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 $y_1, \cdots, y_n$?

Proof. On a $SCR'(m) = 2\sum_{i=1}^{n}(m - y_i)$ et $SCR''(m) = 2 \cdot n$. Donc $SCR'(\widehat{m}) = 0 \Longleftrightarrow \widehat{m} = \frac{1}{n}\sum_{i=1}^{n} y_i$ et $SCR''(m) > 0$ pour tout $m \in \mathbb{R}$: $\widehat{m}$ est un minimum local d'une fonction convexe, c'est donc aussi un minimum global. De plus $\widehat{m}$ est la moyenne empirique des (y_i). □

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 $y_1, \cdots, y_n$?

Proof. 1. pour $n = 2p + 1$, on montre que $\tilde{m} = y_{(p+1)}$. En effet, dans un tel cas

$$SVA(y_{(p+1)}) = \sum_{i=1}^{p}(y_{(p+1)} - y_{(i)}) + \sum_{i=p+2}^{2p+1}(y_{(i)} - y_{(p+1)}) = \sum_{i=p+2}^{2p+1} y_{(i)} - \sum_{i=1}^{p} y_{(i)}.$$

Pour $m \in]y_{(j)}, y_{(j+1)}]$ et $j = p + 1, \ldots, 2p + 1$ (avec, par convention, $y_{(2p+2)} = +\infty$ et les sommes vides nulles),

$$\begin{aligned}
SVA(m) &= \sum_{i=1}^{j}(m - y_{(i)}) + \sum_{i=j+1}^{2p+1}(y_{(i)} - m) \\
&= \sum_{i=1}^{p}(m - y_{(i)}) + \sum_{i=p+2}^{2p+1}(y_{(i)} - m) + \sum_{i=p+1}^{j}(y_{(i)} - m) - \sum_{i=p+2}^{j}(y_{(i)} - m) \\
&= \sum_{i=p+2}^{2p+1} y_{(i)} - \sum_{i=1}^{p} y_{(i)} + 2\sum_{i=p+2}^{j}(m - y_{(i)}) + (m - y_{(p+1)}) \\
&> SVA(y_{(p+1)}).
\end{aligned}$$

Par symétrie, il en est de même si $m \in [y_{(j)}, y_{(j+1)}[$ et $j = 0, \ldots, p$. On en déduit que le minimum de $SVA(.)$ est atteint en $\tilde{m} = y_{(p+1)}$, qui est la médiane des (y_j).

2. Si $n = 2p$, on déduit de ce qui précède que le minimum de SVA doit se trouver entre $y_{(p)}$ et $y_{(p+1)}$. En effet, pour $m \in]y_{(p)}, y_{(p+1)}]$,

$$\begin{aligned}
SVA(m) &= \sum_{i=1}^{p}(m - y_{(i)}) + \sum_{i=p+1}^{2p}(y_{(i)} - m) \\
&= \sum_{i=p+1}^{2p} y_{(i)} - \sum_{i=1}^{p} y_{(i)},
\end{aligned}$$

qui est indépendant de m. Maintenant, pour $m \in [y_{(j)}, y_{(j+1)}[$ et $j = p+1, \ldots, 2p+1$, on a :

$$\begin{aligned}
SVA(m) &= \sum_{i=1}^{j}(m - y_{(i)}) + \sum_{i=j+1}^{2p+1}(y_{(i)} - m) \\
&= \sum_{i=1}^{p}(m - y_{(i)}) + \sum_{i=p+1}^{2p}(y_{(i)} - m) + 2\sum_{i=p+1}^{j}(m - y_{(i)}) \\
&> \sum_{i=p+1}^{2p} y_{(i)} - \sum_{i=1}^{p} y_{(i)}.
\end{aligned}$$

On retrouve bien que le minimum est atteint entre $y_{(p)}$ et $y_{(p+1)}$, donc en la médiane des (y_i). □

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.

Proof. On a $SCR'(a) = 2\sum_{i=1}^{n} x_i(a \cdot x_i - y_i)$ et $SCR''(a) = 2\sum_{i=1}^{n} x_i^2$. Donc $SCR'(\widehat{a}) = 0 \Longleftrightarrow \widehat{a} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$ et $SCR''(a) > 0$ pour tout $a \in \mathbb{R}$: $\widehat{a}$ est un minimum local d'une fonction convexe, c'est donc aussi un minimum global. On remarque que si la série a été recentrée (à savoir que l'on a retranché la moyenne des x_i à chacun des x_i, et la moyenne des y_i à chacun des y_i), alors $\widehat{a} = \widehat{\beta}$, pente de la régression simple. □

Exercice 4.

