REVUE DE STATISTIQUE APPLIQUÉE

J. L. MALLET

Régression sous contraintes linéaires : application au codage des variables aléatoires

Revue de statistique appliquée, tome 28, nº 1 (1980), p. 57-68

<http://www.numdam.org/item?id=RSA_1980__28_1_57_0>

REGRESSION SOUS CONTRAINTES LINEAIRES : APPLICATION AU CODAGE DES VARIABLES ALEATOIRES

J.L. MALLET
Ecole Nationale Supérieure de Géologie (Nancy)

1. NOTION DE REGRESSION LINEAIRE SOUS CONTRAINTES LINEAIRES

1.1. Introduction

Notations

Dans ce qui suit, nous adopterons les notations suivantes :

1) CEVA = classe d'équivalence de variable aléatoire

2) $L^2_{R^n}(\Omega, A, P) = L^2(\Omega, A, P) \underbrace{\times \ldots \times}_{\text{n fois}} L^2(\Omega, A, P)$

3) $\dot{X} = \begin{bmatrix} \dot{X}_1 \\ \vdots \\ \dot{X}_n \end{bmatrix}$ = CEVA appartenant à $L^2_{R^n}(\Omega, A, P)$

4) $H_x = \{\dot{z} \in L^2(\Omega, A, P) \mid z \in R^n : \dot{Z} = z^t \dot{X}\}$
= sous-espace de $L^2(\Omega, A, P)$ engendré par $\dot{X}$

5) $\Pi_x(\dot{Y})$ = projection orthogonale de $\dot{Y} \in L^2(\Omega, A, P)$ sur le sous-espace H_x

6) Pour toutes CEVA $\dot{X} \in L^2_{R^n}(\Omega, A, P)$ et $\dot{Y} \in L^2(\Omega, A, P)$, nous poserons :

$$E_{xx} = E(XX^t) \quad \forall X \in \dot{X}$$
$$E_{xy} = E(XY^t) \quad \forall X \in \dot{X}, Y \in \dot{Y}$$

Mots-Clés : **Régression. Contraintes. Codage.**

Rappels

En utilisant le théorème de la projection orthogonale, on montre facilement que pour toute inverse généralisée E_{xx}^- de E_{xx}, on a :

1) $\Pi_x(\dot{Y}) = \{E_{yx} . E_{xx}^-\} . \dot{X}$

De plus, pour toute inverse généralisée E_{xx}^-, on a :

2) $\| Y - \Pi_x(Y) \|^2 = \| Y \|^2 - E_{yx} . E_{xx}^- . E_{xy}$

Mots-clés : **Régression. Contraintes. Codage linéaire**

Remarque

A toute inverse généralisée E_{xx}^-, on peut associer une inverse généralisée $\tilde{E}_{xx}^-$ vérifiant $\tilde{E}_{xx}^- . E_{xx} . \tilde{E}_{xx}^- = \tilde{E}_{xx}^-$; en effet, il suffit de poser :

$$\tilde{E}_{xx}^- = E_{xx}^- . E_{xx} . E_{xx}^-$$

1.2. Définition de l'opérateur $\Pi_x^c(.)$

Enoncé

Soit $\dot{X}$ une CEVA appartenant à $L^2_{R^n}(\Omega, A, P)$ et soit $\mathcal{C}_x$ une partie convexe de H_x définie de la façon suivante, où D_1 et D_2 désignent des matrices colonnes de tailles respectivement $p \geqslant 0$ et $q \geqslant 0$ tandis que C_1 et C_2 désignent des matrices rectangles à n colonnes et respectivement p et q lignes :

$$\dot{Z} \in \mathcal{C}_x \Longleftrightarrow \dot{Z} = z^t . \dot{X} \quad \text{avec :} \begin{cases} C_1 . z = D_1 \\ C_2 . z \leqslant D_2 \end{cases}$$

Pour toute CEVA scalaire $\dot{Y} \in L^2(\Omega, A, P)$, il existe un vecteur unique noté $\Pi_x^c(\dot{Y})$ tel que :

1) $$\Pi_x^c(Y) \in \mathcal{C}_x$$

2) $$\|\dot{Y} - \Pi_x^c(\dot{Y})\| \leqslant \|\dot{Y} - \dot{Z}\| \quad \forall \dot{Z} \in \mathcal{C}_x$$

Par définition, on dit que $\Pi_x^c(\dot{Y})$ est la projection orthogonale de $\dot{Y}$ sur $\mathcal{C}_x$ et que $\Pi_x^c(.)$ est l'opérateur de projection orthogonale de $L^2(\Omega, A, P)$ sur le convexe $\mathcal{C}_x$.

