

HEC Paris – ESCP Europe

Mathématiques approfondies – Mathématiques II

Corrigé du sujet adapté – 2012

Les questions et sous-questions conservent la numérotation de l'énoncé adapté. On note $\mathbf{1} = (1, \dots, 1)$, $x = (x_1, \dots, x_n)$, $\alpha = (\alpha_1, \dots, \alpha_n)$, et $\langle \cdot, \cdot \rangle$ le produit scalaire canonique de $\mathbb{R}^n$ (identifié à $M_{n,1}(\mathbb{R})$).

Partie I – Quelques résultats statistiques et algébriques

Partie I – Moindres carrés

$y = M\theta + u$, $F = \text{Vect}(x, \mathbf{1})$, K projecteur orthogonal sur F , $G = I - K$.

Corrigé – Partie I.

1. Moyennes et variances empiriques.

- a) s_X^2 est une somme de carrés, donc $s_X^2 \geq 0$. Si $s_X^2 = 0$, tous les x_i sont égaux à $\bar{X}$, ce qui est exclu.

$$s_X^2 > 0.$$

- b) On développe, en utilisant $\sum_i y_i = n\bar{Y}$ et $\sum_i x_i = n\bar{X}$:

$$\sum_i (x_i - \bar{X})y_i = \sum_i x_i y_i - \bar{X} \cdot n\bar{Y}, \quad ns_X^2 = \sum_i x_i^2 - 2\bar{X} \cdot n\bar{X} + n\bar{X}^2.$$

$$\sum_{i=1}^n (x_i - \bar{X})y_i = \sum_{i=1}^n x_i y_i - n\bar{X}\bar{Y}, \quad ns_X^2 = \sum_{i=1}^n x_i^2 - n\bar{X}^2.$$

- c) On a $\sum_i (x_i - \bar{X}) = 0$, donc $\sum \alpha_i = 0$. Par b) appliquée avec $y = x$, $\sum_i (x_i - \bar{X})x_i = ns_X^2$, donc $\sum \alpha_i x_i = 1$. Enfin, $\sum_i (x_i - \bar{X})^2 = ns_X^2$.

$$\sum_{i=1}^n \alpha_i = 0, \quad \sum_{i=1}^n \alpha_i x_i = 1, \quad \sum_{i=1}^n \alpha_i^2 = \frac{ns_X^2}{(ns_X^2)^2} = \frac{1}{ns_X^2}.$$

2. La matrice M .

- a) Les colonnes de M sont x et $\mathbf{1}$. Elles ne sont pas colinéaires, car les x_i ne sont pas tous égaux et $\mathbf{1} \neq 0$.

$$\text{rg}(M) = 2.$$

- b) On calcule

$${}^t M M = \begin{pmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{pmatrix} = \begin{pmatrix} \sum x_i^2 & n\bar{X} \\ n\bar{X} & n \end{pmatrix}.$$

Si ${}^tMM\theta = 0$, alors $\theta {}^tMM\theta = \|M\theta\|^2 = 0$, donc $M\theta = 0$, puis $\theta = 0$ car les colonnes de M sont libres. La matrice carrée ${}^tMM$ est donc injective.

$${}^tMM = \begin{pmatrix} \sum_i x_i^2 & n\bar{X} \\ n\bar{X} & n \end{pmatrix} \text{ est inversible.}$$

3. Projection orthogonale.

- a) On a $\sum \hat{u}_i^2 = \|y - M\hat{\theta}\|^2$, et $\{M\hat{\theta}\} = \text{Im } M = F$. Par le théorème de la projection orthogonale, $\|y - z\|$, pour $z \in F$, est minimal en un unique point $z = p_F(y)$ (projeté orthogonal de y sur F). Comme M est injective (rang 2), il existe un unique $\hat{\theta}$ tel que $M\hat{\theta} = p_F(y)$. Ce point est caractérisé par $y - M\hat{\theta} \in F^\perp$, c'est-à-dire par l'orthogonalité de $y - M\hat{\theta}$ aux colonnes de M : ${}^tM(y - M\hat{\theta}) = 0$.

$$\text{Unique solution } \hat{\theta}, \text{ caractérisée par } {}^tMM\hat{\theta} = {}^tMy.$$

- b) On a ${}^tMy = \begin{pmatrix} \sum x_i y_i \\ n\bar{Y} \end{pmatrix}$. La seconde équation du système donne $n\bar{X}\hat{a} + n\hat{b} = n\bar{Y}$.

