

FEUILLE 2 DE TRAVAUX DIRIGÉS

EXERCICE 1. [*Sélection de modèle avec le critère AIC, sans signal et avec $X = I_n$*]

Considérons le modèle linéaire gaussien suivant

$$Y = \beta^* + \varepsilon,$$

où les ε_i , $1 \leq i \leq n$, sont i.i.d. $\mathcal{N}(0, \sigma^2)$ (σ connu) et $\beta^* \in \mathbb{R}^n$. Cela correspond au cas $X = I_n$ et $p = n$. On considère la famille de modèle $\mathcal{M} = \{m \subset \{1, \dots, p\}\}$.

1. Calculer les estimateurs des moindres carrés $\tilde{\beta}_m$ pour chaque $m \in \mathcal{M}$.

On suppose à présent $\beta^* = 0$ (aucun signal).

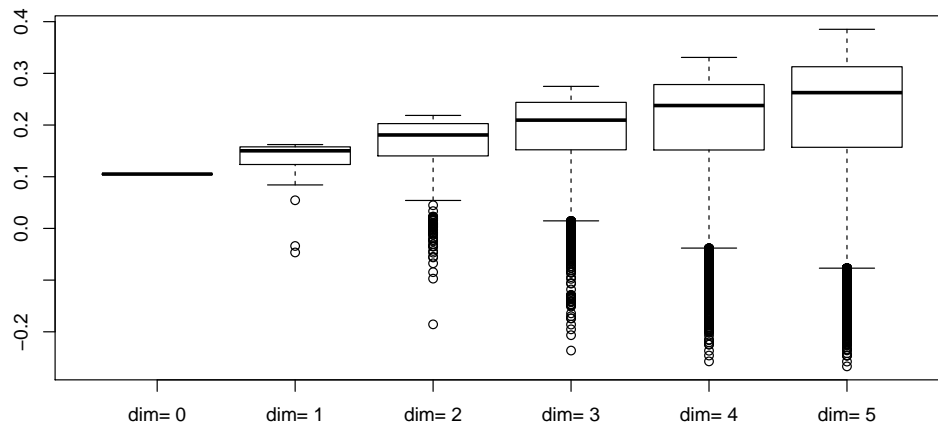
2. Calculer le risque de prédiction $R(\beta^*, \tilde{\beta}_m)$ pour chaque modèle $m \in \mathcal{M}$. Quel modèle faudrait-il choisir ?

3. Calculer

$$\hat{R}_m = n^{-1} \|Y - \tilde{\beta}_m\|^2 + 2|m|\sigma^2/n - \sigma^2$$

et déterminer la loi de $\|Y - \tilde{\beta}_m\|^2$.

4. Soient les familles de points $Z(d) = \{\hat{R}_m, m \in \mathcal{M}, |m| = d\}$ pour $d \in \{0, \dots, 5\}$. Combien y a-t'il de points dans chaque famille ? On donne ci-dessous les boxplots de $Z(d)$, $d \in \{0, \dots, 5\}$ ($n = 35$, $\sigma^2 = 1$). Quel est le modèle sélectionné par le critère AIC ? Commenter.



EXERCICE 2. [*Inégalité de concentration sous exponentielle pour les lois du χ^2*]

On rappelle le résultat de concentration suivant : pour toute fonction $F : \mathbb{R}^q \rightarrow \mathbb{R}$ 1-lipschitzienne et pour $Z \sim \mathcal{N}(0, I_q)$, on a

$$F(Z) - \mathbb{E}F(Z) \leq \sqrt{2\xi}$$

pour une variable $\xi \sim \mathcal{E}(1)$. Montrer que pour toute $U \sim \chi^2(q)$, on a $U \leq (\sqrt{q} + \sqrt{2\xi})^2$.

EXERCICE 3. [*Sélection de modèle dans le cas orthogonal*]

Considérons le modèle linéaire gaussien suivant

$$Y = X\beta^* + \varepsilon,$$

où les ε_i , $1 \leq i \leq n$, sont i.i.d. $\mathcal{N}(0, \sigma^2)$ (σ connu), $\beta^* \in \mathbb{R}^p$ est fixé et X est une matrice réelle $n \times p$. On suppose $p \leq n$ et que les colonnes X_1, \dots, X_p sont **non nulles et orthogonales deux à deux**.