Démonstration

Supposons que $\dot{Z}_1$ et $\dot{Z}_2$ soient deux vecteurs appartenant à $\mathcal{C}_x$ et soit $\dot{Z}$ le vecteur tel que :

$$\begin{cases} Z = \alpha . Z_1 + (1 - \alpha) . Z_2 \\ \text{avec : } \alpha \in [0, 1] \end{cases} \tag{1}$$

Puisque Z_1 et Z_2 appartiennent à $\mathcal{C}_x$, il s'ensuit qu'il existe z_1 et z_2 tels que :

$$\dot{Z}_1 = z_1^t . \dot{X} \quad \text{avec :} \begin{cases} C_1 . z_1 = D_1 \\ C_2 . z_1 \leqslant D_2 \end{cases}$$

$$\dot{Z}_2 = z_2^t . \dot{X} \quad \text{avec :} \begin{cases} C_1 . z_2 = D_1 \\ C_2 . z_2 \leqslant D_2 \end{cases}$$

On vérifie facilement qu'il s'ensuit que :

$$\dot{Z} = z^t . \dot{X}$$

avec : (2)

$$z = \alpha . z_1 + (1 - \alpha) . z_2$$

On peut alors écrire :

$$\begin{aligned} C_1 . z &= \alpha . C_1 . z_1 + (1-\alpha) . C_1 z_2 \\ &= \alpha . D_1 \quad + (1-\alpha) . D_1 \\ &= D_1 \end{aligned}$$

$$\begin{aligned} C_2 . z &= \alpha . C_2 . z_1 + (1-\alpha) . C_2 z_2 \\ &\leqslant \alpha . D_2 \quad + (1-\alpha) . D_2 \qquad \text{car } \alpha \in [0 , 1] \\ &\leqslant D_2 \end{aligned}$$

D'après (2) et par définition de $\mathcal{C}_x$, on en déduit que $Z \in \mathcal{C}_x$ pour tout $\alpha \in [0 , 1]$. donc d'après (1), on a :

$$\left.\begin{array}{l} Z_1 \in C_x \\ Z_2 \in C_x \end{array}\right\} \Rightarrow Z = \alpha . z_1 + (1-\alpha) . Z_2 \in \mathcal{C}_x \quad \forall \alpha \in [0 , 1]$$

En remarquant que cette implication n'est autre que celle définissant un ensemble convexe, on en conclut que $\mathcal{C}_x$ est convexe.

Ceci étant dit, on vérifie facilement que $\mathcal{C}_x$ est fermé dans $L^2 (\Omega , A , P)$, donc d'après le théorème de la projection orthogonale sur un convexe fermé, on sait que pour tout vecteur $\dot{Y} \in L^2 (\Omega , A , P)$, il existe un vecteur unique $\Pi_x^c (\dot{Y})$ appelé "projection orthogonale de $\dot{Y}$ sur $\mathcal{C}_x$ tel que :

1) $$\Pi_x^c (\dot{Y}) \in \mathcal{C}_x$$

2) $$||\dot{Y} - K_x^c (\dot{Y})|| \leqslant ||\dot{Y} - \dot{Z}|| \quad \forall \dot{Z} \in \mathcal{C}_x$$

Interprétation graphique

On trouvera sur la figure 1 une interprétation graphique de la définition précédente. Compte tenu de la façon dont le domaine convexe $\mathcal{C}_x$ est défini, on vérifie facilement que $\mathcal{C}_x$ est en fait un "polygone" convexe tout entier contenu dans H_x .

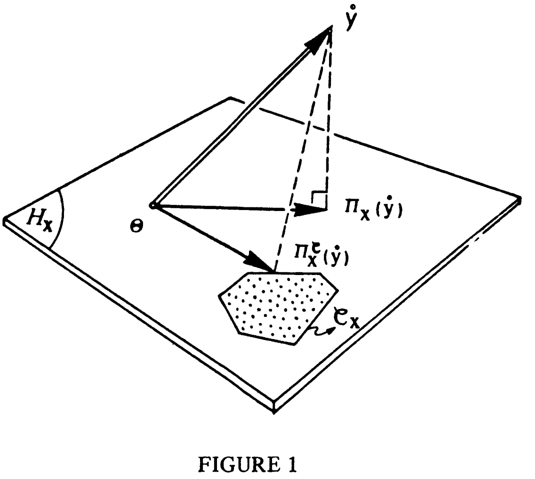

FIGURE 1

2. MISE EN OEUVRE DE LA REGRESSION LINEAIRE SOUS CONTRAINTES LINEAIRES

2.1. Proposition fondamentale

Enoncé

Soient C et D les matrices à p + q lignes telles que :

$$C = \begin{bmatrix} [C_1] \\ [C_2] \end{bmatrix} \qquad D = \begin{bmatrix} [D_1] \\ [D_2] \end{bmatrix}$$

Soit d'autre part E_{xx}^- une inverse généralisée symétrique de E_{xx} et soit $\gamma(u)$ la fonction définie sur R^{p+q} par :

$$\left[\begin{array}{l} \gamma(u) = 2 \,.\, B^t \,.\, u - u^t \,.\, A \,.\, u \\ \text{avec :} \begin{cases} A = C \,.\, \tilde{E}_{xx}^- \,.\, C^t \\ B = C \,.\, E_{xx}^- \,.\, E_{xy} - D \\ \tilde{E}_{xx}^- = E_{xx}^- \,.\, E_{xx} \,.\, E_{xx}^- \end{cases} \end{array}\right.$$