On reporte dans la première, $\sum x_i^2 \hat{a} + n\bar{X}(\bar{Y} - \hat{a}\bar{X}) = \sum x_i y_i$, soit, par 1b), $ns_X^2 \hat{a} = \sum (x_i - \bar{X})y_i$.

$$\hat{a} = \sum_{i=1}^n \alpha_i y_i, \quad \hat{b} = \bar{Y} - \hat{a}\bar{X}.$$

- c) Par a), $Ky = p_F(y) = M\hat{\theta} = M({}^tMM)^{-1}{}^tMy$ pour tout y .

$$K = M({}^tMM)^{-1}{}^tM.$$

- d) $\hat{u} = y - Ky = Gy$. Par ailleurs, $M\theta \in F$, donc $GM\theta = M\theta - KM\theta = 0$, et $Gy = G(M\theta + u) = Gu$.

$$\hat{u} = Gy = Gu.$$

- e) $\hat{u} = Gy \in F^\perp$ et $\mathbf{1} \in F$, donc $\sum \hat{u}_i = \langle \hat{u}, \mathbf{1} \rangle = 0$. G est la matrice, dans une base orthonormale, du projecteur orthogonal sur $F^\perp$: elle est symétrique et vérifie $G^2 = G$. D'où $\sum \hat{u}_i^2 = {}^t(Gy)Gy = {}^t_y G^2 y = {}^t_y Gy$, et de même avec u .

$$\sum_{i=1}^n \hat{u}_i = 0, \quad \sum_{i=1}^n \hat{u}_i^2 = {}^t_y Gy = {}^t_u Gu.$$

Partie II – Le modèle de régression linéaire

Partie II – Estimateurs des moindres carrés

$$Y_i = ax_i + b + U_i, \quad A_n = \sum \alpha_i Y_i, \quad B_n = \bar{Y}_n - A_n \bar{X}.$$

Corrigé – Partie II.

4. Biais et variances.

On utilise $\mathbb{E}(Y_i) = ax_i + b$ et le fait que $Y_1, \dots, Y_n$ sont indépendantes de variance σ^2 .

- a) Par linéarité et 1c), $\mathbb{E}(A_n) = a \sum \alpha_i x_i + b \sum \alpha_i = a$. Ensuite, $\mathbb{E}(\bar{Y}_n) = a\bar{X} + b$, donc $\mathbb{E}(B_n) = a\bar{X} + b - a\bar{X} = b$.

$$\boxed{\mathbb{E}(A_n) = a, \quad \mathbb{E}(B_n) = b.}$$

- b) On écrit $B_n = \sum_i \beta_i Y_i$ avec $\beta_i = \frac{1}{n} - \alpha_i \bar{X}$. Par indépendance, $\mathbb{V}(A_n) = \sigma^2 \sum \alpha_i^2$ et $\mathbb{V}(B_n) = \sigma^2 \sum \beta_i^2$. Avec 1c), $\sum \beta_i^2 = \frac{1}{n} - \frac{2\bar{X}}{n} \sum \alpha_i + \bar{X}^2 \sum \alpha_i^2 = \frac{1}{n} + \frac{\bar{X}^2}{ns_X^2}$.

$$\boxed{\mathbb{V}(A_n) = \frac{\sigma^2}{ns_X^2}, \quad \mathbb{V}(B_n) = \left(1 + \frac{\bar{X}^2}{s_X^2}\right) \frac{\sigma^2}{n}.}$$

- c) Par bilinéarité et indépendance, $\text{Cov}(A_n, B_n) = \sigma^2 \sum \alpha_i \beta_i = \sigma^2 \left(\frac{1}{n} \sum \alpha_i - \bar{X} \sum \alpha_i^2\right)$.

$$\boxed{\text{Cov}(A_n, B_n) = -\frac{\bar{X}\sigma^2}{ns_X^2}.}$$

5. Convergence en probabilité.

Ici $\bar{X} = \bar{X}_n$ et $s_X^2 = s_n^2$ dépendent de n , avec $\bar{X}_n \rightarrow \lambda$ et $s_n^2 \rightarrow \mu^2 > 0$. Soit $\varepsilon > 0$. Par l'inégalité de Bienaymé-Tchebychev (les estimateurs sont sans biais) :

$$\mathbb{P}(|A_n - a| \geq \varepsilon) \leq \frac{\sigma^2}{ns_n^2 \varepsilon^2} \xrightarrow{n \rightarrow +\infty} 0, \quad \mathbb{P}(|B_n - b| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \left(1 + \frac{\bar{X}_n^2}{s_n^2}\right) \xrightarrow{n \rightarrow +\infty} 0,$$

car $ns_n^2 \rightarrow +\infty$ et $1 + \frac{\bar{X}_n^2}{s_n^2} \rightarrow 1 + \frac{\lambda^2}{\mu^2}$.

