Exercices

Econométrie 2

Exercice 1.

(*) Soit $y_1, \cdots, y_n$ des réels. Déterminer le réel $\widehat{m}$ qui minimise $SCR(m) = \sum_{i=1}^{n} |y_i - m|^2$. Que représente $\widehat{m}$ par rapport à $y_1, \cdots, y_n$?

Exercice 2.

(**) Soit $y_1, \cdots, y_n$ des réels. Déterminer le réel $\tilde{m}$ qui minimise la somme des valeurs absolues $SVA(m) = \sum_{i=1}^{n} |y_i - m|$ (on pourra traiter d'abord le cas où $n = 2p+1$ et classer les réels (y_i) sous la forme $y_{(1)} \leq y_{(1)} \leq \ldots \leq y_{(n-1)} \leq y_{(n)}$, puis passer au cas où $n = 2p$). Que représente $\tilde{m}$ par rapport à $y_1, \cdots, y_n$?

Exercice 3.

(*) Soit $(x_1, y_1), \cdots, (x_n, y_n)$ des couples de réels. Déterminer le réel $\widehat{a}$ qui minimise la somme des carrés résiduelle $SCR(a) = \sum_{i=1}^{n} |y_i - a\, x_i|^2$. Que représente $\widehat{a}$ par rapport à $y_1, \cdots, y_n$? Comparer avec la valeur $\widehat{\beta}$ obtenue dans le cadre de la régression linéaire simple classique.

Exercice 4.

(*) Soit Y qui suit un modèle linéaire avec les ε_i des v.a.i.i.d. centrées de variance constante. Soit $T \in \mathbb{R}^n$ un vecteur déterministe. Montrer que

$$\mathbb{E}\left(\|T - Y\|^2\right) = n\sigma^2 + \|T - X\theta\|^2.$$

Exercice 5.

(*) Soit $Y = (Y_i)_{1 \leq i \leq n}$ qui suit un modèle linéaire de régression multiple avec les ε_i des v.a. et la condition **(A0)**. On note $(\widehat{\varepsilon}_i)_{1 \leq i \leq n}$ les résidus. Montrer que $\frac{1}{n} \sum_{i=1}^{n} \widehat{\varepsilon}_i = 0$.

Exercice 6.

(**) [Maximum de vraisemblance] On se place dans le cadre général du modèle linéaire vectoriel de taille $n \geq p+1$,

$$Y = X\,\theta + \varepsilon, \qquad \varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n), \tag{1}$$

où X est une matrice $(n, p+1)$ de rang plein et $\theta \in \mathbb{R}^{p+1}$.

i. Déterminer les estimateurs de θ et σ^2 par maximum de vraisemblance.

ii. Déterminer la statistique du test du rapport de vraisemblance pour le problème de test H_0: ${}^tC\theta = 0$ contre H_1: ${}^tC\theta \neq 0$, où C est un vecteur de $\mathbb{R}^n$. Retrouve-t-on la statistique de Student?

Exercice 7.

(**) [Théorème de Gauss-Markov] On se propose de montrer dans cet exercice que, pour Y qui suit le modèle linéaire $Y = Z\,\theta^* + \varepsilon$, sous **(A0)-(A3)**, l'estimateur des moindres carrés $\widehat{\theta}$ est optimal parmi les estimateurs linéaires sans biais.

L'optimalité veut dire que si $\widetilde{\theta}$ est un autre estimateur linéaire sans biais :

$$\mathrm{cov}\,(\widetilde{\theta}) - \mathrm{cov}\,(\widehat{\theta}) \text{ est une matrice semi-définie positive,}$$

ou encore, ce qui est équivalent, que pour toute combinaison linéaire ${}^tC\,\theta$ des paramètres,

$$\mathrm{Var}\,({}^tC\,\widetilde{\theta}) \geq \mathrm{Var}\,({}^tC\,\widehat{\theta}).$$

i. Posons $\widetilde{\theta} = M\,Y$ où M est une matrice de taille $(p+1, n)$. Montrer que $M\,Z = I_n$.

ii. Ecrire $\widehat{\theta} = T\,P_{[Z]}Y$, et montrer que $M\,P_{[Z]} = T\,P_{[Z]}$.

iii. Montrer que $\widetilde{\theta} = \widehat{\theta} + M\,P_{[Z]^\perp}Y$, la somme étant non-corrélée. Conclure.

