

Les présentations factorielles de la classification

0 - INTRODUCTION GENERALE

Différentes tentatives de rapprochement entre l'Analyse Factorielle et la Classification ont été proposées : N. Howard (1969), J.P. Benzecri (1971), M. Gondran (1975), M. Jambu suivant J.P. Benzecri (1976). Toutes ces tentatives correspondent en fait, comme nous le verrons ci-dessous à des présentations "factorielles" du problème de la Classification. La tentative de telles présentations est en effet grande compte tenu du caractère plus établi de l'histoire de l'Analyse Factorielle en Composantes qui fournit la solution optimale pour le critère de l'inertie expliquée.

Notre but dans ce chapitre est d'analyser chacune de ces tentatives pour préciser sa véritable nature, la généraliser et l'adapter à des situations nouvelles. Ce faisant, nous contribuerons à chacune des approches en les situant, dans un effort de synthèse, les unes par rapport aux autres Ce qui nous permettra de nous rendre compte de l'intérêt relatif de ces différentes tentatives.

Les différents types de rapprochement entre l'Analyse Factorielle et la Classification se distinguent en ce que certains traitent du problème de la recherche d'une classification et que d'autres traitent de celui de la recherche d'un arbre binaire des classifications. Ils se distinguent également de par la nature du critère optimisé : deux critères seront considérés ici ; le premier est basé sur l'inertie expliquée (i.e. variance) et le second sur la notion de proximité entre parties disjointes au sens topologique du terme (i.e. saut minimal). Ces différentes approches se distinguent encore par la forme de l'équation factorielle retenue : s'agit-il de rechercher un système d'axes factoriels dans l'espace de représentation du nuage ou bien, une suite de facteurs dont chacun se trouve défini comme une fonction sur l'ensemble des sommets auquel il est relatif. Ces approches se distinguent enfin de par la nature de la structure projective de l'espace engendré par la solution de l'équation factorielle retenue.

Nous avons organisé le texte en deux parties A et B selon le type de structure interprétée ; A pour une classification et B pour une hiérarchie de classifications. Toutefois, on peut considérer une autre organisation selon le critère en jeu et la forme de l'équation factorielle qui sert à l'interprétation. Dans ce cas, la lecture du chapitre , sans doute plus

aisée, se fera selon le schéma suivant :

$$A11, A12 \rightarrow B1 ; A2 \rightarrow B2.$$

A. PRESENTATION FACTORIELLE DE LA RECHERCHE D'UNE CLASSIFICATION

A.1. Critère de l'inertie expliquée

A.1.1. Recherche d'un système d'axes factoriels

L'étude présentée dans ce paragraphe qui correspond à un travail effectué en 1973 (publié en 1979 (cf. [11]) a pour origine un article de N. Howard ("Least squares Classification and Principal Component Analysis : a comparison" (cf. [5])) où on montre que la classification des "moindres carrés" de l'ensemble indexant l'un des côtés d'un tableau de mesures numériques (Objets ou Individus x Variables) peut d'une certaine façon apparaître comme une analyse en composantes principales avec "contraintes" de l'ensemble indexant l'autre côté du tableau des données. Nous reprenons ici ce travail de façon plus générale, en nous situant au niveau de l'analyse d'un nuage de points dans un espace euclidien ; de façon plus précise, en spécifiant l'influence des parts "factorielle" et "classificatoire" dans l'interprétation des résultats d'une analyse. D'autre part et surtout, nous adaptons cette approche dans le cas de l'analyse d'un tableau de contingence ; cas où elle prend un sens optimal compte tenu de la symétrie des rôles entre lignes et colonnes.

I - INTRODUCTION ET RAPPELS

La méthode de classification conceptuellement la plus proche de l'analyse factorielle d'un nuage

$$\mathcal{N}(I) = \{(M_i, \mu_i) / i \in I\}, \quad (0)$$

où μ_i désigne la masse affectée au sommet M_i , situé dans un espace euclidien E muni d'une métrique q , repose sur la formule suivante de décomposition du moment total d'inertie :

$$\sum_{i \in I} \mu_i \|M_i - G\|^2 = \sum_{1 \leq h \leq k} \left\{ \sum_{i \in I_h} \mu_i \|M_i - G_h\|^2 \right\} + \sum_{1 \leq h \leq k} v_h \|G_h - G\|^2, \quad (1)$$

Dans cette formule, dite d'analyse de la variance, de même type que celles attribuées à Huygens ; $\{I_h / 1 \leq h \leq k\}$ désigne une partition en k classes de I .

A chacune des classes I_h se trouvent associées, d'une part sa masse

$$v_h = \sum_{i \in I_h} \mu_i, \quad (2)$$

et d'autre part, son centre de gravité

$$G_h = \frac{1}{v_h} \sum_{i \in I_h} \mu_i M_i, \quad (3)$$

On a, bien entendu,

$$\mu = \sum_{1 \leq h \leq k} v_h \text{ et } G = \frac{1}{\mu} \sum_{1 \leq h \leq k} v_h G_h, \quad (4)$$

Le moment d'inertie de la classe I_h

$$\mathcal{M}_h = \sum_{i \in I_h} \mu_i \|M_i - G_h\|^2 \quad (5)$$

peut refléter la *cohésion* de la h-ème classe qui sera d'autant plus grande que \mathcal{M}_h est petit.

D'autre part,

$$\mathcal{L}_h = \nu_h \|G_h - G\|^2, \quad (6)$$

qui est le moment d'inertie, par rapport au centre de gravité G du nuage, du centre de gravité G_h de la h-ème classe muni de la masse ν_h de cette dernière, peut refléter combien cette h-ème classe se distingue dans l'ensemble du nuage $\mathcal{N}(I)$. On peut appeler \mathcal{L}_h *distinction* de la h-ème classe. On a

$$\mathcal{M} = \sum_{1 \leq h \leq k} (\mathcal{M}_h + \mathcal{L}_h), \quad (7)$$

Le problème de la classification des moindres carrés, consiste, pour k fixé, à déterminer celles des partitions de I en k classes non vides, qui minimisent $\sum \{\mathcal{M}_h / 1 \leq h \leq k\}$; c'est-à-dire, en vertu de la relation (7), qui rendent maximum $\sum \{\mathcal{L}_h / 1 \leq h \leq k\}$. On cherche aussi une classification dont, globalement (au sens additif du terme), la cohésion des classes est la plus forte.

La solution optimale à ce problème peut ne pas être unique; d'ailleurs il n'existe pas d'algorithme qui conduise à une solution optimale sans envisager presque toutes les partitions en k classes. Toutefois, dans les cas réels où la population étudiée est apte à être classifiée (cf. [10] chapitre 4 § III), les solutions "proches" de l'optimum sont très "voisines" entre elles et on peut accéder à l'une d'entre elles par un algorithme qui ne peut prétendre qu'optimiser localement le critère. Différents algorithmes ont été proposés et étudiés; mais il n'est pas dans notre propos d'en faire allusion ici, où on suppose le problème résolu. C'est donc relativement à une classification optimale, ou du moins "proche" de l'optimale, que nous comparerons l'analyse factorielle à la classification des moindres carrés.