$$\boxed{A_n \xrightarrow{\mathbb{P}} a \quad \text{et} \quad B_n \xrightarrow{\mathbb{P}} b.}$$

6. Estimation de σ^2 .

- a) Par linéarité et 4a), $\mathbb{E}(\hat{U}_i) = ax_i + b - x_i a - b$.

$$\boxed{\mathbb{E}(\hat{U}_i) = 0.}$$

- b) On a $Y_i - \bar{Y}_n = a(x_i - \bar{X}) + U_i - \bar{U}_n$, donc

$$\hat{U}_i = (Y_i - \bar{Y}_n) - A_n(x_i - \bar{X}) = (U_i - \bar{U}_n) - (A_n - a)(x_i - \bar{X}).$$

Par ailleurs, $A_n - a = \sum \alpha_i U_i = \frac{1}{ns_X^2} \sum (x_i - \bar{X})(U_i - \bar{U}_n)$, car $\sum (x_i - \bar{X}) = 0$. On développe le carré :

$$\sum \hat{U}_i^2 = \sum (U_i - \bar{U}_n)^2 - 2(A_n - a) \cdot ns_X^2 (A_n - a) + (A_n - a)^2 ns_X^2.$$

$$\boxed{\sum_{i=1}^n \hat{U}_i^2 = \sum_{i=1}^n (U_i - \bar{U}_n)^2 - ns_X^2 (A_n - a)^2.}$$

- c) $\sum (U_i - \bar{U}_n)^2 = \sum U_i^2 - n\bar{U}_n^2$, d'espérance $n\sigma^2 - n \cdot \frac{\sigma^2}{n} = (n-1)\sigma^2$. Et $\mathbb{E}((A_n - a)^2) = \mathbb{V}(A_n) = \frac{\sigma^2}{ns_X^2}$. Par b) :

$$\mathbb{E} \left(\sum_{i=1}^n \hat{U}_i^2 \right) = (n-2)\sigma^2.$$

Donc :

$$\frac{1}{n-2} \sum_{i=1}^n \hat{U}_i^2 \text{ est un estimateur sans biais de } \sigma^2.$$

Partie III – Hypothèse de normalité et prévision

Partie III – Vecteurs normaux et intervalle de prévision

$U_i \hookrightarrow \mathcal{N}(0, \sigma^2)$ indépendantes.

Corrigé – Partie III.

On utilise : une combinaison linéaire $\sum c_i U_i$ de variables indépendantes de loi $\mathcal{N}(0, \sigma^2)$ suit la loi $\mathcal{N}(0, \sigma^2 \sum c_i^2)$ (stabilité des lois normales).

7. Premiers exemples.

- a) Pour $p \neq 0$, $\sum p_i Y_i = \sum p_i (ax_i + b) + \sum p_i U_i$ suit une loi normale de variance $\sigma^2 \sum p_i^2 > 0$. En revanche, $\sum_{i=1}^n (Y_i - \bar{Y}_n) = 0$: pour $p = \mathbf{1}$, la combinaison est la variable certaine nulle, de variance nulle.

$$(Y_1, \dots, Y_n) \text{ est normal, } (Y_1 - \bar{Y}_n, \dots, Y_n - \bar{Y}_n) \text{ ne l'est pas.}$$

- b) $A_n = \sum \alpha_i Y_i$ et $B_n = \sum \beta_i Y_i$ suivent des lois normales, d'espérances et variances données en 4. Pour $(p, q) \neq (0, 0)$, $pA_n + qB_n = \sum c_i Y_i$ avec $c = (p - q\bar{X})\alpha + \frac{q}{n}\mathbf{1}$. Les vecteurs α et $\mathbf{1}$ sont libres : $\alpha \neq 0$, et si $\alpha = t\mathbf{1}$, alors $\sum \alpha_i = nt = 0$, donc $\alpha = 0$. Par suite, $c = 0$ impose $q = 0$ puis $p = 0$. Donc $c \neq 0$, et $pA_n + qB_n$ suit une loi normale de variance $\sigma^2 \sum c_i^2 > 0$.

$$A_n \hookrightarrow \mathcal{N} \left(a, \frac{\sigma^2}{ns_X^2} \right), \quad B_n \hookrightarrow \mathcal{N} \left(b, \left(1 + \frac{\bar{X}^2}{s_X^2} \right) \frac{\sigma^2}{n} \right), \quad (A_n, B_n) \text{ est normal.}$$

