MAP433 PC4 Corrigé A.Tsybakov

Exercice 1.

$\boxed{Q1.}$ Soit A projecteur sur l'espace linéaire engendré par les colonnes de la matrice (e, X). On le décompose :

$$A = P_{\mathcal{L}} + P_{\mathcal{L}^\perp}$$

où $P_{\mathcal{L}}$ est un projecteur sur l'espace linéaire engendré par le seul vecteur e et $P_{\mathcal{L}^\perp}$ est le projecteur sur son complément orthogonal dans $V = \mathrm{Im}[(e,X)]$.

Notons que $P_{\mathcal{L}} = \frac{ee^T}{n}$. D'après les propriétés de l'estimateur des moindres carrés (MC) :

$$\hat{y} = Ay.$$

$$\Rightarrow \hat{y} = P_{\mathcal{L}} y + P_{\mathcal{L}^\perp} y = \frac{e(e^T y)}{n} + P_{\mathcal{L}^\perp} y.$$

$$\bar{\hat{y}} = \frac{1}{n} e^T \hat{y} = \frac{e^T y}{n} + \frac{1}{n} \underbrace{e^T (P_{\mathcal{L}^\perp} y)}_{=0} = \bar{y}.$$

D'autre part,

$$\hat{y} = X\hat{\theta} + \hat{\theta}_0 e \quad \text{et}$$

$$\bar{\hat{y}} = \frac{1}{n} e^T \hat{y} = \frac{1}{n} e^T X \hat{\theta} + \hat{\theta}_0 = \bar{X}\hat{\theta} + \hat{\theta}_0.$$

$\boxed{Q2.}$ Comme $\bar{y}e \in V$, on peut écrire $\bar{y}e = Au$ pour un vecteur $u \in \mathbb{R}^n$. Notons que $\hat{y} - \bar{y}e \perp y - \hat{y}$. En effet, le produit scalaire

$$(\hat{y} - \bar{y}e,\ y - \hat{y}) = (A(y-u),\ y - Ay) =$$

$$= (A(y-u), (I-A)y) = (y-u, (A-A^2)y) = 0,$$

car $A = A^2$ (A est un projecteur). Donc, d'après le Théorème de Pythagore,

$$\|\hat{y}-\bar{y}e\|^2 + \|y-\hat{y}\|^2 = \|y-\bar{y}e\|^2$$

où $\|\cdot\|$ désigne la norme euclidienne.

$$\Rightarrow \quad R^2 \stackrel{\text{déf}}{=} \frac{\|\hat{y}-\bar{y}e\|^2}{\|y-\bar{y}e\|^2} \leq 1$$

$\boxed{Q3.}$ $y = Z a + \xi$, $Z = (e, X)$, $a = \begin{pmatrix} \theta_0 \\ \theta \end{pmatrix}$.

$\hat{a} = \begin{pmatrix} \hat{\theta}_0 \\ \hat{\theta} \end{pmatrix} = (Z^T Z)^{-1} Z^T y$, l'estimateurs des MC de a.

$$\Rightarrow \quad \hat{a} = a + (Z^T Z)^{-1} Z^T \xi.$$

$$E\left[(\hat{a}-a)(\hat{a}-a)^T\right] =$$

$$= E\left[(Z^T Z)^{-1} Z^T \xi \xi^T (Z^T Z)^{-1}\right] = \sigma^2 (Z^T Z)^{-1}$$

$\mathrm{Var}(\hat{\theta}_j) = E(\hat{\theta}_j - \theta_j)^2$, les éléments diagonaux de

$$\sigma^2 (Z^T Z)^{-1} = \begin{pmatrix} \mathrm{Var}(\hat{\theta}_0) & & & * \\ & \mathrm{Var}(\hat{\theta}_1) & & \\ & & \ddots & \\ * & & & \mathrm{Var}(\hat{\theta}_k) \end{pmatrix}.$$

$\boxed{Q4.}$ L'estimateur des MC $\tilde{\theta}$ dans ce modèle n'est pas forcement égal à celui de la Question 1. Exemple : soit $y_i = \theta x_i + \xi_i$, $\theta \in \mathbb{R}$, $x_i \in \mathbb{R}$.

Alors, $\tilde{\theta} = \dfrac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$ est l'EMC dans ce modèle.