La condensation du nuage $\mathcal{N}(I)$ définie par une telle classification peut être regardée comme résultant de la projection de chacun des points M_i sur le centre de gravité G_h de la classe où il tombe (G_h est considéré ici comme une variété affine de dimension 0). Dans cette condensation $\sum \{\mathcal{L}_h / 1 \leq h \leq k\}$ est la part d'inertie (i.e; variance) retenue ou expliquée et $\sum \{\mathcal{M}_h / 1 \leq h \leq k\}$ est la part perdue.

II - COMPARAISON DE L'INERTIE EXPLIQUÉE PAR UNE CLASSIFICATION ET CELLE PAR UNE ANALYSE FACTORIELLE.

PROPRIÉTÉ. Soit $\{a_r / 1 \leq r \leq p\}$ la suite des p premiers axes factoriels unitaires qui correspondent aux p plus grandes valeurs propres de l'analyse du nuage $\mathcal{N}(I)$ et soit $\{I_h / 1 \leq h \leq k\}$ une partition en k classes non vides de I .

Comparer l'inertie du nuage projeté sur l'espace des p premiers axes factoriels à la part d'inertie retenue par la classification, revient à comparer

$$F = \sum_{1 \leq r \leq p} \left\{ \sum_{1 \leq h \leq k} \left\{ \sum_{i \in I_h} \mu_i [q(a_r, M_i - G_h)]^2 \right\} \right\}, \quad (8)$$

et

$$C = \sum_{r > p} \sum_{1 \leq h \leq k} v_h [q(a_r, G_h - G)]^2, \quad (9)$$

En effet, compte tenu de la décomposition

$$(\forall h=1, \dots, k), (\forall i \in I_h), M_i - G = (M_i - G_h) + (G_h - G);$$

on a, en raison des propriétés barycentriques, la relation

$$\lambda(r) = \sum_{i \in I} \mu_i [q(a_r, M_i - G)]^2 = \sum_{1 \leq h \leq k} \sum_{i \in I_h} \mu_i [q(a_r, M_i - G_h)]^2 + \sum_{1 \leq h \leq k} v_h [q(a_r, G_h - G)]^2, \quad (10)$$

où $\lambda(r)$ désigne l'inertie expliquée par le r -ème axe factoriel. Il s'agit de comparer $\sum\{\lambda(r)/1 \leq r \leq p\}$ à $\sum\{v_h/1 \leq h \leq k\}$ qui peut se mettre sous la forme

$$\sum_{1 \leq r \leq m} \left\{ \sum_{1 \leq h \leq k} v_h [q(a_r, G_h - G)]^2 \right\}, \quad (11)$$

où m désigne la dimension du sous espace de E engendré par l'ensemble des vecteurs $\{(M_i - G)/i \in I\}$; $m > p$.

En se référant à l'expression (10) de $\lambda(r)$, on voit que les deux quantités à comparer ont la partie commune

$$D = \sum_{1 \leq r \leq p} \left\{ \sum_{1 \leq h \leq k} v_h [q(a_r, G_h - G)]^2 \right\}, \quad (12)$$

Il reste donc à comparer F et G définis ci-dessus.

Dans la part d'inertie perdue par la classification, F désigne la fraction due aux p premiers axes factoriels; alors que C indique la portion de l'inertie expliquée par la classification, due aux $(m-p)$ derniers axes factoriels. La quantité D (cf. (12)) définit quant à elle l'inertie expliquée par la classification dans l'espace des p premiers axes factoriels.

La quantité F est d'autant plus petite (resp. C d'autant plus grande) que d'une part, l'aptitude du nuage $\mathcal{N}(I)$ à être classifié est grande dans l'espace des p premiers axes factoriels (resp. dans l'espace des $(m-p)$ derniers axes factoriels) et que d'autre part, l'inertie expliquée par les p premiers axes factoriels unitaires, est petite.

L'inégalité $F < C$ exprime que la variance expliquée par la classification est plus grande que celle du nuage projeté sur l'espace des p premiers axes factoriels. On peut d'ailleurs aisément, au moyen d'un programme, associer aux résultats d'une analyse factorielle et d'une classification des moindres carrés, portant sur un même nuage $\mathcal{N}(I)$ dans un espace euclidien, les deux quantités F et C pour les comparer.

La recherche d'un degré suffisant de précision dans les résultats d'une analyse factorielle, nous entraînerait à prendre une valeur grande de p ; cependant l'interprétation difficile des facteurs (surtout lorsque leur

nombre dépasse 3 ou 4) liée à la nécessité d'une représentation simple des résultats, nous incline à choisir p petit.

D'autre part, dans une optique de classification, le nombre k de classes n'est pas, pour de nombreuses méthodes, fixé arbitrairement ; il est un résultat de l'algorithme. En général k est sensiblement plus grand que le nombre de facteurs interprétables.

La signification d'une même classe d'objets est en général beaucoup plus facile à dégager que celle d'un facteur ; en effet, il est plus aisé d'ajuster un concept à partir d'un profil reflétant un ensemble d'objets "proches" vis à vis d'un ensemble de variables que d'interpréter un facteur qui est une sorte de moyenne pondérée des différentes variables. Ce profil peut par exemple être défini par la suite des moyennes des différentes variables sur l'ensemble d'objets que détermine la classe.

Dans la mesure où la quantité F est sensiblement plus petite que C ; la classification donnera une vision de synthèse plus claire que celle que peut donner l'analyse factorielle. Si par contre, l'inertie expliquée par les tous premiers axes factoriels unitaires est appréciable, on aura une plus grande précision dans l'analyse du nuage projeté. Les deux approches peuvent rivaliser, mais souvent se complètent. F pourra définir la contribution "factorielle" de l'interprétation et C, la contribution "classificatoire". Pour avoir une idée tant soit peu précise des comportements de F et de C, de nombreuses expériences réelles et de simulation sur ordinateur sont nécessaires.

III - CLASSIFICATION DE L'ENSEMBLE DES VARIABLES ET ANALYSE FACTORIELLE CONTRAINTE DE $\mathcal{N}(I)$.

Le nuage $\mathcal{N}(I)$ est ici associé à la description d'un ensemble fini I d'individus ou objets au moyen d'un ensemble fini V de variables descriptives numériques. $J=\{1,2,\dots,m\}$ indexera V et le support de la description est le tableau de mesures (x_{ij}) indexé par $I \times J$.

En associant à la variable $v_j (1 \leq j \leq m)$, la j-ème forme linéaire coordonnée e_j^* de R^m ; on représentera la ligne i du tableau par le point M_i de R^m défini par

$$(M_i - 0) = \sum_{1 \leq j \leq m} x_{ij} e_j, \text{ pour tout } i \in I, \quad (13)$$

où $\{e_j / 1 \leq j \leq m\}$ est la base canonique de R^m .

On affectera au sommet M_i , la masse μ_i définissant d'une certaine façon l'"importance" du i-ème individu. De la sorte, le centre de gravité du nuage est défini par

$$(G-0) = \sum_{1 \leq j \leq m} g_j e_j, \quad (14)$$

où

$$g_j = \frac{1}{\mu} \sum_{1 \leq i \leq n} \mu_i x_{ij}, \quad (15)$$

Enfin, l'espace R^m est supposé muni d'une métrique q, diagonalisé

par rapport à la base canonique, permettant d'évaluer de façon "adéquate" les distances entre individus à partir de leur représentation.