Exercice 8.

(**) Soit θ_1 et θ_2 deux paramètres réels inconnus et soit :

- Y_1 un estimateur sans biais de $\theta_1 + \theta_2$ et de variance σ^2;
- Y_2 un estimateur sans biais de $2\theta_1 - \theta_2$ et variance $4\sigma^2$;
- Y_3 un estimateur sans biais de $6\theta_1 + 3\theta_2$ et de variance $9\sigma^2$,

les estimateurs Y_1, Y_2 et Y_3 étant indépendants. Quels estimateurs de θ_1 et θ_2 proposeriez-vous ? (on pourra utiliser l'exercice précédent).

Exercice 9.

(***) [Moindres carrés généralisés] On suppose le modèle linéaire général mais seulement sous l'hypothèse **(A1)** (erreurs centrées). Cependant, on suppose également connue la matrice de variance-covariance des erreurs (qui n'est donc pas forcément diagonale), qui sera notée Σ, que l'on suppose de rang n.

i. Déterminer alors l'espérance et la matrice de variance-covariance de $\widehat{\theta}$, estimateur des moindres carrés que l'on appeler ici moindres carrés ordinaires.

ii. On considère maintenant, en lieu et place de la distance euclidienne classique dans $\mathbb{R}^n$ utilisée pour les moindres carrés ordinaires, la distance définie par la norme :

$$\|U-V\|_\Sigma = (U-V)'\,\Sigma^{-1}\,(U-V).$$

L'estimateur $\widehat{\theta}_G$ de θ par moindres carrés généralisés minimisera $\|Y - X\,\theta\|_\Sigma$. Montrer que $\widehat{\theta}_G = (X'\,\Sigma^{-1}\,X)^{-1}X'\,\Sigma^{-1}\,Y$, et en déduire son espérance et sa matrice de variance-covariance.

iii. En déduire que l'estimateur $\widehat{\theta}_G$ a une matrice de variance-covariance "plus petite" que $\widehat{\theta}$ (c'est-à-dire que la différence entre leur matrice de variance-covariance est une matrice semi-définie positive). Ce résultat est une variante du Théorème de Gauss-Markov vu à l'exercice précédent

iv. En supposant de plus que les observations sont conjointement gaussiennes, montrer que $\widehat{\theta}_G$ est l'estimateur du maximum de vraisemblance.

Exercice 10.

(*) [Moindres carrés pondérés] On suppose que l'on ne veut pas accorder la même importance à toutes les observations dans un modèle de régression multiple. Pour simplifier, on se place dans le cadre du modèle linéaire (1) avec les ε_i des v.a.i.i.d. centrées de variance constante. Soit $(p_1, \ldots, p_n)$ les poids respectifs que l'on accorde aux différentes variables, avec $0 < p_i$ pour $i = 1, \ldots, n$ et $\sum p_i = n$. Soit Ω la matrice diagonale composée de $(p_1, \ldots, p_n)$ sur la diagonale. Tout comme pour les moindres carrés généralisés, on considérera la norme :

$$\|U-V\|_\Omega = (U-V)'\,\Omega^{-1}\,(U-V).$$

et $\widehat{\theta}_\Omega$, l'estimateur qui minimise $\|Y - X\theta\|_\Omega$ Déterminer l'expression de l'estimateur $\widehat{\theta}_\Omega$ de θ, son espérance et sa variance, ainsi que l'expression de l'estimateur $\widehat{\sigma}^2_\Omega$ de σ^2 la variance des erreurs.

Exercice 11.