(*) Soit $Y = (Y_i)_{1 \le i \le n}$ qui suit un modèle linéaire $Y = Z\,\theta^* + \varepsilon$, avec les ε_i des v.a. centrées de variance constante. Soit $T \in \mathbb{R}^n$ un vecteur déterministe. Montrer que

$$\mathbb{E}\left[\|T - Y\|^2\right] = n\,\sigma^2 + \|T - Z\,\theta^*\|^2.$$

Proof. On a :

$$\begin{aligned}
\mathbb{E}\left[|Y - T\|^2\right] &= \mathbb{E}\left[\|(Y - Z\,\theta) + (Z\,\theta^* - T)\|^2\right] \\
&= \mathbb{E}\left[\|Y - Z\,\theta^*\|^2 + \|T - Z\,\theta^*\|^2 + 2\,{}^t(Z\,\theta^* - T)(Y - Z\,\theta^*)\right] \\
&= \mathbb{E}\left[\|\varepsilon\|^2\right] + \|T - Z\,\theta^*\|^2 + 2\,{}^t(Z\,\theta^* - T)\mathbb{E}\left[\varepsilon\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[\varepsilon_i^2\right] + \|T - Z\,\theta^*\|^2 \\
&= n\,\sigma^2 + \|T - Z\,\theta^*\|^2
\end{aligned}$$

□

Exercice 5.

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

Proof. Sous l'hypothèse **(A0)** on a $\widehat{\varepsilon} = P_{[Z]^\perp}\varepsilon$. On note $\mathbb{1} = {}^t(1,\dots,1)$. On sait que $\mathbb{1} \in [Z]$ donc $\mathbb{1} \perp [Z]^\perp$. Or $\frac{1}{n}\sum_{i=1}^n \widehat{\varepsilon}_i = \frac{1}{n}\langle \mathbb{1}\,,\,\widehat{\varepsilon}\rangle$. Mais comme $\mathbb{1} \perp [Z]^\perp$ alors $\langle \mathbb{1}\,,\,P_{[Z]^\perp}\varepsilon\rangle = 0$ donc $\frac{1}{n}\sum_{i=1}^n \widehat{\varepsilon}_i = 0$. $\square$

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?

Proof. i. En notant $f_{\theta,\sigma^2}(Y)$ la densité de $Y = (Y_1,\dots,Y_n)'$, comme $\mathbb{E}\,(Y) = X\,\theta$ et $\mathrm{Var}\,(Y) = \sigma^2\,I_n$ et comme le bruit est supposé gaussien, alors $f_{\theta,\sigma^2}(Y) = \dfrac{1}{(2\pi\,\sigma^2)^{n/2}}\,\exp\Big\{-\dfrac{1}{2\sigma^2}\,(Y-X\,\theta)'\,(Y-X\,\theta)\Big\}$. Par suite, avec pour $Y \in \mathbb{R}^n$, $L(\theta,\sigma^2) = -2\times\log(f_{\theta,\sigma^2}(Y))$, on a :

$$L(\theta,\sigma^2) = n\,\log(2\pi) + n\,\log(\sigma^2) + \frac{1}{\sigma^2}\,(Y-X\,\theta)'\,(Y-X\,\theta).$$

Pour déterminer un estimateur du maximum de vraisemblance des $p+2$ paramètres, on peut chercher un minimum local de $L(\theta,\sigma^2)$, c'est-à-dire un point critique de cette fonction. On obtient ainsi, en différentiant par rapport à θ et à σ^2 :

$$\begin{cases} \dfrac{\partial L}{\partial\theta}(\theta,\sigma^2) = -\dfrac{1}{\sigma^2}\Big(X'\,(Y-X\,\theta) + (Y-X\,\theta)'\,X\Big) = -\dfrac{2}{\sigma^2}\,X'\,(Y-X\,\theta) \\ \dfrac{\partial L}{\partial\sigma^2}(\theta,\sigma^2) = \dfrac{n}{\sigma^2} - \dfrac{1}{\sigma^4}\,(Y-X\,\theta)'\,(Y-X\,\theta) \end{cases}$$

