

# STATISTIQUE ET ANALYSE DES DONNÉES

G. DROUET D' AUBIGNY

## **L'utilisation des méthodes ordinales en analyse des données**

*Statistique et analyse des données*, tome 1, n° 2 (1976), p. 63-84.

[http://www.numdam.org/item?id=SAD\\_1976\\_\\_1\\_2\\_63\\_0](http://www.numdam.org/item?id=SAD_1976__1_2_63_0)

© Association pour la statistique et ses utilisations, 1976, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/legal.php>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

L'UTILISATION DES METHODES ORDINALES  
en ANALYSE DES DONNEES

par

G. DROUET D'AUBIGNY (\*)

I - INTRODUCTION

Avec le développement des moyens informatiques, de nombreuses propositions méthodologiques d'analyse de données ont marqué la dernière décennie. Leur utilisation en sciences sociales se heurte, le plus souvent, aux difficultés de détermination du type d'information contenue dans les données.

Dans ce cas YOUNG et de LEEUW (21) proposent de déterminer le niveau de mesure des observations, de façon empirique : ils effectuent plusieurs analyses des données étudiées, sous différentes hypothèses concernant l'échelle de mesure. Le niveau retenu est le plus élevé de ceux qui conduisent au meilleur ajustement du modèle descriptif aux données.

Nous montrons qu'une démarche aussi coûteuse s'avère inutile si notre intérêt premier ne réside pas dans l'explicitation du niveau de mesure des données. Le résultat est tout d'abord établi à propos des méthodes d'analyse du triple  $(I, \mathcal{K}_I, P_I)$  - Puis nous montrons qu'il reste valide pour l'ensemble des méthodes exprimables dans notre formulation probabiliste (8,9), telles que l'analyse des différences individuelles ou l'analyse de données de dominance.

Notre étude nous conduit de plus à nous interroger sur l'équivalence couramment admise entre l'analyse d'un tableau de distances euclidiennes et l'étude de la structure de la matrice de Gram associée, dans le cadre de méthodes d'ajustement aux moindres carrés.

## II - DEFINITION ET PROPRIETES D'UN INDICE DE CORRELATION ISOTONE

Les méthodes de regression étudient la liaison entre une variable aléatoire  $Y$ , à valeurs réelles, et un élément aléatoire  $X$ , défini sur le même espace probabilisé  $(\Omega, \mathcal{G}, P)$  -  $X$  est en général à valeurs dans un espace mesurable  $(E, \mathcal{E})$  où  $E$  est un espace vectoriel réel. Lorsque la seule information pertinente sur le phénomène observé est traduite par une relation d'ordre  $\mathcal{R}$  définie sur  $E$ , notre intérêt est limité aux éléments  $\mathcal{U}$  de  $\mathcal{E}$  dont l'indicatrice  $\mathbb{1}_{\mathcal{U}}(\cdot)$  est une fonction isotone de  $(E, \mathcal{R})$  dans  $(\{0,1\}, \succcurlyeq)$ . On définit ainsi une classe  $\mathcal{U}$  de parties de  $E$  qui a structure de sous -  $\sigma$  - treillis unitaire de  $\mathcal{E}$ .

### Définition 1 (2)

- Une classe  $\mathcal{U}$  de parties de  $E$  est un  $\sigma$ -treillis si elle est stable par union et intersection dénombrables. Elle est unitaire si elle contient  $\emptyset$  et  $E$ .
- une fonction  $m$  définie sur  $E$  et à valeurs dans  $\bar{\mathbb{R}}$  est  $\mathcal{U}$ -mesurable ( $m \in \mathcal{M}(\mathcal{U})$  si  $m^{-1}([x, +\infty])$ ) appartient à  $\mathcal{U}$  pour tout réel  $x$ .

Par suite  $m$  est élément de  $\mathcal{M}(\mathcal{U})$  si et seulement si  $m$  est une fonction isotone (i.e, monotone croissante) de  $(E, \mathcal{R})$  dans  $(\mathbb{R}, \succcurlyeq)$  et  $\mathcal{E}$  mesurable.

### Définition 2

Le couple  $(E, \mathcal{U})$  définit un espace mesurable ordonné et toute fonction :

$$x : (\Omega, \mathcal{G}, P) \longrightarrow (E, \mathcal{U})$$

telle que :

$\mathcal{A} = x^{-1}(\mathcal{U})$  soit une partie de  $\mathcal{G}$  est une variable aléatoire ordinale

On vérifie aisément que  $\mathcal{B}$  est un sous- $\sigma$ -treillis unitaire de  $\mathcal{A}$  et non une sous tribu puisque  $\mathcal{B}$  n'est pas stable par dualité. On peut cependant construire la notion d'espérance conditionnelle par rapport à  $\mathcal{B}$  en termes de dérivée de Radon Nikodym (2). Notre exposé étant limité aux méthodes de moindres carrés, nous nous limitons dans la suite à l'étude de variables de carré intégrable (éléments donc de  $L^2_{\mathbb{R}}(\mathcal{A}) = L^2_{\mathbb{R}}(\Omega, \mathcal{A}, P)$ ).