Considérons à présent la représentation de V par un nuage $\mathcal{N}(J)$ de m points de R^m où la suite des coordonnées du j -ème point N_j (représentant la variable v_j) est définie par

$$([x_{ij} - g_j] \sqrt{q_{jj}} / 1 \leq i \leq n), \quad (16)$$

on affectera à N_j la masse unité et on munira R^m de la métrique diagonale suivante

$$q^*(f_i, f_{i'}) = \begin{cases} 0 & \text{si } i' \neq i \\ \mu_i & \text{si } i' = i \end{cases}, \quad (17)$$

où $\{f_i / 1 \leq i \leq n\}$ est la base canonique de R^n .

Avec une telle représentation, le carré de la distance du sommet N_j à l'origine, représente l'inertie de la projection orthogonale du nuage $\mathcal{N}(I)$ sur le j -ème axe canonique porté par e_j ; soit

$$\sum_{i \in I} \mu_i (x_{ij} - g_j)^2 q_{jj}, \quad (18)$$

Dans ces conditions, le problème de la classification des moindres carrés en p classes non vides de V , consiste à déterminer la ou les partition $\{J_r / 1 \leq r \leq p\}$ de J , où, pour tout r , $J_r \neq \emptyset$, de façon à maximiser

$$\sum_{1 \leq r \leq p} m(r) \|H_r - O\|^2, \quad (19)$$

où $m(r)$ est le cardinal de la classe J_r et où H_r est le centre de gravité de la r -ème classe :

$$H_r = \frac{1}{m(r)} \sum_{j \in J_r} N_j, \quad (20)$$

L'expression analytique de (19) est la suivante :

$$\sum_{1 \leq r \leq p} m(r) \sum_{1 \leq i \leq n} \mu_i \left[\frac{1}{m(r)} \sum_{j \in J_r} \xi_{ij} \right]^2, \quad (21)$$

où ξ_{ij} est la mesure centrée réduite $(x_{ij} - g_j) \sqrt{q_{jj}}$.

Nous allons montrer qu'une telle classification, du nuage $\mathcal{N}(J)$ ainsi défini, peut apparaître comme une analyse factorielle sous contraintes de $\mathcal{N}(I)$. Ces contraintes correspondent à imposer une forme donnée aux vecteurs b_r de la base $\{b_r / 1 \leq r \leq p\}$ q -orthonormale, formée de vecteurs propres, qui engendre l'espace factoriel.

L'inertie du nuage $\mathcal{N}(I)$ projeté sur un tel espace est défini par

$$\sum_{1 \leq r \leq p} \sum_{i \in I} \mu_i [q(b_r, M_i - G)]^2 = \sum_{1 \leq r \leq p} \sum_{i \in I} \mu_i \left[\sum_{1 \leq j \leq m} \beta_{rj} (x_{ij} - g_j) q_{jj} \right]^2, \quad (22)$$

où on a posé

$$b_r = \sum_{1 \leq j \leq m} \beta_{rj} e_j, \quad (23)$$

PROPOSITION. La classification des moindres carrés du nuage $\mathcal{N}(J)$, définie ci-dessus, est une analyse factorielle de $\mathcal{N}(I)$ où les axes factoriels $b_r (r=1, \dots, p)$ sont assujettis à la condition suivante :

$$(\forall r=1, \dots, p), (\exists J_r \neq \emptyset); \beta_{rj} = \begin{cases} 0 & \text{si } j \notin J_r \\ \gamma(r)/\sqrt{q_{jj}} & \text{si } j \in J_r \end{cases}, \quad (24)$$

où $\gamma(r) > 0$.

La condition précédente exprime que pour tout r , il existe au moins un indice j , pour lequel $\beta_{rj} \neq 0$ et que, pour un même r , les seules composantes non nulles de b_r , sont respectivement proportionnelles aux nombres $1/\sqrt{q_{jj}}$, $j=1, 2, \dots, m$; le coefficient de proportionnalité $\gamma(r)$ étant positif.

Le caractère orthogonal de la base $\{b_r / 1 \leq r \leq p\}$ impose nécessairement que l'ensemble des parties $\{J_r / 1 \leq r \leq p\}$ forme une partition de J en p classes non vides.

Le caractère q -orthogonal de la base permet de préciser le coefficient de proportionnalité $\gamma(r)$, pour tout r . En effet, en écrivant $q(b_r, b_r) = 1$ on obtient

$$\sum_{j \in J_r} \beta_{rj}^2 q_{jj} = 1, \quad (25)$$

qui donne

$$\gamma(r) = 1/\sqrt{m(r)}, \quad (26)$$

où $m(r) = \text{card}(J_r)$.

Quant au critère optimisé

$$\sum_{1 \leq r \leq p} \sum_{1 \leq i \leq n} [\mu_i q(b_r, M_i - G)]^2, \quad (27)$$

qui représente l'inertie du nuage $\mathcal{N}(I)$ projeté q -orthogonalement sur l'espace engendré par $\{b_r / 1 \leq r \leq p\}$; il prend la forme suivante :

$$\sum_{1 \leq r \leq p} \sum_{1 \leq i \leq n} \mu_i \left[\frac{1}{\sqrt{m(r)}} \sum_{j \in J_r} (x_{ik} - g_j) \sqrt{q_{jj}} \right]^2, \quad (28)$$

qui se ramène bien, immédiatement, à l'expression (21).

C.Q.F.D.

Dans le cas de l'Analyse en Composantes Principales Normée ; la métrique diagonale q est définie par

$$(1/q_{jj}) = s_j^2 = \sum_{i \in I} \mu_i (x_{ij} - g_j)^2, \quad (29)$$

qui représente la variance de la distribution de la variable v_j sur I . Le plus souvent, on a $\mu_i = 1/n$ pour tout i . Dans ces conditions, la suite des composantes du point N_j est la suite des mesures

$$\xi_{ij} = (x_{ij} - \bar{x})/s_j, \quad i=1, \dots, n; \quad (30)$$

centrées réduites de la j -ème variable v_j sur I . Les différents sommets

N_j se trouvent alors à égale distance 1 de l'origine.

Reprenons dans le cas général, la formule suivante de décomposition de l'inertie totale du nuage (I)

$$\sum_{i \in I} \mu_i \sum_{j \in J} \xi_{ij}^2 = \sum_{i \in I} \mu_i \sum_{1 \leq r \leq p, j \in J_r} (\xi_{ij} - \bar{\xi}_i(r))^2 + \sum_{i \in I} \mu_i \sum_{1 \leq r \leq p} m(r) [\bar{\xi}_i(r)]^2, \quad (31)$$

où on a noté

$$\bar{\xi}_i(r) = \frac{1}{m(r)} \sum_{j \in J_r} \xi_{ij}$$

On voit avec cette formule que la classification en p classes recherchée doit être telle que la part d'inertie conservée soit la plus grande lorsqu'on remplace sur chacun des individus i , la suite des n mesures ξ_{ij} par seulement une suite de p mesures, dont chacune $\bar{\xi}_i(r)$, est la moyenne des mesures d'une même classe de variables. Cette moyenne étant équilibrée ; on se rend compte que la classification fournit une approche plus uniforme des diverses variables mises en jeu. Le caractère uniforme de cette approche, qui simplifie considérablement l'interprétation, ne va pas sans une perte de l'inertie retenue par rapport à une analyse factorielle sans contraintes. Mais, il ne faut pas oublier qu'on peut aisément interpréter sensiblement plus de classes que de facteurs ; par ailleurs, des expériences ont pu montrer, qu'au delà d'un certain seuil, un accroissement de la variance retenue peut ne pas s'accompagner d'une interprétation plus nuancée.

