Modèle linéaire gaussien

TD3-MAPI3

2016-2017

Exercice 1

On considère le modèle linéaire

$$Y = X\beta^* + \epsilon$$

où X est de taille $n \times p$ et est égale à

$$\begin{pmatrix} I_p \\ 0_{n-p,p} \end{pmatrix},$$

avec $0_{n-p,p}$ la matrice nulle de taille $(n-p) \times p$. Montrer que les composantes de $\hat{\beta}$, l'estimateur par moindres carrés vu en cours, sont indépendantes.

Exercice 2

On observe n variables aléatoires *iid* $Y_1, ..., Y_n$ de lois $\mathcal{N}(\mu, \sigma^2)$, avec μ et σ^2 inconnus.

1) Construire un intervalle de confiance pour μ, de niveau 95%. On pourra écrire $Y = (Y_1, ..., Y_n)^t$ comme le vecteur observé d'un modèle linéaire et utiliser l'intervalle de confiance vu en cours.

2) Quelles sont les valeurs des bornes de cet intervalle de confiance lorsque $n = 9$ et $y = (0, 0, 0, 0, 1, 2, 2, 2, 2)^t$? On pourra s'aider de la table de probabilité de la loi de Student.

Exercice 3

Cet exercice à pour but d'étudier l'effet de la surparamétrisation sur la qualité des estimateurs dans le modèle linéaire. On dispose de deux variables explicatives $X_1 = (x_{1,1}, ..., x_{1,n})^t$ et $X_2 = (x_{2,1}, ..., x_{2,n})^t$ et on suppose que la réponse Y dépend uniquement de X_1, c'est à dire que le “vrai” modèle linéaire s'écrit

$$Y = X_1\beta_1^* + \epsilon, \tag{1}$$

tel que vu en cours. En revanche, on ignore que la variable X_2 n'est pas influente, et on ajuste le modèle linéaire surparamétré

$$Y = X_1\beta_1^* + X_2\beta_2^* + \epsilon, \tag{2}$$

où l'information $\beta_2^* = 0$ n'est pas accessible. On suppose, pour simplifier les calculs que, pour $i = 1, 2$, $\sum_{j=1}^n x_{i,j} = 0$ et $\sum_{j=1}^n x_{i,j}^2 = 1$. On pose $\rho = \sum_{j=1}^n x_{1,j}x_{2,j}$.

1) Montrer que $|\rho| \leq 1$. On suppose pour la suite que $|\rho| < 1$.

2) On note $\hat{\beta}_1$ et $\hat{\beta}_2$ les estimateurs par moindres carrés de β_1^* et β_2^* dans (2). Calculer ces estimateurs et donner leurs moyennes et variances.

On pourra utiliser

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

3) On note $\tilde{\beta}_1$ l'estimateur par moindres carrés de β_1^* dans (1). Calculer cet estimateur et donner sa moyenne et sa variance.

4) Comparer $\mathbb{E}\left([\hat{\beta}_1 - \beta_1^*]^2\right)$ et $\mathbb{E}\left([\tilde{\beta}_1 - \beta_1^*]^2\right)$ et interpréter cette comparaison.