### Théorème 1 (2)

i) l'ensemble  $L^2_{\mathbb{R}}(\mathcal{B})$  des variables aléatoires, à valeurs réelles,  $\mathcal{B}$ -mesurables vérifie.

$$L^2_{\mathbb{R}}(\mathcal{B}) = L^2_{\mathbb{R}}(\mathcal{A}) \cap \{m \circ X ; m \in \mathcal{M}(\mathcal{U})\}$$

ii)  $L^2_{\mathbb{R}}(\mathcal{B})$  est un cône convexe fermé dans  $L^2_{\mathbb{R}}(\mathcal{A})$

iii)  $L^2_{\mathbb{R}}(\mathcal{B})$  est un treillis  $\sigma$ -complet dans  $L^2_{\mathbb{R}}(\mathcal{A})$

La structure algébrique (iii) de  $L^2_{\mathbb{R}}(\mathcal{B})$  donne un fondement théorique à l'algorithme "inverse-rank-image" de GUTTMAN (11). Nous nous intéressons ici à la structure géométrique (ii) de  $L^2_{\mathbb{R}}(\mathcal{B})$  qui assure l'unicité de la projection orthogonale de tout élément  $Y$  de  $L^2_{\mathbb{R}}(\mathcal{A})$  sur  $L^2_{\mathbb{R}}(\mathcal{B})$  comme point le plus proche de  $Y$ , dans  $L^2_{\mathbb{R}}(\mathcal{B})$ . On peut donc poser la

### Définition 3

On appelle espérance conditionnelle de  $Y$  par rapport à  $\mathcal{B}$ , l'unique point de  $L^2_{\mathbb{R}}(\mathcal{B})$ , le plus proche de  $Y$ , que l'on note  $E(Y/\mathcal{B})$ . Il vérifie donc

$$\int_{\Omega} |Y - E(Y/\mathcal{B})|^2 dP \leq \int_{\Omega} |Y - m \circ X|^2 dP$$

où  $m$  est élément de  $\mathcal{M}(\mathcal{U})$  et  $E(Y/\mathcal{B})$  est identifié à l'une de ses versions suivant l'abus de langage classique.

Une deuxième conséquence importante de la structure géométrique de  $L^2_{\mathbb{R}}(\mathcal{B})$  est que la formule de décomposition de la variance reste valide dans le cadre ordinal.

$$\text{Var}(Y) = \text{Var}(E(Y/\mathcal{B})) + E(|Y - E(Y/\mathcal{B})|^2)$$

L'analogie avec le rapport de corrélation conduit donc à la

Définition 4 (8)

L'indice de corrélation isotone est défini par

$$v_{Y/X}^2 = \rho^2(Y, E(Y/\mathcal{B})) = 1 - \frac{E(|Y - E(Y/\mathcal{B})|^2)}{\text{Var}(Y)}$$

ou  $\rho^2(\dots)$  est le coefficient de corrélation linéaire de PEARSON-BRAVAIS

Cet indice asymétrique a une interprétation immédiate en terme de réduction d'erreur (au sens de BLALOCK (1)) puisque

$$v_{Y/X}^2 = \frac{\text{Var}(E(Y/\mathcal{B}))}{\text{Var}(Y)}$$

ce qui assure que son domaine de variations est  $[0, 1]$

Proposition 1

- i)  $v_{Y/X}^2 = v_{Y/U}^2$  si est seulement si  $U = m \circ X$ ,  $m \in \mathcal{A}(Y)$
- ii)  $v_{Y/X}^2 = 0$  si est seulement si tout élément non dégénéré de  $L_{\mathbb{R}}^2(\mathcal{B})$  a une corrélation linéaire négative avec  $Y$ . Alors  $E(Y)$  appartient à  $E(Y/\mathcal{B})$ .
- iii) S'il existe un élément non dégénéré de  $L_{\mathbb{R}}^2(\mathcal{B})$  dont la corrélation linéaire avec  $Y$  est positive,  $E(Y/\mathcal{B})$  vérifie :

$$v_{Y/X}^2 = \text{Max} \{ \rho^2(Y, V) ; V \in L_{\mathbb{R}}^2(\mathcal{B}) \}$$

Éléments de démonstration

i) est une conséquence directe de la définition. Une démonstration directe de ii) et iii) est donnée dans (8). Plus simplement, il suffit de remarquer que ces deux propriétés sont des conséquences triviales du théorème de dualité convexe.

Les propriétés i) et ii) de la proposition 1 font de  $v_{Y/X}^2$  un indicateur de dépendance positive au sens de LEHMANN. Mais puisque  $\mathcal{B}$  n'est pas une tribu, l'opérateur espérance conditionnelle  $E(\cdot/\mathcal{B})$  n'est plus linéaire.