Dans ces conditions, si $\mathcal{C}_x$ n'est pas vide et si l'on désigne par $\mathcal{C}_x^*$ l'ensemble des vecteurs $u \in R^{p+q}$ de composantes u_i telles que :

$$\mathcal{C}_x^* = \{u \mid u_{p+1} \geqslant 0 , \ldots , u_{p+q} \geqslant 0\}$$

... alors l'ensemble des $\hat{u} \in \mathcal{C}_x^*$ tels que :

$$\gamma(\hat{u}) = \max_{u \in \mathcal{C}_x^*} \gamma(u)$$

... n'est pas vide et pour tout $\hat{u}$ vérifiant cette relation, on a :

1) $$\Pi_x^c(\dot{Y}) = (E_{yx} - \hat{u}^t \,.\, C) \,.\, E_{xx}^- \,.\, \dot{X}$$

2) $$\| \dot{Y} - \Pi_x^c(\dot{Y}) \|^2 - \| \dot{Y} \|^2 - E_{yx} \,.\, \tilde{E}_{xx}^- \,.\, E_{xy} + \gamma(\hat{u})$$

3) $$\| \Pi_x(\dot{Y}) - \Pi_x^c(\dot{Y}) \|^2 = \gamma(\hat{u})$$

Démonstration

Pour tout vecteur $z \in R^n$, on peut écrire :

$$\| \dot{Y} - z^t \dot{X} \|^2 = z^t \,.\, E_{xx} \,.\, z - 2 \,.\, E_{yx} \,.\, z + \| \dot{Y} \|^2 \tag{1}$$

Par ailleurs, si l'on désigne par $\mathcal{C}$ la partie non vide de R^n et par $\phi(z , u)$ la fonction définie sur R^{n+p+q} telles que :

$$z \in \mathcal{C} \Longleftrightarrow \begin{cases} C_1 \,.\, z = D_1 \\ C_2 \,.\, z \leqslant D_2 \end{cases} \tag{2}$$

$$\phi(z , u) = \| Y - z^t \,.\, X \|^2 + 2 \,.\, u^t \,.\, (C \,.\, z - D) \quad \forall \begin{cases} z \in R^n \\ u \in R^{p+q} \end{cases} \tag{3}$$

... alors, d'après le théorème de Lagrange-Kuhn et Tucker (*cf.* LUENGERGER [3] p. 225), on sait que, sous réserve que $\mathcal{C}$ ne voit pas vide, il existe $\hat{z} \in \mathcal{C}$ et $\hat{u} \in \mathcal{C}_x^*$ tels que :

$$\phi(\hat{z}, \hat{u}) = \max_{u \in \mathcal{C}_x^*} \min_{z \in R^n} \phi(z, u) = \min_{z \in \mathcal{C}} \|\dot{Y} - z^t \dot{X}\|^2 = \|\dot{Y} - \hat{z}^t \dot{X}\|^2 \quad (4)$$

Compte tenu de (3), pour tout u fixé, ϕ admet un minimum en z. La minimisation de ϕ par rapport à z n'étant soumise à aucune contrainte, il s'ensuit qu'elle est obtenue pour $\partial\phi/\partial z = \theta_{R^n}$ où θ_{R^n} désigne le vecteur nul de R^n ; compte tenu que d'après (1) et (3), on a :

$$\frac{\partial \phi}{\partial z} = \frac{\partial}{\partial z} \|Y - z^t x\|^2 + 2 \,.\, u^t(Cz - D) = 2 \,.\, E_{xx} \,.\, z - 2 \,.\, E_{xy} + 2 \,.\, C^t \,.\, u$$

... on en déduit que pour chaque vecteur $u \in R^{p+q}$ fixé, le minimum de $\phi(z, u)$ en fonction de z est atteint pour tout point $z = z^*(u)$ tel que :

$$\theta_{R^n} = 2 \,.\, E_{xx} \,.\, z^*(u) - 2 \,.\, E_{xy} + 2 \,.\, C^t \,.\, u$$

En remarquant que pour u fixé, $\phi(z, u)$ admet toujours un minimum, on est assuré que l'ensemble des $z^*(u)$ ainsi définis n'est pas vide et à toute inverse généralisée symétrique E_{xx}^- correspond l'un de ces vecteurs $z^*(u)$ tel que :

$$z^*(u) = \widetilde{E}_{xx}^- \,.\, (E_{xy} - C^t u) \Rightarrow \phi(z^*(u), u) = \min_{z \in R^n} \phi(z, u) \quad (5)$$