(**) [Test de runs] Ce test est utilisé pour tester la présence ou non de corrélations dans les ε_i. On commence d'abord par le décrire dans le cas où l'on observe des variables aléatoires $Y_1, \ldots, Y_n$ dont on veut tester l'indépendance. On les suppose de médiane zéro. On compte parmi Y_1, s, Y_n le nombre R de "paquets" (ou "runs") de (Y_i) consécutifs ayant le même signe.
Par exemple, si $Y_1, \ldots, Y_9 = (1.1, 1.3, -2, -1, 4.5, 1.6, -2.7, -1.3, 4)$, il y a 5 runs pour $n = 9$ données.

i. Montrer que si on suppose qu'aucun des Y_i n'est nul, alors :

$$R = 1 + \sum_{i=1}^{n-1} \mathbb{I}_{Y_i Y_{i+1} < 0} := 1 + \sum_{i=1}^{n-1} Z_i$$

ii. On suppose que les Y_i sont indépendantes et de loi diffuse (c'est-à-dire absolument continue par rapport à la mesure de Lebesgue). Montrer que $\mathbb{E}\,(R) = \dfrac{n+1}{2}$,

iii. Montrer que si $|i-j|>1$, Z_i et Z_j sont indépendantes. Montrer que Z_i et Z_{i+1} sont également indépendantes. En déduire $\mathrm{Var}\,(R)$.

iv. En utilisant le théorème de la limite centrale, construire pour des grands échantillons une statistique libre qui suit une loi normale centrée réduite sous l'hypothèse H_0 d'indépendance et qui tend vers $\pm\infty$ sous les alternatives H_1 d'intrication et de répulsion. Nous laissons au lecteur le soin de deviner le sens de ces deux derniers mots.

Exercice 12.

(**) Soit le modèle de régression linéaire simple de Y par Z,

$$Y_i = \mu + \beta\, Z_i + \varepsilon_i \qquad \text{pour } i = 1, \cdots, n$$

avec des paramètres μ, β et σ^2 inconnus et les ε_i qui sont des v.a.i.i.d. centrées gaussiennes. Cependant, croyant plutôt que le modèle est quadratique, on utilise les estimateurs $\widehat{\theta}$ et $\widehat{\gamma}^{2,n}$ du modèle de régression polynômiale de degré 2 (de paramètres $\theta = (\theta_0, \theta_1, \theta_2)'$ et γ^2). On dira alors que l'on a sur-ajusté le modèle. A-t-on $\mathbb{E}\,(\widehat{\theta_2}^{\,n}) = 0$; $\widehat{\theta_2}^{\,n} \xrightarrow[n\to+\infty]{\mathcal{P}} 0$; $\widehat{\gamma}^{2,n} \xrightarrow[n\to+\infty]{\mathcal{P}} \sigma^2$ quand le nombre de données n tend vers l'infini ? Si z est une valeur donnée de la variable Z, quelle est la variance de $\widehat{Y}$ en z calculée à l'aide de l'estimateur $\widehat{\theta}$? Comparer au cas de la régression simple.

Exercice 13.

(**) A l'inverse de l'exercice précédent, on suppose maintenant que l'on a un modèle de régression polynômiale de degré 2 tel que

$$Y_i = \theta_0 + \theta_1\, Z_i + \theta_2\, Z_i^2 + \varepsilon_i \qquad \text{pour } i = 1, \cdots, n$$

avec des paramètres $(\theta_0, \theta_1, \theta_2)$ inconnus et les ε_i qui sont des v.a.i.i.d. centrées gaussiennes. Cependant, pour analyser ces données, on utilise un modèle de régression simple. On dira alors que l'on a sous-ajusté le modèle. En se plaçant dans le cas particulier où $Z_i = i$ pour $i = 1, \ldots, n$, vérifier que l'estimateur de la variance ne converge pas et que $|\widehat{Y}_{n+1} - Y_{n+1}| \xrightarrow[n\to+\infty]{\mathcal{P}} +\infty$.