Il est donc impossible d'évaluer  $v_{Y/X}^2$  sans construire  $E(Y/\mathcal{B})$ . On retrouve le problème classique en régression non linéaire.

Lorsque l'on a observé un échantillon pondéré  $\{(X_i, Y_i), p_i\} ; i = 1 \dots n$  issu de  $(X, Y)$ , on peut construire l'indice de corrélation isotone empirique

$$\hat{\rho}_{Y/X}^2 = 1 - S_{Y/X}^2$$

où  $S_{Y/X}^2 = \text{Inf} \{ S_{Y/X}^2(m) ; m \in \mathcal{M}(u) \}$

$$S_{Y/X}^2(m) = \frac{\sum_{i=1}^n p_i |y_i - m(x_i)|^2}{\sum_{i=1}^n p_i |y_i - \bar{y}|^2} = \frac{\|y - m(x)\|_{D_n}^2}{\|y - \bar{y} \cdot J\|_{D_n}^2}$$

où l'on note  $\|z\|_{D_n}^2$  la norme euclidienne du vecteur  $z$  de  $\mathbb{R}^n$  pour la métrique des poids.

$$D_n = \text{diag}(p_i), p_i > 0, \quad \sum_{i=1}^n p_i = 1, \quad \bar{y} = \sum_{i=1}^n p_i y_i$$

et  $J$  est le vecteur de  $\mathbb{R}^n$  de composantes toutes égales à 1.

L'évaluation de  $S_{Y/X}^2$  (appelé STRESS) se fait en utilisant l'algorithme pool-adjacent-violator ([2], [13]).

L'essentiel de la discussion future repose sur le résultat très simple suivant, démontré dans un contexte et sous une forme différentes par MAYER (14) :

### Proposition 2

Si l'espérance conditionnelle de  $Y$  sachant  $\mathcal{G}$  est une fonction affine de  $X$ , alors  $\hat{\rho}_{Y/X}^2$  converge presque sûrement vers  $\rho^2(X, Y)$

### Elements de preuve

Le résultat iii) de la proposition 1 nous assure - sous nos hypothèses - que lorsque  $\hat{\rho}_{X/Y}^2$  converge, c'est vers le meilleur prédicteur linéaire de  $Y$  - La preuve consiste donc à montrer la convergence presque sûre de

Remarques :

- i) Le coefficient de corrélation linéaire empirique est un estimateur fortement consistant de  $\rho^2(X,Y)$  et donc la proposition 2 assure que pour  $n$  assez grand :

$$P(|\hat{v}_{Y/X}^2 - r_{XY}^2| > \varepsilon) < \varepsilon \text{ pour tout réel } \varepsilon \text{ positif}$$

et puisque

$$r_{XY}^2 \leq \hat{v}_{Y/X}^2 = \hat{v}_{Y/U}^2 \quad \text{où } U \text{ réalise } S_{Y/X}^2$$

on peut écrire pour  $n$  grand

$$r_{UY}^2 - \varepsilon \leq \hat{v}_{Y/X}^2 - \varepsilon \leq r_{XY}^2 \leq \hat{v}_{Y/U}^2$$

ce qui permet d'utiliser  $\hat{v}_{Y/X}^2$  sous l'hypothèse linéaire, comme estimateur biaisé (par excès) de  $\rho^2(X/Y)$ .

- ii) La comparaison des comportements de  $\hat{v}_{Y/X}^2$  et de l'indice de détermination métrique proposé par YOUNG (20) est éclairée par le résultat précédent. Le biais dû au choix de  $\hat{v}_{Y/X}^2$  étant asymptotiquement négligeable, l'écart entre les estimations des deux indices traduit en fait le degré de non respect de l'hypothèse linéaire. Les simulations de SHERMAN et YOUNG (20), interprétées dans cet esprit, montrent, s'il en était besoin, la robustesse des méthodes linéaires aux violations de l'hypothèse de linéarité.

### III - APPLICATION A L'ANALYSE DU TRIPLE

L'expérimentation en sciences sociales conduit très souvent à l'analyse d'un triple  $(I, \delta_I, P_I)$  où  $I$  est un ensemble d'objets auxquels est affectée une pondération:

$$P_I : i \rightsquigarrow p_i, \quad p_i > 0, \quad \sum_{i=1}^n p_i = 1, \quad \text{si } |I| = n.$$

et  $\delta_I$  est un indice de dissimilarité entre objets de  $I$  :

$$\delta_I : (i, e) \rightsquigarrow \delta_{ie} \quad i, e \in I$$

dont l'échelle de mesure - nous excluons à dessein le cas de données nominales - est l'un des trois suivants.

i) Les données sont d'échelle ordinale si l'information contenue dans le tableau de données  $\Delta$  est préservée par toute bijection isotone

$$t_I = m_0 \delta_I$$

ii) Les données sont d'échelle d'intervalle si l'information contenue dans le tableau  $\Delta$  est préservée pour toute bijection affine.