Posons $$\phi(u) = \phi(z^*(u), u) = \min_{z \in R^n} \phi(z, u) \quad (6)$$

... alors d'après (1), (3) et (5), on a :

$$\phi(u) = [z^*(u)]^t \,.\, E_{xx} \,.\, [z^*(u)] - 2 \,.\, E_{yx} \,.\, [z^*(u)] + \|Y\|^2 + 2u^t \,.\, \{C \,.\, [z^*(u)] - D\}$$

$$\Rightarrow \phi(u) = (E_{yx} - u^t C) \,.\, \widetilde{E}_{xx}^- \,.\, E_{xx} \,.\, \widetilde{E}_{xx}^- \,.\, (E_{xy} - C^t u)$$
$$- 2 \,.\, E_{yx} \,.\, \widetilde{E}_{xx}^- \,.\, (E_{xy} - C^t u) + \|Y\|^2 + 2u^t \,.\, \{C \,.\, \widetilde{E}_{xx}^- \,.\, (E_{xy} - C^t u) - D\}$$

Puisque E_{xx}^- est par hypothèse symétrique, on peut remarquer que $\widetilde{E}_{xx}^-$ l'est également si bien qu'en développant l'expression ci-dessus et en effectuant des mises en facteur adéquates, on obtient :

$$\phi(u) = \|\dot{Y}\|^2 - E_{yx} \widetilde{E}_{xx}^- E_{xy} + \gamma(u) \quad (7)$$

On peut donc écrire :

$$\max_{u \in \mathcal{C}_x^*} \phi(u) = \|\dot{Y}\|^2 - E_{yx} \widetilde{E}_{xx}^- E_{xy} + \max_{u \in \mathcal{C}_x^*} \gamma(u)$$

Compte tenu de (6), il s'ensuit que l'on a :

$$\max_{u \in \mathcal{C}_x^*} \phi(z^*(u), u) = \|Y\|^2 - E_{yx} \widetilde{E}_{xx}^- E_{xy} + \max_{u \in \mathcal{C}_x^*} \gamma(u) \quad (8)$$

Ceci étant dit, supposons qu'il existe $u^* \in \mathcal{C}_x^*$ tel que :

$$\gamma(u^*) = \max_{u \in \mathcal{C}_x^*} \gamma(u) \quad (9)$$

D'après (8), on est alors assuré d'avoir :

$$\phi(z^*(u^*), u^*) = \max_{u \in \mathcal{C}_x^*} \phi(z^*(u), u) \quad (10)$$

Compte tenu de (6), il s'ensuit que :

$$\phi(z^*(u^*), u^*) = \max_{u \in \mathcal{C}_x^*} \min_{z \in R^n} \phi(z, u)$$

En comparant cette dernière relation avec la relation (4), on s'aperçoit que u* et z*(u*) constitue une solution de (4), et l'on peut donc poser :

$$\begin{cases} \hat{u} = u^* \\ \hat{z} = z^*(\hat{u}) \end{cases} \tag{11}$$

Par définition de $\Pi_x^c(\dot{Y})$ et compte tenu de (2), (5) et (4), on peut alors écrire les relations suivantes :

$$\left.\begin{array}{ll} 1) & \begin{cases} \hat{u} = u^* \text{ vérifiant } (9) \\ \hat{z} = E_{xx}^- . (E_{xy} - C^t . \hat{u}) \in \mathcal{C} \end{cases} \\ \\ 2) & \min_{z \in \mathcal{C}} \|\dot{Y} - z^t \dot{X}\|^2 = \|\dot{Y} - \hat{z}^t \dot{X}\|^2 \\ \\ 3) & \Pi_x^c(\dot{Y}) = \hat{z}^t . \dot{X} \end{array}\right\} \tag{12}$$

D'après (4), (11), (10), (8) et (9), il s'ensuit que l'on a :

$$\begin{aligned} \|\dot{Y} - \Pi_x^c(\dot{Y})\|^2 &= \|\dot{Y} - \hat{z}^t . \dot{X}\|^2 & \\ &= \phi(\hat{z}, \hat{u}) & \text{d'après (4)} \\ &= \phi(z^*(u^*), u^*) & \text{d'après (11)} \\ &= \max_{u \in \mathcal{C}_x^*} \phi(z^*(u), u) & \text{d'après (10)} \\ &= \|Y\|^2 - E_{yx} \widetilde{E}_{xx}^- E_{xy} + \gamma(\hat{u}) & \text{d'après (8) et (9)} \end{aligned}$$

Compte tenu de (12) et de la dernière relation ci-dessus, on conclut que les propriétés 1) et 2) de l'énoncé sont vraies :

$$\begin{cases} 1) & \Pi_x^c(\dot{Y}) = \{(E_{yx} - \hat{u}^t . C) . E_{xx}^-\} . \dot{X} \\ 2) & \|\dot{Y} - \Pi_x^c(\dot{Y})\|^2 = \|\dot{Y}\|^2 - E_{yx} \widetilde{E}_{xx}^- E_{xy} + \gamma(\hat{u}) \end{cases} \tag{13}$$

Par ailleurs, on remarque sur la figure 1 que les points $\{\dot{Y}, \Pi_x(\dot{Y}), \Pi_x^c(\dot{Y})\}$ constituent un triangle rectangle, donc d'après le théorème de Pythagore, on peut écrire :

$$\|\Pi_x(\dot{Y}) - \Pi_x^c(\dot{Y})\|^2 = \|\dot{Y} - \Pi_x^c(\dot{Y})\|^2 - \|\dot{Y} - \Pi_x(\dot{Y})\|^2 \tag{14}$$