D'autre part, pour le modèle $y_i = \theta_0 + \theta x_i + \xi_i$ (avec la constante $\theta_0 \neq 0$), l'EMC s'écrit, d'après la Question 3 :

$$\hat{a} = (Z^T Z)^{-1} Z^T y, \quad \text{où} \quad Z = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$

$$Z^T Z = \begin{pmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix}. \text{ Notons } \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

$$(Z^T Z)^{-1} = \frac{1}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2} \begin{pmatrix} \sum_i x_i^2 & -\sum_i x_i \\ -\sum_i x_i & n \end{pmatrix}$$

$$= \frac{1}{n(\overline{x^2} - \bar{x}^2)} \begin{pmatrix} \overline{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix},$$

$$Z^T y = \begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{pmatrix} = \begin{pmatrix} n\bar{y} \\ \sum_i x_i y_i \end{pmatrix}.$$

$$\hat{a} = \begin{pmatrix} \hat{\theta}_0 \\ \hat{\theta} \end{pmatrix} = \frac{1}{n(\overline{x^2} - \bar{x}^2)} \begin{pmatrix} n\bar{y}\,\overline{x^2} - \bar{x}\sum_i x_i y_i \\ -n\bar{x}\bar{y} + \sum_i x_i y_i \end{pmatrix}.$$

$$\Rightarrow \hat{\theta} = \frac{1}{n(\overline{x^2} - \bar{x}^2)}\left(-n\bar{x}\bar{y} + \sum_i x_i y_i\right) \neq \frac{\sum x_i y_i}{n\,\overline{x^2}}.$$

En effet, généralement,

$$\overline{x^2}\, n^2 \bar{x} \bar{y} \neq n \bar{x}^2 \sum_i x_i y_i$$

$$\Updownarrow$$

$$\overline{x^2}\, n\bar{y} \neq \bar{x} \sum_i x_i y_i$$

(par exemple, ce n'est pas vrai si $y_1 = y_2 = \dots = y_n$ $\Rightarrow$ $\overline{x^2} \neq \bar{x}^2$ généralement). (Si $x_i \equiv x_0$, un cas dégénéré, on a $\overline{x^2} \sum_i y_i = \bar{x} \sum_i x_i y_i$)

$\boxed{Q5.}$ $\bar{\hat{y}} \neq \bar{y}$ généralement pour ce modèle.

Exemple : $K=1$, $y_i = \theta x_i + \xi_i$, $i = 1, \dots, n$

$$\tilde{\theta} = \frac{\sum_i x_i y_i}{\sum_i x_i^2}, \quad \hat{y} = X\tilde{\theta} = \begin{pmatrix} x_1 \tilde{\theta} \\ \vdots \\ x_n \tilde{\theta} \end{pmatrix} = \begin{pmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{pmatrix}$$

$$\bar{\hat{y}} = \frac{1}{n} \sum_{i=1}^n \hat{y}_i = \bar{x} \tilde{\theta} = \frac{\sum_i x_i \sum_i x_i y_i}{n \sum_i x_i^2}.$$

$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. Généralement,

$$\frac{\sum_i x_i \sum_i x_i y_i}{\sum_i x_i^2} \neq \sum_i y_i$$

Dans la question 2, le R^2 était justifié, car là on avait $e \in V$, ce qui a permit d'appliquer Pythagore. Quand il n'y a pas de terme constant θ_0 dans le modèle, cette justification n'existe plus. Le R^2 n'a pas de sens, notamment on peut avoir $R^2 > 1$.

Exercice 2.

Q1. Les mêmes calculs que dans la Question 3 de l'Exercice 1 donnent :

$$E(\hat{\theta}) = \theta, \qquad \operatorname{cov}(\hat{\theta}) = E\left[(\hat{\theta}-\theta)(\hat{\theta}-\theta)^T\right] = \sigma^2 (X^T X)^{-1}.$$

Q2.

$$\tilde{\theta} = LY$$

$$E(\tilde{\theta}) = \theta \iff E(LX\theta + L\xi) = 0$$

$$\iff \underset{\forall \theta \in \mathbb{R}^K}{LX\theta = \theta,} \iff \boxed{LX = I.}$$

Q3.

$$\tilde{\theta} = LX\theta + L\xi = \theta + L\xi$$

$$\Rightarrow E\left[(\tilde{\theta}-\theta)(\tilde{\theta}-\theta)^T\right] = E\left[L\xi\xi^T L^T\right] = \sigma^2 (LL^T)$$

Or,

$$LL^T = (\overbrace{L-L_0}^{\Delta})(\overbrace{L-L_0}^{\Delta})^T + L_0 L^T + L L_0^T - L_0 L_0^T$$

Posons : $L_0 = (X^T X)^{-1} X^T$.

$$\Rightarrow L_0 L^T = (X^T X)^{-1} \underbrace{X^T L^T}_{=I} = (X^T X)^{-1}.$$

$$L L_0^T = \underbrace{LX}_{=I} (X^T X)^{-1} = (X^T X)^{-1}$$

$$L_0 L_0^T = (X^T X)^{-1}.$$

$$\Rightarrow LL^T = \Delta\Delta^T + (X^T X)^{-1} \geq (X^T X)^{-1},$$

car $\Delta\Delta^T \geq 0$, $\forall$ matrice Δ.