$$t_I = a \delta_I + c, \quad a \in \mathbb{R}^+ \setminus \{0\}, \quad c \in \mathbb{R}$$

iii) Les données sont d'échelle de proportion si l'information contenue dans le tableau  $\Delta$  est préservée par toute bijection linéaire.

$$t_I = a \delta_I, \quad a \in \mathbb{R}^+ \setminus \{0\}$$

Par suite les échelles de mesure ii) et iii) accordent de l'importance à la valeur numérique de l'indice de dissimilarité - conduisant ainsi aux méthodes dites métriques - alors que les échelles ordinales - conduisant aux méthodes dites ordinales (ou non métrique) - ne retiennent que l'ordonnance induite par  $\delta_I$  sur  $I$ .

### III-1 Le problème posé en analyse du triple

Etant donné le triple  $(I, \delta_I, P_I)$  nous nous proposons de représenter  $I$  dans un espace vectoriel normé  $V$  (mais notons  $\| \cdot \|_V$  la norme sur  $V$ )

$$X : I \longrightarrow V$$

$$i \longmapsto X(i) = (x_k(i))_{k=1 \dots r}, \quad r = \dim(V),$$

de telle sorte que

i) la dimension de  $V$  soit minimale

ii) l'ajustement entre les données  $\Delta$  et les distances interpoints reconstruites  $D$  soit optimale (au sens des moindres carrés)

$$D = (d_{ie})_{i,e \in I} \quad d_{ie} = \|X(i) - X(e)\|_V$$

Lorsque l'on connaît  $r$  nous sommes confrontés à un problème de regression simple, qui s'exprime en termes de moindres carrés:

$$(P_1) \quad \min_{D_p} \|D - T\|^2 = \text{Min} !$$

où  $D_p$  est la métrique des poids dans  $\mathbb{R}^p$ ,  $p = \frac{n(n-1)}{2}$ ,  $D$  et  $T$  étant identifiés à des vecteurs de  $\mathbb{R}^p$  ainsi

$$T = (t_b)_{b \in B} \quad \mathcal{B} = \{ (i, l) ; i, l \in I, \quad i \neq l \},$$

$$D_p = \text{diag}(P_b), \quad \text{où } P_b = P_i P_l \quad \text{si } b = (i, l).$$

Lorsque  $\delta_I$  est mesurée sur une échelle ordinale, le problème  $(P_1)$  est résolu par l'algorithme de KRUSKAL (13) qui calcule

$$(1) \quad d(D(I), \mathcal{C}) = \text{MIN} \left[ \text{MIN} \left( \frac{\|D - T\|_{D_p}^2}{\|D - d.J\|_{D_p}^2} ; T \in \mathcal{C} \right) ; D \in D(I) \right]$$

où le cône convexe



### III-2 Comparaison à l'analyse factorielle du triple $(I, \delta_I, P_I)$

Soit  $\Pi$  le projecteur  $D_n$  orthogonal sur l'hyperplan  $H_J$  orthogonal à la droite des constantes (engendrée par le vecteur  $J$ ) de  $\mathbb{R}^n$ :

$$\Pi = (I_n - D_n J^t J) \text{ où } I_n \text{ est la matrice identité sur } \mathbb{R}^n.$$

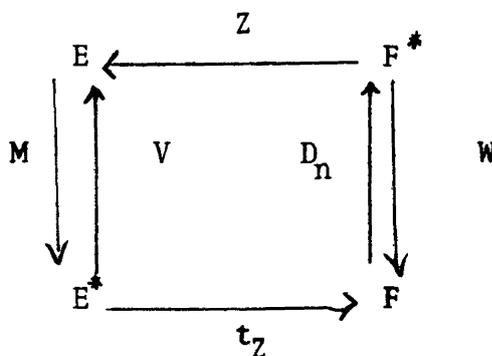
La méthode de double centrage de TORGERSON (19) associe au tableau d'indices de distances  $\Delta$  la matrice de GRAMM

$$W = -\frac{1}{2} \Pi \Delta^{(2)} \Pi \quad \Delta^2 = (\delta_{il}^2)_{i,l \in I}$$

#### Remarque

La matrice  $W$  est une matrice de coefficients d'association, au sens de GOVER (10) qui peut ne pas être semi-définie positive. Nous n'abordons pas ce problème ici et renvoyons le lecteur à (3) ou (8) pour une discussion des difficultés alors rencontrées.