En remarquant que d'après 1.1, on a :

$$\|\dot{Y} - \Pi_x(\dot{Y})\|^2 = \|\dot{Y}\|^2 - E_{yx} \widetilde{E}_{xx}^- E_{xy}$$

. . . on en conclut d'après (13) et (14) que la propriété (3) de l'énoncé est vraie :

$$\|\Pi_x(\dot{Y}) - \Pi_x^c(\dot{Y})\|^2 = \gamma(\hat{u})$$

Ceci étant dit, on peut remarquer que, la matrice E_{xx} étant non négative, il en est de même de A et par conséquent, dans R^{p+q+1}, la surface d'équation $s = \gamma(u)$ est :

– soit un paraboloïde ayant sa concavité tournée vers les s négatifs (si A est définie positive),

– soit un cylindre parabolique de génératrices "parallèles" au "plan" des u et ayant sa concavité tournée vers les s négatifs (si A possède des valeurs propres nulles).

Il s'ensuit que, dans R^{p+q+1}, la surface d'équation $s = \gamma(u)$ présente un maximum (pas forcément unique s'il s'agit du cylindre parabolique) et par conséquent, l'ensemble des $\hat{u} = u^*$ vérifiant (9) n'est pas vide.

2.2. Algorithme de calcul itératif

Remarquons tout d'abord que le gradient de la fonction $\gamma(u)$ est un vecteur de R^{p+q} tel que :

$$\text{grad}\, \gamma|_u = 2 \,.\, (B - Au)$$

Par conséquent, si l'on désigne par a_{ij} et b_i les composantes de A et B, on a :

$$\begin{aligned} \text{grad}\, \gamma|_u = \theta_{R^{p+q}} &\Leftrightarrow Au = B \\ &\Leftrightarrow a_{ii} \,.\, u_i + \sum_{j=1}^{i-1} a_{ij} \,.\, u_j + \sum_{j=i+1}^{p+q} a_{ij} \,.\, u_j = b_i \\ &\Leftrightarrow u_i = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} \,.\, u_j - \sum_{j=i+1}^{p+q} a_{ij} \,.\, u_j \right) \end{aligned}$$

Pour obtenir un vecteur $\hat{u} \in C_x^*$ vérifiant la relation :

$$\gamma(\hat{u}) = \max_{u \in \mathcal{C}_x^*} \gamma(u)$$

... on peut alors utiliser l'algorithme de "montée" décrit sur la figure 2 et connu sous le nom de "gradient réduit" ; on démontre facilement que cet algorithme est convergent.

3. APPLICATION AU CODAGE DES V.A.

3.1. Introduction

Considérons d'une part une V.A. scalaire Y^α définie sur (Ω, A, P) et à valeurs dans $(R, \mathbf{B}_R)$ et soit d'autre part $\{A_1^\alpha, \ldots, A_{n_\alpha}^\alpha\}$ la suite des ensembles A-mesurables associés à une suite croissante de nombres réels donnés

$$\{y_1^\alpha \leqslant y_2^\alpha \leqslant \ldots \leqslant y_{n_\alpha - 1}^\alpha\}$$

de telle façon que :

$$\begin{aligned} A_1^\alpha &= \{\omega \in \Omega \mid -\infty < Y^\alpha(\omega) \leqslant y_1^\alpha\} \\ A_2^\alpha &= \{\omega \in \Omega \mid y_1^\alpha \leqslant Y^\alpha(\omega) < y_2^\alpha\} \\ &\vdots \\ A_{n_\alpha}^\alpha &= \{\omega \in \Omega \mid y_{n_\alpha - 1}^\alpha \leqslant Y^\alpha(\omega) \leqslant +\infty\} \end{aligned}$$

Il est évident que la famille $\{A_1^\alpha, \ldots, A_{n_\alpha}^\alpha\}$ constitue une partition de Ω et l'on définit facilement la sous-tribu $\mathcal{A}_\alpha$ engendrée par cette famille.

Ceci étant dit, afin d'accéder aux méthodes non linéaires de traitement de données, il est courant d'associer à la V.A. Y^α son "codage disjonctif" constitué par la V.A. X^α telle que :

$$X^\alpha = \begin{bmatrix} X_1^\alpha \\ \vdots \\ X_{n_\alpha}^\alpha \end{bmatrix} \qquad \text{avec : } X_j^\alpha(\omega) = 1_{A_j^\alpha}(\omega) = \begin{cases} 1 \text{ si } \omega \in A_j^\alpha \\ 0 \text{ sinon} \end{cases}$$

Dans ces conditions, si l'on pose :

$$\tilde{Y}^\alpha = [\tilde{y}^\alpha]^t \,.\, X^\alpha \quad \text{et} \quad [\tilde{y}^\alpha] = \begin{bmatrix} \tilde{y}_1^\alpha \\ \tilde{y}_2^\alpha \\ \vdots \\ \tilde{y}_{n_\alpha - 1}^\alpha \\ \tilde{y}_{n_\alpha}^\alpha \end{bmatrix} \qquad \text{avec : } \begin{cases} -\infty < \tilde{y}_1^\alpha < y_1^\alpha \\ y_{j-1}^\alpha \leqslant \tilde{y}_j^\alpha < y_j^\alpha \\ y_{n_\alpha - 1}^\alpha \leqslant \tilde{y}_{n\alpha}^\alpha < +\infty \end{cases}$$