IV - ANALYSE FACTORIELLE CONTRAINTE DE J ET CLASSIFICATION DE L'ENSEMBLE DES OBJETS.

Les notations sont celles du paragraphe précédent. En désignant par $\{c_r / 1 \leq r \leq p\}$ une base q -orthogonale, le critère à maximiser dans une analyse factorielle de $\mathcal{N}(J)$, se met sous la forme

$$\sum_{1 \leq r \leq p, j \in J} [q^*(c_r, N_j - 0)]^2, \quad (32)$$

où q^* a été précisé en (17).

L'expression (32) se met analytiquement, sous la forme suivante :

$$\sum_{1 \leq r \leq p, j \in J} \left[\sum_{i \in I} \mu_i \gamma'_{ri} (x_{ij} - g_j) \sqrt{q_{jj}} \right]^2, \quad (33)$$

où on a noté

$$c_r = \sum_{1 \leq i \leq n} \gamma'_{ri} f_i, \quad (34)$$

$\{f_i / 1 \leq i \leq n\}$ désignant, bien entendu, la base canonique de R^n .

PROPOSITION. La classification des moindres carrés du nuage $\mathcal{N}(I)$, est une analyse factorielle de $\mathcal{N}(J)$ où les axes factoriels c_r ($r=1, \dots, p$) sont assujettis à la condition suivante :

$$(\forall r=1, \dots, p), (\exists I_r \neq \emptyset) ; \gamma'_{ri} = \begin{cases} 0 & \text{si } i \notin I_r, \\ \delta(r) & \text{si } i \in I_r \end{cases}, \quad (35)$$

où $\delta(r) > 0$.

Comme dans le cas dual précédent (cf. §3 ci-dessus), la condition exprime que pour tout r , il existe au moins un indice i , pour lequel $\gamma'_{ri} \neq 0$ et que, pour un même r , les seules composantes non nulles de c_r , sont égales à un même nombre positif.

Le caractère orthogonal de la base $\{c_r / 1 \leq r \leq p\}$ implique que l'ensemble des parties $\{I_r / 1 \leq r \leq p\}$ définit une partition de I en p classes non vides ; la r -ème étant indiquée par c_r .

D'autre part, le caractère q^* -orthonormal de la base, va nous permettre de préciser la valeur du coefficient positif $\delta(r)$.

$$q^*(c_r, c_r) = \sum_{i \in I_r} (\delta(r))^2 \mu_i = 1, \quad (36)$$

d'où

$$\delta(r) = 1/\sqrt{v(r)} \text{ avec } v(r) = \sum_{i \in I_r} \mu_i, \quad (37)$$

Dans ces conditions, l'expression (33) du critère maximisé devient

$$\sum_{1 \leq r \leq p} v(r) \sum_{1 \leq j \leq m} \left[\frac{1}{v(r)} \sum_{i \in I_r} \mu_i (x_{ij} - g_j) \right]^2 q_{jj}, \quad (38)$$

laquelle, pouvant manifestement se mettre sous la forme

$$\sum_{1 \leq r \leq p} v(r) \|G_r - G\|^2, \quad (39)$$

avec les notations du paragraphe 2 précédent ; G_r est le centre de gravité de la classe I_r et G , le centre de gravité du nuage $\mathcal{N}(I)$, total.C.Q.F.D.

V - CAS SYMETRIQUE D'UN TABLEAU DE CONTINGENCE.

On a pu se rendre compte d'une certaine dissymétrie dans la représentation et l'analyse entre le traitement des lignes, représentant l'ensemble des objets ou individus et le traitement des colonnes, représentant l'ensemble des variables descriptives numériques (cf. §3 et §4). Cette dissymétrie est profondément due à la différence de nature entre ce que représente une ligne et ce que représente une colonne. Il y a en effet autant de différence entre une variable et un objet qu'entre un appareil de mesure et l'individu sur lequel se trouve effectuée la mesure. Dans les enquêtes ou études expérimentales, l'ensemble I des individus ou objets est défini à partir d'un échantillon choisi aléatoirement qu'on espère "représentatif" de la population étudiée ; tandis que pour le choix de l'ensemble V des variables de description, chacun de ses éléments est déterminé avec minutie par le spécialiste des données en question et derrière une même variable, il y a toute la connaissance et l'appréhension d'un domaine scientifique donné.

La dissymétrie entre le traitement des lignes et des colonnes va disparaître dans le cas de l'analyse métrique d'un tableau de contingence : analyse des correspondances. Il s'agira dans notre cas de montrer qu'une telle analyse factorielle, avec des contraintes à préciser, sur l'un des côtés du tableau $I \times J$ peut apparaître comme classification des moindres carrés, au sens de la métrique du χ^2 , sur l'autre côté du tableau.

Commençons par rappeler la représentation géométrique du nuage $\mathcal{N}(I)$ (resp. $\mathcal{N}(J)$) associé à l'ensemble des lignes (resp. colonnes) du tableau de contingence $I \times J$; ce qui nous permettra de préciser nos notations pour la suite.

La table de contingence représente le croisement de deux variables définissant chacune une partition sur la population étudiée. I (resp. J) désigne ici l'ensemble des modalités ou valeurs de l'une des variables (resp. de l'autre) ; posons $n = \text{card}(I)$ et $m = \text{card}(J)$. A l'intersection de la ligne i et de la colonne j du tableau de contingence indexé par $I \times J$, est le nombre k_{ij} de sujets (d'un échantillon de la population étudiée, en général) qui possèdent la i -ème modalité du premier caractère et la j -ème du second. On substitue à ce tableau de nombres entiers, celui des fréquences relatives $f_{ij} = (k_{ij}/k_{..})$ où $k_{..}$ est l'effectif de l'échantillon concerné : $k_{..} = \sum\{k_{ij}/(i,j) \in I \times J\}$. On complète le tableau par deux marges ; une marge colonne qui contiendra les proportions $p_{i.} = \sum\{f_{ij}/j \in J\}$, $i=1, \dots, n$, une marge ligne qui contiendra celles $p_{.j} = \sum\{f_{ij}/i \in I\}$, $j=1, \dots, m$. Ces fréquences relatives sont dites marginales ; $p_{i.}$ (resp. $p_{.j}$) est une mesure de l'"importance" (numérique) de la i -ème (resp. j -ème) modalité du premier (resp. second) caractère.

Pour analyser métriquement l'ensemble des modalités du premier caractère par rapport aux modalités du second caractère, on est conduit à représenter I par le nuage suivant de R^m :

$$\mathcal{N}(I) = \{(f_{.j}^i, p_{i.}) / i \in I\}, \quad (40)$$

où à $i \in I$, on a associé le point de R^m

$$f_{.j}^i = (f_{.1}^i, \dots, f_{.j}^i, \dots, f_{.m}^i), \quad (41)$$

dont la suite des coordonnées est la suite des proportions conditionnelles $f_{.j}^i = (f_{ij}/p_{i.})$: proportion, dans la i -ème classe définie par la première variable partition, des individus possédant la j -ème modalité du second caractère, on retient ainsi pour la description de i son "profil" à travers J . On préserve l'importance de présence de i en affectant le sommet $f_{.j}^i$ du poids $p_{i.}$.