8. Image d'un vecteur gaussien.

- a) Pour $p \neq 0$, $\sum p_i T_i = {}^t p S U = \sum_j c_j U_j$ avec $c = {}^t S p \neq 0$, car ${}^t S$ est inversible. Cette variable suit une loi normale de variance $\sigma^2 \|c\|^2 > 0$.

$$(T_1, \dots, T_n) \text{ est normal.}$$

- b) Par bilinéarité de la covariance et indépendance des U_k , $\text{Cov}(T_i, T_j) = \sigma^2 \sum_k s_{ik} s_{jk} = \sigma^2 (S^t S)_{ij}$. Si S est orthogonale, $S^t S = I$, donc $\text{Cov}(T_i, T_j) = 0$ pour $i \neq j$. Le résultat admis (le vecteur étant normal par a)) donne :

$$T_1, \dots, T_n \text{ sont mutuellement indépendantes (et de loi } \mathcal{N}(0, \sigma^2)).$$

9. Loi de la somme des carrés des résidus.

- a) Par 3a) et 3b), pour tout y , $\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = ({}^tMM)^{-1}{}^tMy$ avec $\hat{a} = \sum \alpha_i y_i$, $\hat{b} = \bar{Y} - \hat{a}\bar{X}$. En remplaçant y par $Y(\omega)$, on obtient $\begin{pmatrix} A_n \\ B_n \end{pmatrix} = ({}^tMM)^{-1}{}^tMY$, puis $M \begin{pmatrix} A_n \\ B_n \end{pmatrix} = KY$ par 3c). Donc $\hat{U} = GY = G(M\theta + U) = GU$, comme en 3d.

$$\boxed{\hat{U} = GU.}$$

- b) G est symétrique réelle : par le théorème spectral, il existe R orthogonale et D diagonale telles que $G = RD^tR$. G représente le projecteur orthogonal sur $F^\perp$, de dimension $n - 2$. Ses valeurs propres sont donc 1 (sous-espace propre $F^\perp$, de dimension $n - 2$) et 0 (sous-espace propre F , de dimension 2). On peut ordonner les vecteurs propres pour avoir $D = \text{diag}(1, \dots, 1, 0, 0)$.

$$\boxed{D \text{ a pour éléments diagonaux } 1 \text{ (} n - 2 \text{ fois) et } 0 \text{ (2 fois).}}$$

- c) Par 8, avec $S = {}^tR$ orthogonale, $Z_1, \dots, Z_n$ sont indépendantes, de loi $\mathcal{N}(0, \sigma^2)$. Alors $\frac{Z_i}{\sigma} \hookrightarrow \mathcal{N}(0, 1)$ et $\frac{Z_i^2}{2\sigma^2} \hookrightarrow \gamma\left(\frac{1}{2}\right)$. Par stabilité des lois γ (variables indépendantes par le lemme des coalitions), $\frac{1}{2\sigma^2} \sum_{i=1}^{n-2} Z_i^2 \hookrightarrow \gamma\left(\frac{n-2}{2}\right)$.

$$\boxed{\sum_{i=1}^{n-2} Z_i^2 \hookrightarrow \Gamma\left(2\sigma^2, \frac{n-2}{2}\right).}$$

- d) Comme G est symétrique et $G^2 = G$, on a $\sum \hat{U}_i^2 = {}^tUG^2U = {}^tUGU = {}^tURD^tRU = {}^tZDZ = \sum_{i=1}^{n-2} Z_i^2$.

$$\boxed{\sum_{i=1}^n \hat{U}_i^2 \hookrightarrow \Gamma\left(2\sigma^2, \frac{n-2}{2}\right).}$$

- e) La variable $Q = \frac{1}{\sigma^2} \sum \hat{U}_i^2$ suit la loi $\Gamma\left(2, \frac{n-2}{2}\right)$, qui ne dépend pas de a, b, σ^2 . Sa fonction de répartition F_Q est continue et strictement croissante sur $]0, +\infty[$ (densité strictement positive), et réalise une bijection de $]0, +\infty[$ sur $]0, 1[$. On pose $c_n = F_Q^{-1}(1 - p) > 0$. Alors $\mathbb{P}(Q \geq c_n) = 1 - F_Q(c_n) = p$.

$$\boxed{\exists c_n > 0 \text{ (indépendant de } a, b, \sigma^2), \quad \mathbb{P}\left(\sum_{i=1}^n \hat{U}_i^2 \geq c_n \sigma^2\right) = p.}$$