... il est évident que, si les constantes $\{y_1^\alpha, \ldots, y_{n\alpha-1}^\alpha\}$ sont "bien choisies", alors on a approximativement :

$$\tilde{y}^\alpha \simeq y^\alpha$$

On dit alors que la V.A. étagée $\tilde{Y}^\alpha$ est un "codage" de la V.A. Y^α tandis que les constantes $\{\tilde{y}_1^\alpha, \ldots, \tilde{y}_{n\alpha}^\alpha\}$ sont appelées les "codes" de ce codage. On ne manquera pas de remarquer que l'on a :

$$\tilde{y}_1^\alpha \leqslant \tilde{y}_2^\alpha \leqslant \ldots \leqslant \tilde{y}_{n\alpha}^\alpha$$

3.2. Codage par analyse en composantes principales

Rappels

Soit $\{Y^1, \ldots, Y^N\}$ une famille de N V.A. scalaires définies sur (Ω, A, P) et à valeurs dans $(R, \mathbf{B}_R)$ et soient d'autre part $\{X^1, \ldots, X^N\}$ et $\{\tilde{Y}^1, \ldots, \tilde{Y}^N\}$ les V.A. associées définies au paragraphe précédent. Si l'on désigne par $\{\dot{\phi}_0, \dot{\phi}_1, \ldots\}$ les composantes principales généralisées de Carroll associées à la famille $\{\dot{X}^1, \ldots, \dot{X}^N\}$ et si l'on désigne par $r(\dot{\phi}_1 \mid \dot{X}^\alpha)$ le coefficient de corrélation multiple de $\dot{\phi}_1$ avec $\dot{X}^\alpha$, alors en supposant les composantes $\{\dot{\phi}_i\}$ rangées par ordre décroissant de valeurs propres, on peut vérifier que l'on a :

1) $$\dot{\phi}_0 = \dot{1}_\Omega$$

2) $$\sum_{\alpha=1}^{N} r^2\,(\dot{\phi}_1 \mid \dot{X}^\alpha) \text{ est maximum}$$

3) $$\Pi_{X^\alpha}(\dot{\phi}_1) = E^{\mathcal{A}_\alpha}(\dot{\phi}_1) = \dot{\hat{Y}}^\alpha = \text{fonction mesurable de } \tilde{y}^\alpha$$

Compte tenu de la nature de l'opérateur Π_{x^α} on en déduit que Y^α est de la forme suivante :

$$\dot{\hat{Y}}^\alpha = [\hat{y}^\alpha]^t \cdot \dot{X}^\alpha$$

$$[y^\alpha] = \begin{bmatrix} \hat{y}_1^\alpha \\ \vdots \\ \hat{y}_{n\alpha}^\alpha \end{bmatrix}$$

Certains auteurs (MASSON, SAPORTA, etc.) suggèrent d'utiliser la V.A. $\hat{Y}^\alpha$ définie ci-dessous comme un nouveau codage (optimal) de la V.A. Y^α :

$$\hat{y}^\alpha = [\hat{y}^\alpha]^t \cdot X^\alpha$$

Remarque

Les constantes $\{\tilde{y}_1^\alpha\}$ étaient initialement telles que :

$$\tilde{y}_1^\alpha \leqslant \tilde{y}_2^\alpha \leqslant \ldots \leqslant \tilde{y}_{n\alpha}^\alpha$$

Par contre les constantes $\{\hat{y}_i^\alpha\}$ne sont soumises à aucune contrainte, et si l'on désire avoir

$$\hat{y}_1^\alpha \leqslant \hat{y}_2^\alpha \leqslant \ldots \leqslant \hat{y}_{n\alpha}^\alpha$$

... il est nécessaire de définir $\hat{Y}^\alpha$ autrement que par la méthode proposée ci-dessus.

Recherche d'un codage croissant

Considérons la matrice C_2^α à $p_\alpha = n_\alpha - 1$ lignes et n_α colonnes telle que :

$$C_2^\alpha = \underbrace{\begin{bmatrix} 1 & -1 & & & & & & \\ & 1 & -1 & & & \text{O} & & \\ & & 1 & -1 & & & & \\ & & & \ddots & \ddots & & & \\ & & & & 1 & -1 & & \\ & \text{O} & & & & 1 & -1 & \\ & & & & & & 1 & -1 \end{bmatrix}}_{n_\alpha} \; \Big\updownarrow n_\alpha - 1$$