Relativement à l'analyse de $\mathcal{N}(I)$, on munit l'espace ambiant R^m , dont la base canonique est notée $\{e_j / 1 \leq j \leq m\}$ de la métrique q suivante

$$q(e_j, e_h) = \begin{cases} 0 & \text{si } h \neq j \\ 1/p_{.j} & \text{si } h = j \end{cases} \quad (42)$$

Rappelons que l'introduction de cette métrique, dite du χ^2 , se justifie pour des raisons algébrique et statistique. la raison algébrique est définie par le condition dite de l'"équivalence distributionnelle" qui exprime que les distances entre éléments de I ainsi que celles entre éléments de J, pour une représentation analogue de $\mathcal{N}(J)$ relativement à I, restent invariantes lorsqu'on remplace deux éléments i' et i'' de I de masses respectives $p_{i'}$ et $p_{i''}$, qui ont la même représentation dans \mathbb{R}^m , pour un seul point i de masse $p_i = p_{i'} + p_{i''}$. D'autre part, avec une telle métrique (42) le carré de la distance entre i_1 et i_2 de I est exactement la distance du χ^2 , par rapport à la loi de probabilité $\{p_{.j}/j \in J\}$ entre les deux distributions $\{f_j^{i_1}/j \in J\}$ et $\{f_j^{i_2}/j \in J\}$. Enfin, le moment total d'inertie du nuage $\mathcal{N}(I)$ (resp. $\mathcal{N}(J)$) est, au facteur k .. près, la statistique du χ^2 attachée au tableau de contingence.

Il va sans dire que le nuage $\mathcal{N}(J)$ associé à J se trouve défini comme suit

$$\mathcal{N}(J) = \{ (f_I^j, p_{.j}) / j \in J \}, \tag{43}$$

où f_I^j est le point de \mathbb{R}^n

$$(f_1^j, \dots, f_i^j, \dots, f_n^j), \tag{44}$$

dont la suite des coordonnées est celle des proportions conditionnelles $f_i^j = (f_{ij}/p_{.j})$.

Conformément à (42), l'espace ambiant \mathbb{R}^n du nuage $\mathcal{N}(J)$, dont la base canonique est notée $\{f_i / 1 \leq i \leq n\}$, sera muni de la métrique q suivante

$$q^*(f_i, f_{i'}) = \begin{cases} 0 & \text{si } i' \neq i \\ 1/p_{i_1} & \text{si } i' = i \end{cases}, \tag{45}$$

Dans ces conditions, relativement à la recherche d'une classification des moindres carrés de J, la formule (1) de décomposition de l'inertie du nuage $\mathcal{N}(J)$ devient

$$\sum_{j \in J} p_{.j} \|f_I^j - g_I\|^2 = \sum_{1 \leq r \leq p} [\sum_{j \in J_r} p_{.j} \|f_I^j - g_I^r\|^2] + \sum_{1 \leq r \leq p} p(J_r) \|g_I^r - g_I\|^2, \tag{46}$$

où $\{J_r / 1 \leq r \leq p\}$ désigne une partition en p classes non vides de J. Le poids de la r -ème classe est noté

$$p(J_r) = \sum_{j \in J_r} p_{.j}, \tag{47}$$

son centre de gravité est le point

$$g_I^r = \frac{1}{p(J_r)} \sum_{j \in J_r} p_{.j} f_I^j ; \tag{48}$$

La suite des coordonnées du point g_I^r peut être notée

$$(f_1^{Jr}, \dots, f_i^{Jr}, \dots, f_n^{Jr}), \quad \text{où}$$

$$(\forall i \in I), f_i^{J_r} = f(i, J_r) / p(J_r) = \left(\sum_{j \in J_r} f_{ij} \right) / \left(\sum_{j \in J_r} p_{.j} \right), \quad (49)$$

qui représente une proportion relative.

Le centre de gravité g_I du nuage total $\mathcal{N}(J)$ est le point dont la i -ème coordonnée peut se mettre sous la forme (49) :

$$p_{i.} = \left(\sum_{j \in J} f_{ij} \right) / \left(\sum_{j \in J} p_{.j} \right)$$

Le critère maximisé, seconde somme du second membre de (46), peut être explicité comme suit

$$\sum_{1 \leq r \leq p} p(J_r) \left\{ \sum_{i \in I} \frac{1}{p_{i.}} (f_i^{J_r} - p_{i.})^2 \right\}, \quad (50)$$

Par ailleurs, l'inertie du nuage $\mathcal{N}(I)$ projeté sur le sous espace, dont l'origine est placé au centre de gravité du nuage, engendré par une base q -orthonormale $\{a_r / 1 \leq r \leq p\}$, est

$$\sum_{1 \leq r \leq p, i \in I} \left\{ \sum p_{i.} [q(a_r, f_J^i - g_J)]^2 \right\}, \quad (51)$$

En exprimant chacun des vecteurs a_r par rapport à la base canonique de R^m ;

$$(\forall r = 1, \dots, p), a_r = \sum_{1 \leq j \leq m} \alpha_{rj} e_j, \quad (52)$$

la condition de normalisation s'écrit

$$(\forall r = 1, \dots, p), \sum_{1 \leq j \leq m} (\alpha_{rj}^2 / p_{.j}) = 1 \quad (53)$$

et on a

$$q(a_r, f_J^i - g_J) = \sum_{1 \leq j \leq m} \frac{\alpha_{rj}}{p_{.j}} (f_{ij} - p_{.j}), \quad (54)$$

Si on impose maintenant à la suite des axes factoriels $(a_r / 1 \leq r \leq p)$ la condition suivante

$$(C): \quad (\forall r = 1, \dots, p), (J_r \neq \emptyset); \alpha_{rj} = \begin{cases} 0 & \text{si } j \notin J_r \\ \delta(r) p_{.j} & \text{si } j \in J_r, \end{cases}$$

où $\delta(r) > 0$; l'ensemble des parties $\{J_r / 1 \leq r \leq p\}$ définit une partition de J dont chaque classe se trouve indiquée par un axe factoriel.

La condition de normalisation (53) permet de préciser $\delta(r)$;

$$(\forall r = 1, \dots, p), \delta(r) = 1 / \sqrt{p(J_r)}; \quad (55)$$

de sorte que le critère maximisé (51) se met sous la forme

$$\sum_{1 \leq r \leq p, i \in I} \sum_{j \in J_r} p_{i.} \left[\sum_{j \in J_r} \frac{(f_{ij} - p_{.j})^2}{p(J_r)} \right] = \sum_{1 \leq r \leq p, i \in I} \left[\frac{f(i, J_r) - p_{i.} p(J_r)}{\sqrt{p_{i.} p(J_r)}} \right]^2, \quad (56)$$

et l'expression (50) peut aisément se réduire à cette dernière forme ; d'où le théorème

THEOREME. *Relativement à un même tableau de contingence $I \times J$ de dimension $n \times m$; l'analyse factorielle des correspondances de dimension p ($p < m$) de I à travers J , sous la contrainte définie par la condition (C) ci-dessus, est équivalente à une classification des moindres carrés de J en p classes non vides.*

A.1.2. Recherche d'un système de facteurs.