L'opérateur associé à  $W$  est le projecteur aux moindres carrés sur la variété linéaire  $V$  de  $\mathbb{R}^n$ , support du nuage de points, que l'on cherche à décrire (C.F. OBENCHAIM (15)) - Aussi la démarche adoptée en analyse factorielle du triple  $(I, \delta_I, P_I)$  consiste à construire  $X$  en interprétant l'opérateur  $W \circ D_n$  dans le schéma de dualité (3)



$$W = {}^t Z \circ M \circ Z$$

$$V = Z \circ D_n \circ {}^t Z$$

$$M = V^{-1}$$

où la matrice centrée  $Z = Y\Pi$  est associée à un tableau fictif de données  $Y$  figurant des caractères latents en nombre inconnu, mesurés sur  $I$ , qui ont permis la construction de  $\Delta$ . Le modèle implicite du processus d'élaboration des données doit être mis en question par l'utilisateur éventuel. D'un point de vue mathématique, il conduit à juger de la qualité de la représentation  $X$  par le critère usuel en analyse en composantes principales (3):

$$\lambda = \frac{I_V^2}{I_g}$$

rapport des inerties du nuage  $\mathcal{N} = \{(Y(i), p_i) ; i \in I\}$  de  $\mathbb{R}^n$  par rapport à la variété affine  $D_n$  - orthogonale à  $V$  d'une part et au barycentre  $g$  de  $\mathcal{N}$  d'autre part. Il est immédiat de vérifier qu'ici

$$I_g = \frac{1}{2} \sum_{i,l} p_i p_l d_{il}^2 = \frac{1}{2} \sum \{ p_i p_l d_{il}^2 ; (i,l) \in I \times I \}$$

$$I_V = \frac{1}{2} \sum_{i,l} p_i p_l d_{il}^2 = \frac{1}{2} \sum \{ p_i p_l d_{il}^2 ; (i,l) \in I \times I \}$$

ce qui conduit à un critère  $\lambda$  bien peu satisfaisant dans le cas présent.

L'inadéquation de cette démarche à la situation apparaît, d'un point de vue statistique, en terme d'invariance. Pour être reconnue valide, une méthode d'analyse doit respecter le niveau d'invariance des données. Cependant alors que l'analyse en composantes principales n'est invariante que par symétrie et translations,  $W$  et  $\Delta^{(2)}$  sont deux statistiques maximales  $L(p)$  - invariantes (15), où nous avons noté  $L(p)$  le groupe des transformations affines de plein rang, opérant sur  $F$ .

La méthode affine invariante la plus proche (en termes de critère optimisé) de l'analyse factorielle du triple, semble être celle proposée par OBENCHAIN. (16). On se propose de minimiser

$$\| \Delta^{(2)} - D^{(2)} \|_{D_p}^2 = 4 \| W D_n - X X^t \|_{D_p}^2 + 2 ({}^t J D_p e)^2 + 2n {}^t e D_p e$$

avec

i)  $\|W D_n - {}^t X X\|_{D_p}^2$  Le critère minimisé en analyse en composantes principales.

ii)  $e$  le vecteur de  $\mathbb{R}^D$  comptabilisant les écarts entre distances carrées des  $n$  points représentatifs de  $I$  à leur barycentres, calculées respectivement dans les configurations  $Y$  et  $X$ .

C'est dire que les méthodes affines invariantes accordent plus d'importance à l'ajustement des distances aux barycentres que l'approche factorielle. Celle-ci, comme toute méthode d'analyse de la structure d'une matrice de GRAMM, optimise en fait l'ajustement des angles entre vecteurs représentatifs.

De plus, en écrivant

$$\begin{aligned} \|\Delta^{(2)} - D^{(2)}\|_{D_p} &= \sum \{ p_b (\delta_b^2 - d_b^2) ; b \in B \} \\ &= \sum \{ (p_b (\delta_b + d_b)^2) \cdot (\delta_b - d_b)^2 ; b \in B \} \end{aligned}$$

on constate qu'un tel critère accorde plus de poids aux grandes distances, que la méthode affine invariante de de LEEUW (7), qui minimise

$$\|\Delta - D\|_{D_p} = \sum \{ p_b (\delta_b - d_b)^2 ; b \in B \}$$

critère qui respecte au mieux les données traitées.

Dès lors que l'on admet la différence de finalité entre méthodes d'analyse d'un tableau de distances d'une part et de description de la structure d'une matrice de GRAMM, il convient de définir une méthode ordinaire dans le second cas. Une telle méthode d'analyse factorielle ordinaire fut proposée par GUTTMAN et LINGOES (12) qui minimisent le "coefficient d'aliénation":

$$K = [1 - r^2(W, T)]^{1/2}$$

où  $r$  est le coefficient de corrélation linéaire et  $T$  est un tableau obtenu par permutation des termes du tableau  ${}^t X X$  :

$$T = m({}^t X X)$$

en utilisant l'algorithme "inverse-rank-image" de GUTTMAN, pour construire  $m$ , à chaque itération.

Nous étudions actuellement un modèle équivalent construisant  $X$  en termes de régression isotone aux moindres carrés, par maximisation de  $\hat{v}^2_{t_{XX}/WD_n}$

La configuration  $X$  ainsi construite réalise le minimum de

$$\|WD_n - T\|_{D_p}^2 \quad \text{où } T = m(t_{XX}) \text{ et } m \text{ est une fonction isotone de } (E, ) \text{ dans } (R, ).$$

L'application de la proposition 2 montre l'utilisation possible de cette méthode dans un but de protection contre les violations de l'hypothèse linéaire, dues à des transformations isotones des données.