Soit d'autre part $\theta_{R^{n\alpha}}$ le vecteur nul de $R^{n\alpha}$ et posons :

$$D_2^\alpha = \theta_{R^{n\alpha}}$$

Ceci étant dit, soit $\mathcal{C}_{x^\alpha}$ le cône convexe de H_{x^α} tel que :

$$\dot{Z} \in \mathcal{C}_x \Longleftrightarrow \dot{Z} = z^t \cdot \dot{X}^\alpha \quad \text{avec} \quad C_2^\alpha \cdot z \leqslant D_2^\alpha$$

$$\Longleftrightarrow \dot{Z} = z^t \cdot \dot{X}^\alpha \quad \text{avec} \quad z_1 \leqslant z_2 \leqslant \ldots \leqslant z_{n\alpha}$$

Dans ces conditions, considérons les C.E.V.A. $\hat{\hat{\dot{Y}}}{}^{\alpha}_{1} = \Pi^{c}_{x^{\alpha}}(\dot{\phi}_1)$ et $\hat{\hat{\dot{Y}}}{}^{\alpha}_{-1} = \Pi^{c}_{x^{\alpha}}(-\dot{\phi}_1)$ projections orthogonales de $\dot{\phi}_1$ et $-\dot{\phi}_1$ sur $\mathcal{C}_{x^{\alpha}}$. On est alors assuré que pour $\epsilon = \pm 1$, on a :

1) $$\begin{cases} \hat{\hat{\dot{Y}}}{}^{\alpha}_{\epsilon} = [\hat{\hat{y}}{}^{\alpha}_{\epsilon}]^{t} \,.\, X \\ \hat{\hat{y}}{}^{\alpha}_{\epsilon,1} \leqslant \hat{\hat{y}}{}^{\alpha}_{\epsilon,2} \leqslant \ldots \leqslant \hat{\hat{y}}{}^{\alpha}_{\epsilon,n\alpha} \end{cases}$$

2) $$\|\hat{\hat{\dot{Y}}}{}^{\alpha}_{\epsilon} - \epsilon \,.\, \hat{\dot{Y}}^{\alpha}\| \quad \text{est minimum}$$

Nous proposons de prendre pour codage optimal de Y^{α} la V.A. $\hat{\hat{Y}}{}^{\alpha}_{\epsilon} = [\hat{\hat{y}}{}^{\alpha}_{\epsilon}]^{t} \,.\, X^{\alpha}$ pour laquelle on a :

$$\|\hat{\hat{\dot{Y}}}{}^{\alpha}_{\epsilon} - \epsilon \,.\, \hat{\hat{\dot{Y}}}{}^{\alpha}\| \leqslant \|\hat{\hat{\dot{Y}}}{}^{\alpha}_{-\epsilon} + \epsilon \,.\, \hat{\hat{\dot{Y}}}{}^{\alpha}\|$$

Si nous posons :

$$\gamma_{\epsilon}(u) = 2 \,.\, B^{t}_{\epsilon} \,.\, u^{t} \,.\, A \,.\, u$$

$$\text{avec :} \begin{cases} A = [C^{\alpha}_{2}] \,.\, \tilde{E}^{-}_{x^{\alpha}x^{\alpha}} \,.\, [C^{\alpha}_{2}]^{t} \\ B_{\epsilon} = \epsilon \,.\, [C^{\alpha}_{2}] \,.\, E^{-}_{x^{\alpha}x^{\alpha}} \,.\, E_{x^{\alpha}\phi_1} \end{cases}$$

. . . alors, d'après le théorème précédent, on a :

$$\gamma_{\epsilon}(\hat{u}_{\epsilon}) = \max_{u \in \mathcal{C}^{*}_{x}} \gamma_{\epsilon}(u)$$

$$\Rightarrow \begin{cases} \hat{\hat{\dot{Y}}}{}^{\alpha}_{\epsilon} = (\epsilon \,.\, E_{\phi 1 X^{\alpha}} - \hat{u}^{t}_{\epsilon} \,.\, C^{\alpha}_{2}) \,.\, E_{x^{\alpha}x^{\alpha}} \,.\, \dot{X}^{\alpha} \\ \|\hat{\hat{\dot{Y}}}{}^{\alpha}_{\epsilon} - \epsilon \,.\, \hat{\dot{Y}}^{\alpha}\|^{2} = \gamma_{\epsilon}(\hat{u}_{\epsilon}) \end{cases}$$

On choisira donc pour ϵ la valeur +1 ou −1 telle que :

$$\gamma_{\epsilon}(\hat{u}_{\epsilon}) \leqslant \gamma_{-\epsilon}(\hat{u}_{-\epsilon})$$

4. BIBLIOGRAPHIE

CAZES P. et TURPIN P.Y. (1971). – Régression sous contraintes, application à l'estimation de la courbe granulométrique d'un aérosol. *Revue de Statistique Appliquée,* 1971, vol. XIX, n° 4.

CAZES P. (1978). – Méthodes de régression, la régression sous-contrainte. Les cahiers de l'analyse des données, vol. 2, Dunod.

LUENBERGER D.G. (1969). – Optimization by vector space methods. Wiley.

CARROL J.D. (1968). – Generalisation of canonical correlation analysis to three or more sets of variables. – *Poc. Amer. Ass.,* 1968 (p. 227-228).