On considère la famille de modèles $\mathcal{M} = \{m \subset \{1, \dots, p\}\}$. Pour chaque $m \in \mathcal{M}$, on considère $\tilde{\beta}_m$ l'estimateur des moindres carrés associé, c'est-à-dire un vecteur $\tilde{\beta}_m \in \mathbb{R}^p$ tel que $X\tilde{\beta}_m = \Pi_{S_m}(Y)$, où Π_{S_m} désigne la projection orthogonale sur $S_m = \text{Vect}(X_j, j \in m)$.

Pour $\lambda > 0$, on choisit

$$\hat{m}_\lambda \in \arg \min_{m \in \mathcal{M}} \left\{ n^{-1} \|Y - X\tilde{\beta}_m\|^2 + \lambda |m| \sigma^2 / n \right\}$$

1. Calculer $\tilde{\beta}_m$ puis $\|Y - X\tilde{\beta}_m\|^2$. En déduire que pour tout $\lambda > 0$,

$$\|Y - X\tilde{\beta}_m\|^2 + \lambda |m| \sigma^2 = \|Y\|^2 - \sum_{j \in m} ((X_j^t Y / \|X_j\|)^2 - \lambda \sigma^2).$$

2. Montrer que \hat{m}_λ peut être obtenu par "seuillage dur" :

$$\hat{m}_\lambda = \left\{ 1 \leq j \leq p \text{ tel que } \left(\frac{X_j^t Y}{\sigma \|X_j\|} \right)^2 \geq \lambda \right\}.$$

On suppose désormais $\beta^* = 0$ (aucun signal). Le modèle idéal est donc $m^* = \emptyset$ et on cherche à voir si $|\hat{m}_\lambda|$ est grand ou pas.

3. Calculer $\mathbb{E}|\hat{m}_\lambda|$ en fonction de $2\bar{\Phi}(\lambda^{1/2})$ pour $\bar{\Phi}(z) = \mathbb{P}(Z > z)$, $Z \sim \mathcal{N}(0, 1)$.

4. Que dire du modèle sélectionné par le critère AIC ?

5. Pour $\lambda = 2K \log p$, montrer que lorsque p et n tendent vers l'infini,

$$\mathbb{E}|\hat{m}_\lambda| \sim \frac{p^{1-K}}{(\pi K \log p)^{1/2}}$$

Que dire du choix $K < 1$?

EXERCICE 4. [*Estimateur des moindres carrés pénalisé comme minimum d'une fonction*]
 Dans le cadre général d'un modèle de régression linéaire gaussien de grande dimension, on considère l'estimateur des moindres carrés $\tilde{\beta}_m$ sur chaque modèle $m \in \mathcal{M}$, pour $\mathcal{M} = \{m \subset \{1, \dots, p\}\}$. Choisissons, pour tout $\lambda > 0$,

$$\hat{m} \in \arg \min_{m \in \mathcal{M}} \{n^{-1} \|Y - X\tilde{\beta}_m\|^2 + \lambda |m| \sigma^2 / n\}.$$

1. Montrer que $\|\tilde{\beta}_{\hat{m}}\|_0 = |\hat{m}|$.
2. Montrer que

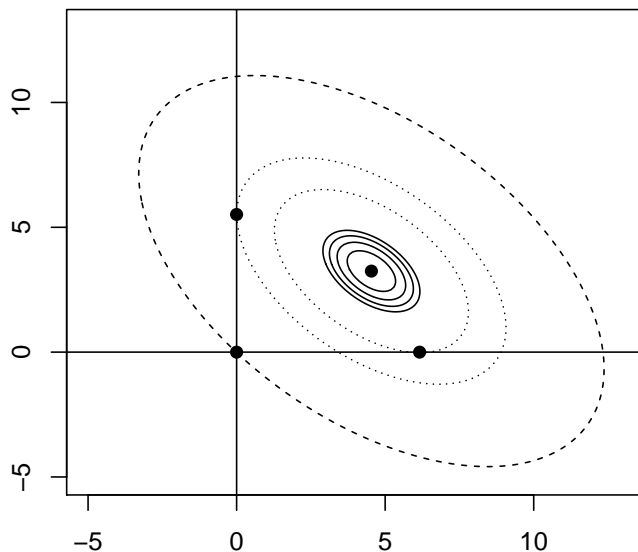
$$\tilde{\beta}_{\hat{m}} \in \arg \min_{\beta \in \mathbb{R}^p} F(\beta), \quad \text{où } F(\beta) = n^{-1} \|Y - X\beta\|^2 + \lambda \|\beta\|_0 \sigma^2 / n.$$

EXERCICE 5. [*Représentation graphique de $\tilde{\beta}_{\hat{m}}$ en dimension $p = n = 2$*]