(noter que la différentiation par rapport à θ se fait matriciellement) et ainsi, en résolvant $\dfrac{\partial L}{\partial\theta}(\theta,\sigma^2) = 0$ et $\dfrac{\partial L}{\partial\sigma^2}(\theta,\sigma^2) = 0$, on obtient :

$$\begin{pmatrix} \tilde\theta \\ \tilde\sigma^2 \end{pmatrix} = \begin{pmatrix} \widehat\theta \\ \frac{1}{n}\,\|Y - X\,\tilde\theta\|^2 \end{pmatrix}$$

(on vérifie également que la matrice hessienne de l'application $L(\theta,\sigma^2)$ est définie positive et donc que l'on a bien un minimum). Seul l'estimateur de la variance est différent de celui obtenu par la méthode des moindres carrés.

ii. Soit le problème de test $\begin{cases} H_0: & {}^tC\,\theta = 0 \\ H_1: & {}^tC\,\theta \neq 0 \end{cases}$. Notons Θ_0 le s.e.v. de $\mathbb{R}^{p+1}$ tel que $\Theta_0 = \big\{\theta \in \mathbb{R}^{p+1},\ {}^tC\,\theta\big\} = \{C\}^\perp$ et $\Theta_1 = \mathbb{R}^{p+1}\setminus\Theta_0$. Pour simplifier le problème, on supposera que C est un vecteur tel que ${}^tC\,C = \big\|C\big\|^2 = 1$.
La statistique du rapport de vraisemblance pour ce problème de test s'écrit: $\qquad \widehat{LR} = \dfrac{\sup_{\theta\in\Theta_0,\sigma^2>0} L(\theta,\sigma^2)}{\sup_{\theta\in\Theta_1,\sigma^2>0} L(\theta,\sigma^2)}$.
Notons $\{C\}^\perp = \big[(C_1,\dots,C_p)\big]$, sev engendré par les p vecteurs C_i, que l'on peut choisir comme une base orthonormale. On peut écrire que $\theta = ({}^tC\,\theta)\,C + \sum_{k=1}^p {}^tC_k({}^tC_k\,\theta$. Donc sous H_0, comme $({}^tC\,\theta) = 0$, $X\,\theta = X\left(\sum_{k=1}^p C_k{}^tC_k\right)\theta$. Notons $X^{(0)}$ la matrice de taille (n,p) et de rang p engendrée par $X\left(\sum_{k=1}^p C_k{}^tC_k\right)$. On en déduit que sous H_0 le modèle s'écrit:

$$Y = X^{(0)}\,\theta^{(0)} + \varepsilon, \quad \text{où } [X^{(0)}] \text{ est un sev de } [X] \text{ de dimension } p \text{ et } \theta^{(0)} \in \mathbb{R}^p.$$

En conséquence, pour $\sup_{\theta\in\Theta_0,\sigma^2>0} L(\theta,\sigma^2) = L(\widehat\theta^{(0)},\widehat\sigma^{(0)2})$, avec $\widehat\theta^{(0)} = ({}^tX^{(0)}\,X^{(0)})^{-1\,t}X^{(0)}Y$ et $\widehat\sigma^{(0)2} = \frac{1}{n}\,\|X^{(0)}\,\widehat\theta^{(0)}\|^2$.
Ceci entraîne que

$$\sup_{\theta\in\Theta_0,\sigma^2>0} L(\theta,\sigma^2) = \frac{1}{(2\pi)^{n/2}}\,\frac{1}{(\widehat\sigma^{(0)2})^{n/2}}\,e^{-n/2}.$$

On a également $\sup_{\theta\in\Theta_0,\sigma^2>0} L(\theta,\sigma^2) = \frac{1}{(2\pi)^{n/2}}\frac{1}{(\widehat{\sigma}^2)^{n/2}} e^{-n/2}$, car prendre le sup sur $(\Theta_1, \sigma^2 > 0)$ c'est prendre le sup sur $\mathbb{R}^{p+1}\times]0,\infty[$ du fait que Θ_0 est de mesure nulle dans $\mathbb{R}^{p+1}$. Par conséquent:

$$\widehat{LR} = \Big(\frac{\widehat{\sigma}^{(0)2}}{\widehat{\sigma}^2}\Big)^{-n/2}.$$