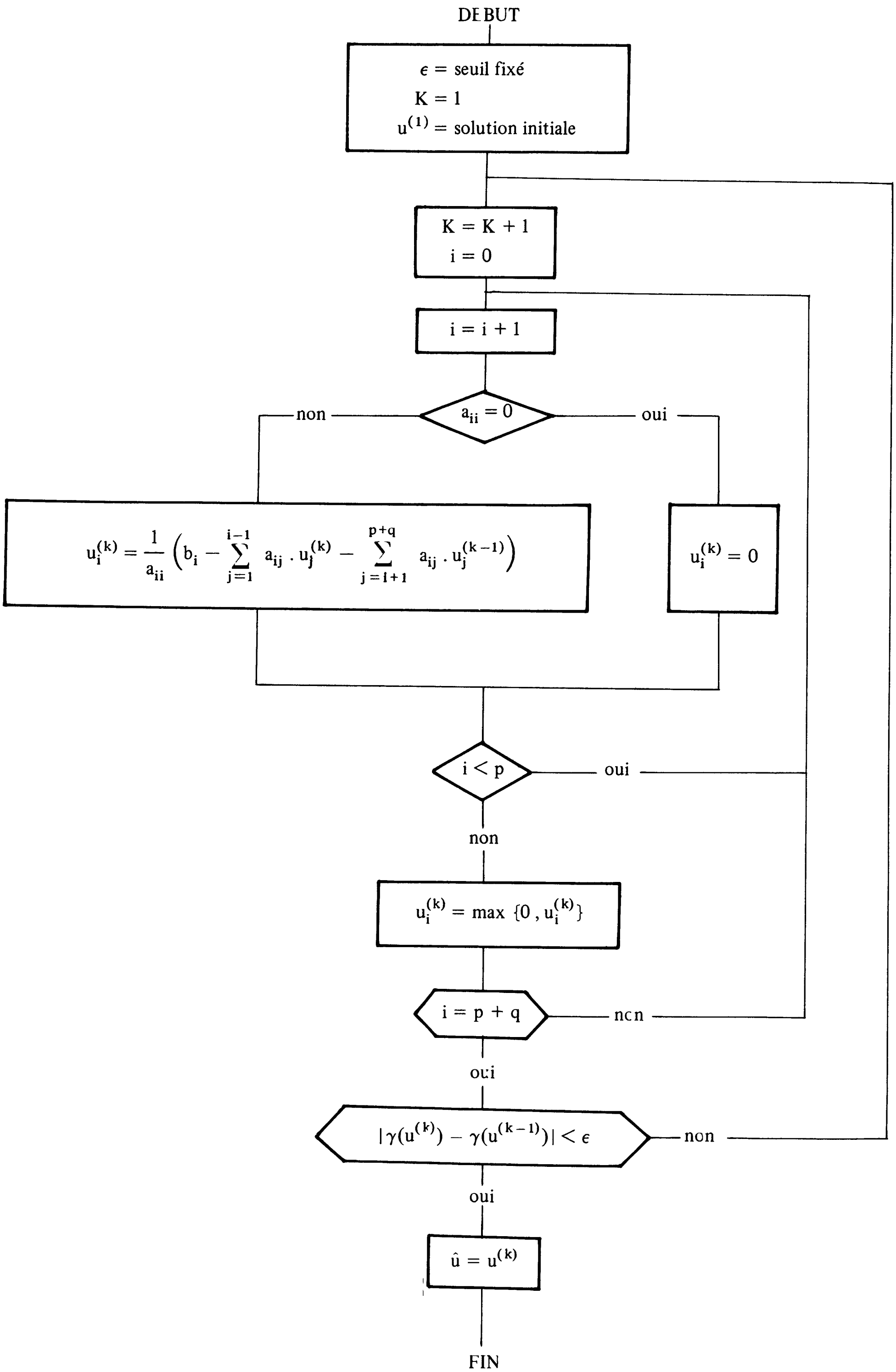
DEBUT
ε = seuil fixé
K = 1
u(1) = solution initiale
K = K + 1
i = 0
i = i + 1
non
aii = 0
oui
ui(k) = 1/aii (bi − Σj=1..i−1 aij . uj(k) − Σj=i+1..p+q aij . uj(k−1))
ui(k) = 0
i < p
oui
non
ui(k) = max {0 , ui(k)}
i = p + q
non
oui
|γ(u(k)) − γ(u(k−1))| < ε
non
oui
û = u(k)
FIN

FIGURE 2

TENENHAUSS M. (1976). – Analyse en composantes principales d'un ensemble de variables nominales ou numériques – *Cahiers de Recherche du CESA*, N° 40/1976.

SAPORTA A. (1975). – 1) Liaison entre plusieurs ensembles de variables et codage de données qualitatives – Thèse 3e cycle, université Paris VI.
2) Dependance et codage de deux variables aléatoires – *Rev. Stat. Appliquée 23*, N° 1.

MASSON M. (1974). – Processus linéaires et analyse de données non linéaires – Thèse Doc. d'état université Paris VI.