Conclusion : $LL^T \geq (X^TX)^{-1}$ pour toute matrice L telle que $LX = I$.

[Th. de Gauss-Markov]

$$\Rightarrow \boxed{Cov(\tilde{\theta}) \geq cov(\hat{\theta}), \quad \forall L : LX = I.}$$

Q4. $\|\theta\|^2 = \text{Trace}(\theta\theta^T)$.

$$\Rightarrow \text{Trace}(Cov(\tilde{\theta})) \geq \text{Trace}(Cov(\hat{\theta}))$$

vu la question 3.

Exercice 3

Q1. L'EMC est solution de $(X^TX)\theta = X^Ty$.
Si $k > n$, la matrice X^TX n'est pas de rang plein, car $\text{Rang}(X) \leq n < k$. Il existe un sous espace linéaire de dimension $\geq k-n$ de solutions θ.

Q2. $\|y - X\theta\|^2 + \lambda \|\theta\|^2 \to \min\limits_{\theta \in \mathbb{R}^k}$, $\lambda > 0$.

La condition nécessaire et suffisante de minimum de cette fonction convexe est l'annulation de gradient :

$$-2X^T(y - X\theta) + 2\lambda\theta = 0$$

$$\Leftrightarrow \quad \theta(X^TX + \lambda I) = X^Ty,$$

d'où

$$\hat{\theta}_\lambda = (X^TX + \lambda I)^{-1} X^Ty.$$

(Notons que $X^TX + \lambda I$ est toujours inversible, car $\lambda > 0$ et $X^TX \geq 0$).

$\boxed{Q3.}$ $E(\hat{\theta}_\lambda) = E[(X^TX+\lambda I)^{-1}X^T(X\theta+\xi)]$

$$= (X^TX+\lambda I)^{-1}X^TX\theta$$

$$\hat{\theta}_\lambda - \theta = (X^TX+\lambda I)^{-1}X^TX\theta - \theta + (X^TX+\lambda I)^{-1}X^T\xi$$

$$\Rightarrow \quad Cov(\hat{\theta}_\lambda) = E[(X^TX+\lambda I)^{-1}X^T\xi\xi^TX^T(X^TX+\lambda I)^{-1}]$$

$$= \sigma^2(X^TX+\lambda I)^{-1}X^TX\ (X^TX+\lambda I)^{-1}$$

Bias: $E(\hat{\theta}_\lambda)-\theta = \quad [(X^TX+\lambda I)^{-1}X^TX - I]\theta$

$$= -\lambda\ (X^TX+\lambda I)^{-1}\theta$$

$\Rightarrow$ $\hat{\theta}_\lambda$ est biaisé.

$\boxed{Q4.}$ $K=1$, Rang $X=1$. $y_i = \theta x_i + \xi_i$

$$\hat{\theta}_\lambda = \frac{1}{\sum_{i=1}^n x_i^2+\lambda}\sum_{i=1}^n x_i y_i, \qquad \hat{\theta}^{MC} = \frac{\sum_i x_i y_i}{\sum_{i=1}^n x_i^2}.$$

$$\Rightarrow \hat{\theta}_\lambda = c\,\hat{\theta}^{MC}, \quad c = \frac{\|x\|^2}{\|x\|^2+\lambda}, \quad 0<c<1.$$

$$E(\hat{\theta}_\lambda-\theta)^2 = E(c\hat{\theta}^{MC}-\theta) = c^2E(\hat{\theta}^{MC}-\theta)^2 + (1-c)^2\theta^2.$$

$$\min_{0<c<1}[c^2\, Var(\hat{\theta}^{MC}) + (1-c)^2\theta^2] < Var(\hat{\theta}^{MC}).$$

$$\min_{0<c<1}[c^2a + (1-c^2)b] \Rightarrow c_{opt} = \frac{b}{a+b}$$

$$c_{opt}^2\, a + (1-c_{opt})^2 b = \frac{ab}{a+b} < \min(a,b).$$