Telle quelle, on ne retrouve pas du tout la statistique de Student. Mais sous H_0,

$$\frac{\widehat{\sigma}^{(0)2}}{\widehat{\sigma}^2} - 1 = \frac{\widehat{\sigma}^{(0)2} - \widehat{\sigma}^2}{\widehat{\sigma}^2} = \frac{1}{n}\frac{\|P_{[X^{(0)}]^\perp}\varepsilon\|^2 - \|P_{[X]^\perp}\varepsilon\|^2}{\widehat{\sigma}^2} = \frac{1}{n}\frac{\|P_{[X^{(0)}]^\perp\cap[X]}\varepsilon\|^2}{\widehat{\sigma}^2}$$

en utilisant Pythagore. Or $[X^{(0)}]^\perp \cap [X]$ est un sev de dimension 1 dans $\mathbb{R}^n$. On peut montrer que $\frac{\left({}^tC\,\varepsilon\right)^2}{(C\,{}^tX\,X\,{}^tC)^{-1}} = \|P_{[X^{(0)}]^\perp\cap[X]}\varepsilon\|^2$: on retrouve le carré de la statistique de Student (on rappelle que si E est un sev de $\mathbb{R}^n$ ayant une base orthonomrale $(e_1,\ldots,e_k)$, la projection de u sur E est égale à $\sum_{i=1}^k < e_i, u > \, e_i = \sum_{i=1}^k ({}^te_i\,u)\,e_i$).

□

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 $M\,Z = I_{p+1}$.

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.

Proof. (i) $\widetilde{\theta}$ est sans biais donc $\forall \theta^* \in \mathbb{R}^{p+1}$, $\mathbb{E}\,[\widetilde{\theta}] = \theta^*$. C'est-à-dire: $M\,Z\,\theta^* = \theta^*$ car $\mathbb{E}\,[M\,Y] = \mathbb{E}\,[M\,(Z\,\theta^* + \varepsilon)] = \mathbb{E}\,[M\,Z\,\theta^*] + \mathbb{E}\,[M\,\varepsilon)] = M\,Z\,\theta^*$ car $\mathbb{E}\,[\varepsilon] = 0$ d'après **(A1)**. 1 est donc valeur propre de la matrice $M\,Z$ et le sous-espace propre est $\mathbb{R}^{p+1}$ tout entier, d'où $M\,Z$ est diagonalisable et sa matrice diagonale est I_{p+1}, donc $M\,Z = P\,I_{p+1}\,P^{-1}$ où P est la matrice de passage, et ainsi $M\,Z = I_{p+1}$.

(ii) $\widehat{\theta} = ({}^tZ\,Z)^{-1}{}^tZ\,Z({}^tZ\,Z)^{-1}{}^tZY$ donc $T = ({}^tZ\,Z)^{-1}{}^tZ$.
$M\,P_{[Z]} = M\,Z({}^tZ\,Z)^{-1}{}^tZ = ({}^tZ\,Z)^{-1}{}^tZ$
$T\,P_{[Z]} = ({}^tZ\,Z)^{-1}{}^tZ\,Z({}^tZ\,Z)^{-1}{}^tZ = ({}^tZ\,Z)^{-1}{}^tZ.$

(iii) $\widetilde{\theta} = M(P_{[Z]}\,Y + P_{[Z]^\perp}\,Y) = \widehat{\theta} + M\,P_{[Z]^\perp}\,Y$, la somme étant non-corrélée car $P_{[Z]}\,Y$ et $P_{[Z]^\perp}\,Y$ le sont. En effet $\mathrm{cov}\,(P_{[Z]}\,Y, P_{[Z]^\perp}\,Y) = \mathrm{cov}\,(P_{[Z]}\,\varepsilon, P_{[Z]^\perp}\,\varepsilon) = \mathbb{E}\,\big(P_{[Z]}\,\varepsilon\,{}^t(P_{[Z]^\perp}\,\varepsilon)\big) = P_{[Z]}\,\mathbb{E}\,\big(\varepsilon\,{}^t\varepsilon\big)\,P_{[Z]^\perp} = \sigma^2\,P_{[Z]}\,P_{[Z]^\perp} = 0$. On en déduit

$$\mathrm{cov}\,(\widetilde{\theta}) = \mathrm{cov}\,(\widehat{\theta}) + \mathrm{cov}\,(M\,P_{[Z]^\perp}\,Y).$$

ce qui donne le résultat.

□