On considère le modèle linéaire gaussien $Y = X\beta^* + \varepsilon$ avec $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_2)$, $\beta \in \mathbb{R}^2$ et une matrice X de la forme

$$X = \begin{pmatrix} 1 & \cos \theta \\ 0 & \sin \theta \end{pmatrix}, \quad \theta \in]0, \pi/2]$$

1. Calculer le minimum de la fonction $\beta \in \mathbb{R}^2 \mapsto \|Y - X\beta\|^2$?
2. A quoi ressemblent les lignes de niveaux de la fonction $\beta \in \mathbb{R}^2 \mapsto \|Y - X\beta\|^2$?
3. Pour $\theta = \pi/3$ et $\beta^* = (5, 3)$, la figure ci-dessous représente certaines lignes de niveaux de $\beta \mapsto \|Y - X\beta\|^2$. Placer $\tilde{\beta}_m$ pour chaque modèle $m \in \mathcal{M} = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$.
4. Posons $\hat{m}_\lambda \in \arg \min_{m \in \mathcal{M}} \{n^{-1} \|Y - X\tilde{\beta}_m\|^2 + \lambda |m| \sigma^2 / n\}$. Comment évolue $\tilde{\beta}_{\hat{m}_\lambda}$ lorsque λ croît de 0 à l'infini ?
5. Pour chaque $r \in \{0, 1, 2\}$, décrire $\arg \min_{\beta \in B_0(r)} \{\|Y - X\beta\|^2\}$ pour $B_0(r) = \{\beta \in \mathbb{R}^2 \mid \|\beta\|_0 \leq r\}$ la "boule" L^0 de rayon r . Comment évolue cet arg min lorsque r augmente ?
6. A l'aide de l'exercice 6, réinterpréter $\tilde{\beta}_{\hat{m}_\lambda}$ comme le minimum d'une fonction F .
7. Lorsque p est grand, que penser de la complexité de ce problème de minimisation ?



EXERCICE 6. [*Identifiabilité sous hypothèse de sparsité*]

Pour $p \geq 3$ et $n \geq 2$, on note X la matrice $n \times p$ qui a pour j -ème colonne

$$X_j = (\cos(2\pi j/p), \sin(2\pi j/p), 0, \dots, 0).$$

On considère le modèle linéaire donné par la loi de $Y = X\beta^* + \varepsilon$, où $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$ avec σ connu et $\beta^* \in \mathbb{R}^p$ inconnu.

1. montrer que le modèle n'est pas identifiable en β^* .
2. montrer qu'en supposant que β^* vérifie $\|\beta^*\|_0 \leq 1$ (i.e., $s_{max} = 1$ avec les notations du cours), le modèle devient identifiable en β^* .
3. Plus généralement, pour un entier $s_{max} \leq p/2$, montrer que pour une matrice X $n \times p$ quelconque, le modèle devient identifiable en β^* sous l'hypothèse de sparsité $\|\beta^*\|_0 \leq s_{max}$, dès que $X^{(m)}$ est de plein rang pour tous les sous-modèles $m \subset \{1, \dots, p\}$ avec $|m| = 2s_{max}$.

EXERCICE 7. [*Constante d'isométrie restreinte pour X*]

Soit X une matrice $n \times p$ quelconque. Pour chaque $1 \leq j \leq p$, la j -ième colonne de X est notée X_j et est normalisée de la façon suivante : $\|X_j\|^2 = 1$. On définit la constante d'isométrie restreinte pour X par

$$\kappa(s) = \min_{\beta \in \mathbb{R}^p \setminus \{0\}, \|\beta\|_0 \leq s} \left\{ \frac{\|X\beta\|}{\|\beta\|} \right\}$$

pour tout entier $s \geq 1$ plus petit que p .

1. Calculer $\kappa(1)$. Faire le lien entre $\kappa(p)$ et la plus petite valeur propre de $X^t X$.
2. Montrer que le modèle linéaire gaussien correspondant à X est identifiable en β^* si et seulement si $\kappa(p) > 0$.
3. Sous l'hypothèse de sparsité $\|\beta^*\|_0 \leq s_{max}$ ($s_{max} \leq p/2$), montrer que le modèle est identifiable en β^* si $\kappa(2s_{max}) > 0$.
4. En développant $\|X\beta\|^2$, montrer qu'en posant

$$\rho = \max_{j \neq j'} |X_j^t X_{j'}|,$$

on a $(\kappa(s))^2 \geq 1 - \rho(s-1)$ pour tout $1 \leq s \leq p$. Comment interpréter ce résultat ?