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 $\mathbf{z}_i$, $i \in \{1,\dots,n\}$. 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

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

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 $i \in \{1,\dots,n\}$, 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 $i \in \{1,\dots,n\}$, 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 $i \in \{1,\dots,n\}$, 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 $\mathbf{u} \in \mathbb{R}^q$,

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

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 $\theta \in \Theta$,
$$\mathbb{E}_\theta\left[(T(Z_1,\ldots,Z_r)-\theta)^2\right] = \mathbb{E}_\theta\left[\left(T(Z_1,\ldots,Z_r)-\widetilde{T}(Y_r)\right)^2\right] + \mathbb{E}_\theta\left[\left(\widetilde{T}(Y_r)-\theta\right)^2\right]$$
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 $\text{signe}(g(x)) \geq 0$ 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 $x \in \mathsf{X}$,
$$g_e(x) := p_e(\eta(x)) .$$

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

$$H_e^-(\eta) := \inf_{p\in\mathbb{R}\ :\, p\,(2\eta-1)\leq 0} C_e(\eta, p)\ , \quad \eta \in]0, 1[\ ,$$

$$\psi_e(\theta) := H_e^-\left(\frac{1+\theta}{2}\right) - H_e\left(\frac{1+\theta}{2}\right) = 1 - H_e\left(\frac{1+\theta}{2}\right), \quad \theta \in [0, 1[\ .$$

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

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