Remarque :

L'intérêt pratique d'une telle méthode a été souligné par GUTTMAN (12) : son utilisation pour la description des relations d'indépendance d'une famille de caractères ordinaux à partir de la matrice des corrélations de KENDALL s'est révélée très encourageante. Cependant le coût accru de mise en oeuvre de ce modèle ainsi que la robustesse des méthodes linéaires, signalée plus haut, limite son emploi au cas de fortes non-linéarités.

#### IV - EXTENSION DE CES RESULTATS

L'extension des méthodes de régression isotone, à l'étude de la dépendance d'un vecteur aléatoire  $Y$  à valeurs dans  $\mathbb{R}^m$ , de carré intégrable, par rapport à une variable aléatoire ordinaire  $X$ , défini sur le même espace probabilisé  $(\Omega, \mathcal{G}, P)$  nécessite une hypothèse complémentaire. Nous supposons ici l'existence d'une forme linéaire  $\langle \cdot, W \rangle$  sur  $L^2_{\mathbb{R}^m}(\mathcal{G})$ , telle que soit élément de  $L^2_{\mathbb{R}}(\mathcal{G})$  la variable aléatoire.

$$Z : \omega \rightsquigarrow Z(\omega) = \langle Y(\omega), W \rangle$$

La mise en oeuvre du modèle linéaire ordinal consiste à construire la meilleure approximation de  $\langle \cdot, w \rangle$  au sens des moindres carrés, en étudiant la régression isotone de  $Z$  en  $X$ .

#### Remarque

La régression multiple et la régression polynomiale isotones entrent dans ce modèle qui, de façon plus générale, couvre tous les modèles de mesures conjointes (YOUNG (20)) - Ainsi en posant

$$Y_{il}^j = \begin{cases} 1 & \text{si l'observation } (i, l) \text{ appartient à la classe } j, j \in J \\ 0 & \text{sinon} \end{cases} \quad |J| = m$$

où les  $m$  classes sont déterminées par un plan d'expérience à deux facteurs (correspondant aux indices  $i$  et  $j$ ), la régression isotone de  $Z$  en  $X$  permet alors un ajustement ordinal du modèle d'analyse de variance étudié par KRUSKAL (13), de LEEUW (6), YOUNG et de LEEUW (21) :

$$Z_{il} = \langle Y_{il}, w \rangle = a_i + b_l + \xi_{il}$$

où  $\xi_{il}$  est un terme d'erreur.

#### IV-1 L'analyse multiple du triple

Le problème abordé par ce type d'analyse, plus connu des psychométriciens sous le nom d'analyse des différences individuelles, est l'étude du triple.

$$(I, \{\delta_I^j ; j \in J\}, \{p_I^j ; j \in J\})$$

où l'on dispose de  $m$  replications ( $|J| = m$ ) de la mesure de la dissimilarité  $\delta_I$  entre éléments de  $I$ . Nous nous proposons de représenter  $I$  dans un espace vectoriel normé  $V$  de dimension  $r$ , et de représenter  $J$  dans le dual  $V^*$  de  $V$  :

$$X : I \longrightarrow V$$

$$Y : J \longrightarrow V^*$$

$$i \rightsquigarrow X(i) = (x_k(i))_{k=1..r}, j \rightsquigarrow Y(j) = (y_k(j))_{k=1..r}$$

de telle sorte que

- i) la dimension  $r$  de  $V$  soit minimale
- ii) l'ajustement entre les données  $\Delta = \{\Delta^j ; j \in J\}$  et les distances interpoints reconstruites  $D$  dans l'espace commun  $V$ , soit optimale (au sens des moindres carrés).

$$D = (d_{i1})_{i,1 \in I} \quad d_{i1} = \|X(i) - X(1)\|_V$$

- iii)  $Y(j)$  traduit, dans  $V^*$ , la contribution (au sens le Loading) de la donnée  $\Delta^j$  dans l'espace commun  $V$ .

Lorsque  $r$  est connu, la détermination de  $X$  est un problème de régression multiple de  $D$  en fonction des  $T^j = f(\Delta^j)$  où l'expression de  $f$  dépend du niveau d'échelle des mesures  $\Delta^j$ . On est donc conduit à poser la

Définition 5-

On appelle indice de corrélation ordinaire multiple

$$\sqrt[2]{Y/\{X_j, j \in J\}} = \rho^2(Y, E(Y/\mathcal{P}))$$

où  $\mathcal{P}$  est le sous  $\sigma$ -treillis de  $\mathcal{Q}$  engendré par les variables ordinales  $\{X_j, j \in J\}$ , c'est-à-dire le plus petit sous  $\sigma$ -treillis de  $\mathcal{Q}$  contenant les  $\sigma$ -treillis engendrés par les  $X_j$

de plus si nous notons  $\sqrt[2]{Y/\{X_j, j \in J\}}$  l'indice empirique de corrélation ordinaire multiple, défini comme dans le cas univarié par

$$\sqrt[2]{Y/\{X_j, j \in J\}} = 1 - \frac{\|Y - T\|_{\mathbb{P}}^2}{\|Y - y.J\|_{\mathbb{P}}^2}$$