Pour l'interprétation "factorielle" d'une classification des moindres carrés du nuage $\mathcal{N}(I)$, on recherchera ici un système de facteurs dont chacun est présenté comme une fonction sur I . La forme retenue de l'équation factorielle sera donc ici, la suivante :

$$\sum_{i \in I} \mu_i q(i', i) \phi(i') = \lambda \phi(i), \tag{1}$$

dans cette formule, on a, pour simplifier, noté i au lieu de $(M_i, -G)$ (cf. notation du paragraphe A.1.1.) ; on peut d'ailleurs supposer une fois pour toutes et sans restreindre la généralité qu'on a placé le centre de gravité G du nuage à l'origine. q est, rappelons-le, la métrique dont se trouve muni l'espace ambiant du nuage. En vérité et contrairement au paragraphe A.1.1. précédent, on "oubliera" l'espace ambiant du nuage ; c'est-à-dire l'autre côté du tableau des données, pour ne retenir que le tableau des proximités définis par les produits scalaires

$$\{q(i', i) = q(M_i, -G, M_{i'} - G) / (i, i') \in I \times I\}, \tag{2}$$

Toutefois, on précisera la correspondance entre facteur et axe factoriel associé.

Cette partie de notre étude apparaîtra, d'une certaine façon, comme un cas particulier de celle du paragraphe B.1.1. où on cherche à présenter factoriellement un arbre binaire de classification.

1. Définition de paramètres liés à la classification.

Désignons par

$$(I_1, I_2, \dots, I_r, \dots, I_p), \tag{3}$$

la suite des classes ordonnée selon un critère qui apparaîtra bientôt. Comme ci-dessus (§ A.1.1.), on associera à chacune des classes sa masse et son centre de gravité

$$(\forall r=1, \dots, p), \nu(r) = \sum_{i \in I_r} \mu_i, G_r = \sum_{i \in I_r} \frac{\mu_i}{\nu(r)} M_i, \tag{4}$$

A la suite (3) de parties disjointes, associons la suite croissante de parties

$$(I_1, I_1 \cup I_2, I_1 \cup I_2 \cup I_3, \dots, I_1 \cup I_2 \cup \dots \cup I_p), \tag{5}$$

faisons correspondre à chacune de ces parties sa masse $N(r)$ et son centre de gravité H_r ,

$$(\forall r=1, \dots, p), N(r) = \sum_{s \leq r} v(s), H_r = \sum_{i \in I_1 \cup \dots \cup I_r} \frac{u_i}{N(r)} M_i = \sum_{s \leq r} \frac{v(s)}{N(r)} G_s, \quad (6)$$

Enfin, à la suite des couples de parties disjointes de la forme $(I_1 \cup \dots \cup I_r, I_{r+1})$, on associera la quantité

$$\lambda(r) = \frac{N(r)v(r+1)}{N(r+1)} \|G_{r+1} - H_r\|^2, \quad (7)$$

qui représente la part d'inertie perdue dans la réunion de I_{r+1} à $I_1 \cup I_2 \cup \dots \cup I_r$; $r=1, \dots, p$.

On suppose précisément que la suite (3) des classes est ordonnée par valeurs croissantes de $\lambda(r)$. Ce choix qui, comme nous le verrons au paragraphe B.1.1, est possible; ne restreint en rien la généralité. Il a pour raison une présentation plus fidèlement factorielle de la classification.

2. Base orthonormale de fonctions au sens de Helmert.

Désignons par P l'ensemble des entiers $\{1, 2, \dots, r, \dots, p\}$ muni de la mesure $\{v(r)/1 \leq r \leq p\}$; P aura pour rôle dans la suite de coder les classes de la partition $\{I_r/1 \leq r \leq p\}$.

L'ensemble R^P des fonctions à valeurs réelles sur P est alors muni du produit scalaire

$$\langle \phi, \psi \rangle = \sum_{1 \leq r \leq p} v(r) \phi(r) \psi(r), \quad (8)$$

pour tout couple (ϕ, ψ) de fonctions numériques sur P .

L'espace des fonctions de moyenne nulle sur P :

$$\{\phi \in R^P / \sum_{1 \leq r \leq p} v(r) \phi(r) = 0\}, \quad (9)$$

est de dimension $(p-1)$. Une base orthonormale peut en être fournie par le système de fonctions de Helmert (cf. [4]) qu'on notera $(\phi^0, \phi^1, \dots, \phi^{(p-1)})$, défini comme suit:

ϕ^0 est la fonction constante, égale à 1 sur P

ϕ^1 prend des valeurs non nulles sur $\{1, 2\}$ et $\phi^1(2) > 0$.

ϕ^2 prend des valeurs non nulles sur $\{1, 2, 3\}$ et $\phi^2(3) > 0$.

...

ϕ^r prend des valeurs non nulles sur $\{1, 2, \dots, r+1\}$ et $\phi^r(r+1) > 0$.

...

$\phi^{(p-1)}$ prend des valeurs non nulles sur $\{1, 2, \dots, p\}$ et $\phi^{(p-1)}(p) > 0$.

Cette suite orthonormale de fonctions peut être déterminée de proche en proche et on a

$$\begin{aligned} \phi^r(s) &= -\sqrt{v(r+1)}/\sqrt{N(r)N(r+1)}, \text{ pour } 1 \leq s \leq r \\ \phi^r(r+1) &= \sqrt{N(r)}/\sqrt{v(r+1)N(r+1)} \\ \phi^r(s) &= 0, \text{ pour } s > r+1; \end{aligned} \quad (10)$$

pour tout $r=1,2,\dots,(p-1)$.

L'orthogonalité à ϕ^0 de la suite $(\phi^1, \phi^2, \dots, \phi^{(p-1)})$ exprime que chacune des fonctions ϕ^r , $1 \leq r \leq (p-1)$, est de moyenne nulle sur P. Ainsi, cette suite de fonctions $(\phi^r / 1 \leq r \leq (p-1))$ constitue une base orthonormale de l'espace (9) des fonctions numériques de moyenne nulle sur P.

Etendons naturellement la définition de chacune des fonctions de la suite $(\phi^r / 1 \leq r \leq p)$ à I en posant

$$\phi^r(i) = \phi^r[\text{classe}(i)] ;$$

c'est-à-dire,

$$(\forall s=1, \dots, p), (\forall i \in I_s), \phi^r(i) = \phi^r(s). \tag{11}$$

De cette façon, pour tout $r=1, \dots, p-1$; la fonction ϕ^r indique la $(r+1)$ ème classe de la suite (3) des classes, par rapport à la réunion des classes qui précèdent ; soit I_{r+1} par rapport à $I_1 \cup \dots \cup I_r$.

3. Présentation factorielle de la classification

Relativement à la suite des fonctions $(\phi^1, \phi^2, \dots, \phi^{(p-1)})$ sur I établie ci-dessus (cf. formules (10) et (11)), considérons la matrice des proximités suivante entre éléments de I

$$\{Q(i, i') / (i, i') \in I \times I\}, \tag{12}$$

où

$$(\forall (i, i') \in I^2), Q(i, i') = \sum_{1 \leq r \leq (p-1)} \lambda(r) \phi^r(i) \phi^r(i'), \tag{13}$$

En désignant par $s(i, i')$ l'indice le plus grand de la classe contenant l'un des deux objets i ou i' , la somme précédente se réduit à

$$(\forall (i, i') \in I^2), Q(i, i') = \sum_{r \geq [s(i, i') - 1]} \lambda(r) \phi^r(i) \phi^r(i'), \tag{14}$$

où $\lambda(r)$ a été défini par la formule (7) ci-dessus.