10. Meilleur prédicteur linéaire sans biais.

- a) On note $C = \{r \in \mathbb{R}^n : \sum r_i = 1, \sum x_i r_i = x_{n+1}\}$. Par 1c), $\sum r_i^* = 1$ et $\sum x_i r_i^* = \bar{X} + (x_{n+1} - \bar{X}) = x_{n+1}$, donc $r^* \in C$. Soit $r \in C$ et $h = r - r^*$. Alors $\sum h_i = 0$ et $\sum x_i h_i = 0$, donc $\sum \alpha_i h_i = \frac{1}{ns_X^2} \sum (x_i - \bar{X}) h_i = 0$, et

$$\langle r^*, h \rangle = \frac{1}{n} \sum h_i + (x_{n+1} - \bar{X}) \sum \alpha_i h_i = 0.$$

Par le théorème de Pythagore, $g(r) = \|r^* + h\|^2 = g(r^*) + \|h\|^2 \geq g(r^*)$, avec égalité si et seulement si $h = 0$.

$$\boxed{g \text{ admet sous les contraintes un minimum absolu atteint uniquement en } r^* .}$$

- b) $\mathbb{E}(\hat{Y}_{n+1}^{(r)}) = a \sum r_i x_i + b \sum r_i$ et $\mathbb{E}(Y_{n+1}) = ax_{n+1} + b$. L'égalité pour tout (a, b) équivaut (prendre $(a, b) = (1, 0)$ puis $(0, 1)$) à $r \in C$. Par indépendance, $\mathbb{V}(\hat{Y}_{n+1}^{(r)}) = \sigma^2 g(r)$, minimale sur C en r^* uniquement, par a). Enfin,

$$\hat{Y}_{n+1}^{(r^*)} = \bar{Y}_n + (x_{n+1} - \bar{X})A_n = A_n x_{n+1} + B_n.$$

$$\boxed{\hat{Y}_{n+1}^{(r^*)} = A_n x_{n+1} + B_n \text{ est le prédicteur linéaire sans biais de variance minimale.}}$$

11. Intervalle de prévision.

- a) Comme $r^* \in C$, $\sum r_i^* Y_i = ax_{n+1} + b + \sum r_i^* U_i$, donc $W = Y_{n+1} - (A_n x_{n+1} + B_n) = U_{n+1} - \sum_{i=1}^n r_i^* U_i$. C'est une combinaison linéaire non nulle des variables indépendantes $U_1, \dots, U_{n+1}$. Avec $\sum \alpha_i = 0$, on obtient $\sum (r_i^*)^2 = \frac{1}{n} + \frac{(x_{n+1} - \bar{X})^2}{ns_X^2}$.

$$\boxed{W \hookrightarrow \mathcal{N}(0, \sigma^2 v_n), \quad v_n = 1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{X})^2}{ns_X^2}.$$

- b) $\frac{W}{\sigma \sqrt{v_n}} \hookrightarrow \mathcal{N}(0, 1)$, donc, pour $d > 0$, $\mathbb{P}(|W| \leq d\sigma) = 2\Phi\left(\frac{d}{\sqrt{v_n}}\right) - 1$. Cette quantité vaut p si et seulement si $\Phi\left(\frac{d}{\sqrt{v_n}}\right) = \frac{1+p}{2} \in \left]\frac{3}{4}, 1\right[$, ce qui a une unique solution (Φ est une bijection de $\mathbb{R}$ sur $]0, 1[$).

$$\boxed{d_n = \sqrt{v_n} \Phi^{-1}\left(\frac{1+p}{2}\right) = \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{X})^2}{ns_X^2}} \Phi^{-1}\left(\frac{1+p}{2}\right).$$

- c) On note $E_1 = \left[\sum \hat{U}_i^2 \geq c_n \sigma^2\right] = \left[\sigma \leq \sqrt{\frac{1}{c_n} \sum \hat{U}_i^2}\right]$ et $E_2 = [|W| \leq d_n \sigma]$. On a $\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2) - \mathbb{P}(E_1 \cup E_2) \geq 2p - 1$. Sur $E_1 \cap E_2$, $|W| \leq d_n \sqrt{\frac{1}{c_n} \sum \hat{U}_i^2}$. On pose $\hat{\Sigma}_n = \sqrt{\frac{1}{c_n} \sum_{i=1}^n \hat{U}_i^2}$, qui ne dépend que des Y_i , des x_i et de c_n .

$$\boxed{\mathbb{P}\left(Y_{n+1} \in [A_n x_{n+1} + B_n - d_n \hat{\Sigma}_n, A_n x_{n+1} + B_n + d_n \hat{\Sigma}_n]\right) \geq 2p - 1.$$

Ce n'est pas un intervalle de confiance au sens usuel : il encadre la valeur d'une variable aléatoire Y_{n+1} , et non un paramètre inconnu (a , b ou σ^2) du modèle. C'est un intervalle de prévision.