapsto C(\eta, p)$ est décroissante sur $\mathbb{R}^-$ puisque la dérivée est négative sur $\mathbb{R}^-$. De même, si $\eta \leq \frac{1}{2}$, $p \longmapsto C(\eta, p)$ est croissante. Ainsi, on conclut que :

$$\forall \eta \in]0, 1[\qquad \inf_{p(2\eta-1)\leq 0} C_e(\eta, p) = C_e(\eta, 0) = 1. \tag{7}$$

9. L'égalité $H_e(\eta) = H_e(1 - \eta)$ découle directement de l'expression démontrée en question 3.
   On démontre la dernière égalité en séparant les cas $\eta \geq \frac{1}{2}$ et $\eta \leq \frac{1}{2}$.
   • Supposons dans un premier temps que $\eta \geq \frac{1}{2}$, alors :

$$\psi(|2\eta - 1|) = \psi(2\eta - 1) = 1 - H_e(\eta).$$

   • A l'inverse, si $\eta \leq \frac{1}{2}$, dans ce cas

$$\psi(|2\eta - 1|) = \psi(1 - 2\eta) = 1 - H_e(1 - \eta) = 1 - H_e(\eta).$$

10. On évalue la fonction ψ_e en 0. On obtient que :

$$\psi_e(0) = 1 - H_e\left(\frac{1}{2}\right) = 1 - 2\sqrt{\frac{1}{2}\frac{1}{2}} = 0.$$

   Pour étudier la convexité de ψ_e, on écrit

$$\psi_e(\theta) = 1 - H_e\left(\frac{1+\theta}{2}\right) = 1 - \sqrt{\frac{1+\theta}{2}\frac{1-\theta}{2}} = 1 - \sqrt{1-\theta^2}.$$

   La fonction précédente est bien entendu convexe (facilement vérifiable par dérivation).
11. On applique la question 7 avec la fonction ψ_e qui est convexe et vérifie $\psi_e(0) = 0$. On obtient alors en utilisant la question 9

$$\begin{aligned}\psi_e(R_{0-1}(g) - R^*_{0-1}) &\leq \mathbb{E}\left[\psi_e\left(|2\eta(X) - 1|\right) \mathbb{1}_{\{g(X)(2\eta(X)-1)\leq 0\}}\right] \\ &\leq \mathbb{E}\left[\left(1 - H_e(\eta(X))\right) \mathbb{1}_{\{g(X)(2\eta(X)-1)\leq 0\}}\right] \\ &= \mathbb{E}\left[\{1 - C_e(\eta(X), p_e(\eta(X)))\} \mathbb{1}_{\{g(X)(2\eta(X)-1)\leq 0\}}\right]\end{aligned}$$

On observe que le premier terme satisfait toujours (voir question 8)

$$\mathbb{1}_{\{g(X)(2\eta(X)-1)\leq 0\}} \leq C_e(\eta(X), g(X))\mathbb{1}_{\{g(X)(2\eta(X)-1)\leq 0\}}.$$

On obtient donc, en utilisant $C(\eta, g) - C(\eta, p(\eta)) \geq 0$

$$\begin{aligned}
\psi_e(R_{0-1}(g) - R^*_{0-1}) &\leq \mathbb{E}\left[\{C_e(\eta(X), g(X)) - C_e(\eta(X), p_e(\eta(X)))\}\ \mathbb{1}_{\{g(X)(2\eta(X)-1)\leq 0\}}\right] \\
&\leq \mathbb{E}\left[C_e(\eta(X), g(X)) - C_e(\eta(X), p_e(\eta(X)))\right] \\
&= R_e(g) - R^*_e\ .
\end{aligned}$$