Le résultat de la proposition 2 s'étend alors de façon immédiate au cas présent

Le problème réside en fait en la construction de l'espérance conditionnelle  $T$ . nous avons proposé (8) d'utiliser, dans le cadre ordinal, des techniques d'analyse multi-critère, afin de construire une relation d'ordre médian, traduisant l'information portée par le  $\sigma$ -treillis, suivant divers critères d'agrégation.

Mais l'application de ces méthodes à l'analyse du triple devient très vite fort coûteuse, aussi les approches de YOUNG et de LEEUW (21) et de LEEUW (7) nous semble t-elles plus prometteuses.

La construction de  $Y$  utilise le modèle linéaire ordinal. En effet si l'on associe à tout  $j$  dans  $J$  une métrique euclidienne de matrice  $W(j)$ , a pour tout  $b$  dans  $B$  ( $b = (i, 1)$ )

$$(2) \left[ \alpha_b^j \right]^2 = \|X(i) - X(1)\|_{W(j)}^2$$

Alors

$$d_b^2 = \sum_{j \in J} p_j [d_b^j]^2 = \|X(i) - X(1)\|_W^2$$

définit une distance euclidienne sur  $V$  associée à la forme quadratique

$$W = \sum_{j \in J} p_j W(j) \text{ où l'on suppose donnée une pondération } P_j \text{ sur } J$$

De la même façon que dans le modèle INDSCAL de CARROLL (20), nous interprétons  $W$  comme la métrique sur l'espace commun  $V$ , induisant la distance  $D$  et nous imposons aux  $W(j)$  d'être diagonales. Si nous notons  $Y(j)$  le vecteur des termes diagonaux de  $W(j)$  et

$$z_k(b) = (x_k(i) - x_k(1))^2, \quad Z(b) = (z_k(b))_{k=1 \dots r}, \quad b = (i, 1)$$

L'expression (2) s'écrit :

$$[d_b^j]^2 = \langle Z(b), Y(j) \rangle$$

Ainsi en maximisant  $\hat{\nu}_{D/\Delta^j}^2$  construit-on  $Y(j)$  pour tout  $j$  dans  $J$ . De plus le tableau de décomposition de la variance associée aux  $\hat{\nu}_{D/\Delta^j}^2$  et à  $\hat{\nu}_{D/\Delta^j, j \in J}^2$  permet-il un meilleur jugement de la qualité

des représentations construites. La proposition 2 montre qu'une telle méthode généralise le modèle INDSCAL, dont il partage les faiblesses, liées à la contrainte non testée de diagonalité des  $W(j)$ .

#### IV-2 L'analyse en composantes principales ordinale

De nombreuses méthodes d'analyse de données de dominance ont été proposées et leur expression en termes de regression isotone est donnée dans (8). Nous nous limitons ici à l'approche de SHEPARD et KRUSKAL (18).

Les données de dominance se présentent sous l'une des deux formes suivantes :

- i) un tableau  $\Delta$  résume un ensemble  $J$  de classements des éléments d'un ensemble fini  $I$ .
- ii) une famille de tableaux de comparaisons par paires des éléments de  $I$

$$\{ \Delta^j, j \in J \} \text{ nous est donnée .}$$

Nous supposerons dans la suite la transitivité des comparaisons par paires de sorte que l'équivalence de i) et ii) est assurée.

On se propose de construire, dans un espace euclidien  $V$  de dimension  $r$  minimale, une représentation  $X$  de  $I$ . L'ensemble  $J$  est décrit dans le dual  $V^*$  de  $V$  de telle sorte que les distances à l'origine des projections des points représentatifs  $X(i)$  sur la droite  $D_j$ , associée dans  $V$  à la forme linéaire  $Y(j)$  respectent au mieux les classements donnés.

Ainsi, si l'on impose à  $Y(j)$  d'être normé à 1, on demande au produit scalaire

$$Z_{ij} = \langle X(i), Y(j) \rangle$$

d'être une fonction isotone des données  $\sigma_{ij}$ .

On impose de plus les contraintes usuelles en analyse en composantes principales i) de centrage des  $X(i)$

i) de centrages des  $X(i)$

$$t_{X J_n} = 0 \quad .$$

ii) de sphéricité du nuage des  $X(i)$

$$t_{XX} = nI$$

qui assurent l'unicité de la représentation.