3.1. Extension linéaire de la définition de Q.

Pour i fixé, la fonction $Q(i', i)$ est constante par rapport à i' , pour i' décrivant une même classe I_r . Puisque la classification correspond à projeter chacun des points M_i sur le centre de gravité de la classe à laquelle il appartient ; on posera

$$Q(i', i) = Q(G_r, i) \quad \text{si } i' \in I_r, \tag{15}$$

de sorte qu'on peut poser

$$Q\left(\sum_{i' \in I_r} \alpha_{i'} i', i\right) = \sum_{i' \in I_r} \alpha_{i'} Q(i', i), \tag{16}$$

pour tout système $\{\alpha_{i'} / i' \in I_r\}$ de coefficients numériques. Le second membre de (16) peut d'ailleurs se mettre sous la forme

$$\beta_r Q(G_r, i) \quad \text{où } \beta_r = \sum_{i' \in I_r} \alpha_{i'}, \tag{16'}$$

On étendra ensuite Q de façon linéaire à l'ensemble des combinaisons linéaires de la forme $\sum_r \beta_r (G_r - G)$, en posant

$$Q \left[\sum_{1 \leq r \leq p} \beta_r (G_r - G) \right] = \sum_{1 \leq r \leq p} \beta_r Q(G_r - G), \quad (17)$$

On étendra enfin Q à l'ensemble de l'espace engendré par $\{M_i - G\} / i \in I$ en posant pour

$$(M - G) = \sum_{i \in I} \alpha_i (M_i - G), \text{ où pour tout } i, \alpha_i \text{ est un nombre réel,}$$

$$Q(M - G) = \sum_{1 \leq r \leq p} \beta_r Q(G_r - G)$$

où

$$\beta_r = \sum_{i \in I_r} \alpha_i, \quad (18)$$

3.2. THEOREME. La suite des fonctions $(\phi^1, \phi^2, \dots, \phi^{(p-1)})$ définit la suite des facteurs normalisés de moyenne nulle de la matrice des proximités sur I définie par (14) ; la valeur propre associée à ϕ^r étant $\lambda(r), r=1, \dots, (p-1)$. L'axe factoriel unitaire associé à ϕ^r au sens de la métrique Q est défini par

$$\frac{(G_{r+1} - H_r)}{\|G_{r+1} - H_r\|} \quad (r=1, \dots, (p-1))$$

où la norme $\|\cdot\|$ reste relative à q .

L'équation factorielle associée à Q se met sous la forme

$$(\forall i \in I), \sum_{i' \in I} \mu_{i'} Q(i', i) \Psi(i') = \lambda \Psi(i), \quad (19)$$

En remplaçant $Q(i, i')$ par le second membre de (13) et en posant $\Psi = \phi^s$, le premier membre de (19) se met sous la forme

$$\sum_{i' \in I} \mu_{i'} \left[\sum_{1 \leq r \leq (p-1)} \lambda(r) \phi^r(i) \phi^r(i') \right] \phi^s(i'), \quad (20)$$

En inversant à présent les deux signes sommes, l'expression précédente devient

$$\sum_{1 \leq r \leq (p-1)} \lambda(r) \phi^r(i) \left[\sum_{i' \in I} \mu_{i'} \phi^r(i') \phi^s(i') \right], \quad (21)$$

Compte tenu du caractère orthonormal de la famille de fonctions $\{\phi^r / 1 \leq r \leq (p-1)\}$; la quantité entre crochets n'est différente de 0 que si $r=s$ où elle vaut 1. La quantité (20) se réduit par conséquent à $\lambda(s) \phi^s(i)$; ce qui prouve que la fonction ϕ^s est facteur, solution de (19), relativement à la valeur propre $\lambda(s)$.

Reprenons l'expression (21) sous la forme condensée définie par le premier nombre de (19) ; soit

$$\sum_{i' \in I} \mu_{i'} Q(i', i) \phi^S(i'), \quad (22)$$

qui peut, compte tenu de la définition de ϕ^S , se mettre sous la forme

$$\sum_{1 \leq r \leq p} \left\{ \sum_{i \in I_r} \mu_i Q(i', i) \right\} \phi^S(r), \quad (23)$$

qui s'écrit, compte tenu de l'extension linéaire de Q (cf. § 3.1. précédent),

$$\sum_{1 \leq r \leq p} v(r) Q(G_r, i) \phi^S(r) \quad (24)$$

et en tenant compte de la définition de ϕ^S et de l'extension linéaire de Q (cf. (10), (17)) le premier membre de (19) se réduit à

$$\sqrt{M(s, s+1)} Q(G_{s+1} - H_s, i), \quad (25)$$

où on a noté

$$M(s, s+1) = N(s) v(s+1) / N(s+1), \quad (26)$$

et où H_s a été défini au paragraphe 1 ci-dessus (cf. formule (6)).

Compte tenu du second membre de l'équation factorielle (cf. (19)) et de la valeur de $\lambda(s)$ (cf. formule (7)), l'axe factoriel unitaire au sens de q, associé au facteur ϕ^S au sens Q, est défini par

$$A_{s+1} = (G_{s+1} - H_s) / \|G_{s+1} - H_s\|, \quad (27)$$

on, a en effet

$$\phi^S = \frac{1}{\sqrt{\lambda(s)}} Q(A_{s+1}), \quad (28)$$

et on a du même coup démontré les relations

$$(\forall j=1, \dots, s) (\forall i \in I_j), Q(A_{s+1}, i) = \frac{-v(s+1)}{N(s+1)} \|G_{s+1} - H_s\|$$

$$(\forall i \in I_{s+1}), Q(A_{s+1}, i) = \frac{N(s)}{N(s+1)} \|G_{s+1} - H_s\|$$

$$(\forall j=s+2, \dots, p) (\forall i \in I_j), Q(A_{s+1}, i) = 0, \quad (29)$$

Ces dernières formules peuvent s'obtenir directement après un patient calcul, à partir des relations (7), (10) et (13). On commencera pour cela par établir les relations

$$Q(G_j, G_h) = \frac{-vh}{N(h-1) N(h)} \left[\frac{N(h-1)}{N(h)} \|G_h - H_{(h-1)}\|^2 + \sum_{r \geq h} \left[\frac{v(r+1)}{N(r+1)} \|G_{r+1} - H_r\|^2 \right] \right],$$

pour $j < h$ et

$$Q(G_j, G_j) = \frac{N(j-1)}{N(j)} \|G_j - H_{(j-1)}\|^2 + \sum_{r \geq j} \left[\frac{v(r+1)}{N(r+1)} \|G_{r+1} - H_r\|^2 \right], \quad 1 \leq j, h \leq p. \quad (30)$$

D'autre part, on peut vérifier que le système d'axes factoriels $\{A_s / 2 \leq s \leq p\}$ est Q-orthonormal. C.Q.F.D.

A.2. Critère du "saut minimal" ; recherche d'un système de facteurs.

I - INTRODUCTION

Nous nous proposons de reprendre ici l'intéressant travail de M. Gondran (cf. [2]), mais avec une présentation formelle dégagée de certaines contingences algorithmiques et qui respecte mieux la nature du rapprochement voulu entre l'analyse factorielle et une même classification d'un ensemble I muni d'un indice de distance ou de proximité. Cette classification est celle d'un niveau donné de l'arbre binaire des classifications sur I , bâti pas à pas à partir du critère du "saut minimal", et que nous présenterons rapidement au paragraphe suivant.