Le modèle linéaire ordinal permet de construire  $X$  et  $Y$ , en posant

$$z_j = t_{XY(j)} \quad .$$

La régression isotone de  $z_j$  en  $\Delta^j$  maximise l'indice  $\psi^2_{D/T_j}$  et les auteurs proposent de juger la qualité de la représentation par l'indice

$$\psi^2 = \sum_{j \in J} p_j \left[ 1 - \psi^2_{D/T_j} \right]^2$$

où  $T_j$  est une fonction isotone de  $\Delta^j$ . Un tel critère nous semble moins adapté au problème que l'indice empirique de corrélation ordinale, ce qui pourrait expliquer la sensibilité de cette méthode aux phénomènes de dégénérescence signalés par les auteurs (18).

#### Remarques

- L'utilisation de la proposition 2 dans ce cadre ne présente qu'un intérêt théorique, compte tenu du coût accru de cette méthode comparativement à une analyse en composantes principales linéaire. La robustesse de celle-ci aux non-linéarités vient conforter ce jugement.
- L'introduction d'une pondération  $P_I$  sur l'ensemble des objets classés bien que non étudiée par les auteurs, est immédiate et présente un grand intérêt pratique en sciences sociales.

V - REFERENCES

- (1) BLALOCK, H.M. (1974) - Measurement in the social sciences - theories and strategies - Aldine publ. comp. - CHICAGO .
- (2) BARLOW, R.E and Alt (1972) - Statistical inference under order restrictions - Wiley and sons - New York .
- (3) CAILLEZ, F and Alt. (1976) - Introduction à l'analyse des données - S.M.A.S.H. - 9, rue Duban - 75016 PARIS .
- (4) COOMBS, C.H (1964) - A theory of data - Wiley - New York .
- (5) COOPER, L.G. (1972) - A new solution to the additive constant problem in metric multidimensional scaling - PSYCHOMETRIKA - Vol 37 n° 3 pp 311-322 .
- (6) de LEEUW, J (1968) - Non metric linear model - Mimeographed RN 00369 D<sup>pt</sup> of data theory - Leiden - the Netherlands .
- (7) de LEEUW, J (1976) - Application of convex analysis to multidimensional scaling - Invited paper - European meeting of statisticians - Grenoble .
- (8) DROUET D'AUBIGNY, G. (1975) - Description statistique des données ordinales : Analyse multidimensionnelle thèse de 3° cycle - GRENOBLE .
- (9) DROUET D'AUBIGNY, G. (1976) - Description multidimensionnelle des données ordinales - SEMINAIRE I.R.I.A. en classification automatique (à paraître) .
- (10) GOWER, J.C. (1966) - Some distance properties of latent root and vector methods used in multivariate analysis - BIOMETRIKA n° 53 pp 325-338 .

- (11) GUTTMAN, L. (1968) - A general non metric technique for finding the smallest coordinate space for a configuration of points - PSYCHOMETRIKA - Vol 33 n° 4 pp469-506 .
- (12) GUTTMAN, L. et  
LINGOES J.C. (1967) - Non metric factor analysis a rank reducing alternative to linear factor analysis - MULTIVARIATE BEHAVIORAL RESEARCH - october 1967 pp 485-505 .
- (13) KRUSKAL, J.B. (1964) - Multidimensional scaling for optimising goodness of fit to a non-metric hypothesis - PSYCHOMETRIKA vol 29 pp 1-27 et vol 29 pp 115-129 .
- (14) MAYER, L.S (1973) - Estimating a correlation coefficient when one variable is not directly observed - J.A.S.A. n° '68 pp 420-421 .
- (15) OBENCHAIN, R.N. (1971) - Multivariate procedures invariant under linear transformations - ANNALS OF MATHEMATICAL STATISTICS - Vol 42 n° 5 pp 1569-1578 .
- (16) OBENCHAIN, R.N. (1973) - Affine multidimensional scaling - Mimeographed BELL LABORATORIES - MURPHY HILL - New Jersey USA .
- (17) SAMMON, J.W. (1969) - A non linear mapping for data structure analysis I.E.E.E. TRANS ON COMPUTERS Vol. C 18 n° 5 MAY pp 401-409 .
- (18) SHEPARD, R.N et  
KRUSKAL J.B. (1974) - A non-metric variety of linear factor analysis PSYCHOMETRIKA Vol 39 n° 2 pp 123-156 .

- (19) TORGERSON, W.S (1958) - Theory and methods of scaling - Wiley - New York
- (20) YOUNG, F.W. (1972) - A model for polynomial conjoint analysis algorithms  
dans - Multidimensional scaling - SHEPARD - ROMNEY  
NERLOVE edi. SEMINAR PRESS - New York .
- (21) YOUNG, F.W., de LEEUW, J (1975) - Non-metric individual differences  
multidimensional scaling - PSYCHOMETRIC  
LABORATORY REPORT n° 146  
University of north Carolina .

G. DROUET D'AUBIGNY

Département Statistiques

Bt Sciences Humaines et Mathématiques

Université des Sciences Sociales de Grenoble

47X - 38040 GRENOBLE-CEDEX