Sans restreindre en rien la généralité ; nous supposerons que I est muni d'un indice de proximité "s" à valeurs dans l'intervalle $[0, 1]$. Ceci nous conduira à munir cet ensemble de deux opérations V et \wedge (i.e. "max" et "min") qu'on notera par l'"addition" \oplus et par la "multiplication" $*$. L'élément neutre pour \oplus (resp. $*$) est 0 (resp. 1). Le triplet $([0, 1], \oplus, *)$ définit dans ces conditions un semi-anneau. Ces notations sont beaucoup plus conformes à l'usage que celles de [2] ; elles ont d'autre part, un rôle important dans la reconnaissance de la forme retenue de l'équation factorielle pour l'interprétation de la classification.

En exprimant la vraie nature de cette forme de l'équation factorielle, nous verrons qu'il s'agit de la recherche d'un système de facteurs dont chacun est défini comme une fonction sur I ; alors que dans [2]. On parle de façon un peu abstraite de la recherche d'un système de vecteurs de S^n , où S représente $R^+ \cup \{\infty\}$ et n est le cardinal de I .

D'autre part, nous nous dégagerons d'un intermédiaire algorithmique non nécessaire dans la construction de cette présentation factorielle de la classification et qui est l'"arbre de longueur minimum" (cf. [2]). En effet, la notion essentielle est celle de proximité ultramétrique sur I associée à l'arbre binaire des classifications basé sur le critère du "saut minimal" (cf. § ci-dessous) auquel plusieurs algorithmes peuvent conduire. De la sorte, certaines démonstrations deviendront plus directes.

Enfin, bien qu'il s'agisse de l'interprétation d'une classification ; il ne faut pas oublier que cette dernière est celle d'un même niveau de l'arbre binaire des classifications mentionné ci-dessus. D'un certain point de vue ce paragraphe A2 constitue donc un passage entre la section A et celle B, où il s'agit de présenter la recherche d'un arbre des classifications. D'ailleurs nous verrons au paragraphe B.2.1. que dans ce dernier cas ; ce type d'approche devient plus naturel.

II - ARBRE DES CLASSIFICATIONS ET MATRICE ASSOCIEE DES PROXIMITES ULTRAMETRIQUES.

On considère un ensemble I d'objets de cardinal n , qu'on peut noter sous la forme $I = \{1, 2, \dots, i, \dots, n\}$, muni d'un indice de proximité "s". s est défini comme une application de $I \times I$ dans l'intervalle $[0, 1]$; associant à chaque couple (i, i') de $I \times I$, un nombre $s(i, i')$, compris entre 0 et 1 ($0 \leq s(i, i') \leq 1$), qui "mesure" la "ressemblance" entre i et i' . La fonction $s(i, i')$ respecte les deux conditions

- a) $(\forall (i,i') \in I \times I), s(i,i') = s(i',i)$; condition de symétrie.
- b) $(\forall i \in I), s(i,i) = 1$; condition d'échelle.

En tenant compte de la condition (a), on peut restreindre la définition de s à l'ensemble des paires d'objets distincts de I , c'est-à-dire, à l'ensemble $P_2(I)$ des parties à deux éléments de I . On supposera pour simplifier que la fonction s est injective sur ce dernier ensemble ; c'est-à-dire qu'il n'existe pas deux paires distinctes sur lesquelles l'indice s prend la même valeur. Dans ce dernier cas en effet, l'arbre que nous allons associer à s est nécessairement binaire.

La définition algorithmique la plus simple de l'arbre des classifications basé sur le critère du "saut minimal" est sans doute, la suivante :

On démarre avec la partition "discrète" où chaque classe contient un seul élément de l'ensemble I à organiser. A chaque niveau on réunit les deux classes les plus proches au sens de l'indice de proximité suivant entre deux parties disjointes quelconques K et L de I

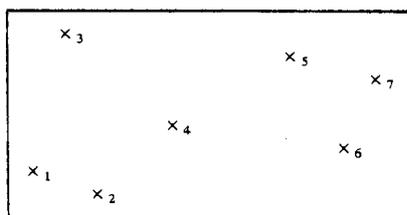
$$p(K,L) = \max\{s(i,j) / (i,j) \in K \times L\}, \tag{1}$$

De la sorte, dans le cas ci-dessus mentionné où s est injective sur $P_2(I)$; si le niveau 0 est celui défini par la partition discrète à n classes, celui 1, définit une partition à $(n-1)$ classes. Le dernier niveau est celui d'indice $(n-1)$, il correspond à la partition "grossière" à une seule classe.

Considérons à titre d'illustration, la configuration géométrique suivante de 7 points et la matrice des proximités associée. Pour établir cette dernière, nous avons posé, en utilisant le cm comme unité

$$(\forall (i,j) \in I^2, s(i,j) = [7 - d(i,j)] / 7,$$

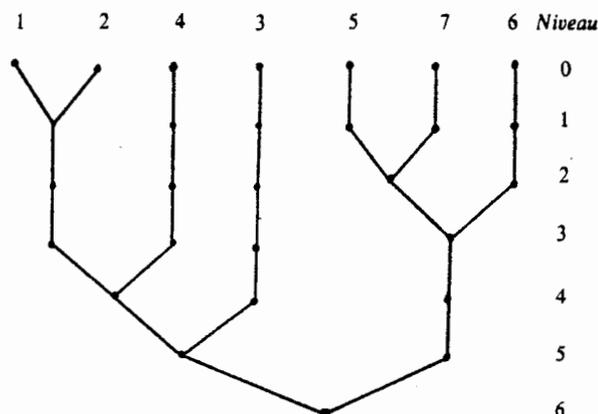
où $d(i,j)$ est la distance entre i et j . A une homothétie près on obtient la figure suivante :



	1	2	3	4	5	6	7
1	1.						
2	.857	1.					
3	.671	.586	1.				
4	.657	.714	.664	1.			
5	.264	.336	.364	.607	1.		
6	.186	.300	.193	.500	.743	1.	
7	.093	.179	.171	.436	.814	.800	1.

"matrice des proximités"

L'arbre binaire des classifications associé par l'algorithme précédent, est le suivant.



De nombreux algorithmes d'inspirations informatiques différentes conduisent à un tel arbre des classifications. On peut commencer par citer celui de Mme Nicolau (cf. [12]) qui utilise le plus directement une notion de proximité quelconque, en particulier celle (1) ci-dessus pour aboutir à la représentation polonaise de l'arbre des classifications. L'algorithme de Roux (cf. [13]) qui fait suite aux travaux de N. Jardine et de R. Sibson (cf. [7]) fait appel à la notion de "distance ultramétrique inférieure maxima". L'algorithme "lexicographique" que nous avons proposé (cf. [10]) met en évidence que la seule donnée nécessaire pour la construction d'un tel arbre, a un caractère ordinal : il s'agit de la "préordonnance" sur I ; c'est-à-dire, le préordre total sur $P_2(I)$ associé à l'indice de proximité sur I . On peut également citer l'algorithme de S.C. Johnson (cf. [8]). Enfin l'algorithme de M. Gondran (cf. [2]) utilise de façon privilégiée l'arbre de longueur minimale" de J.B. KRUSKAL (cf. [9]) de façon conforme à J.C. Gower et G.J.S. Ross (cf. [3]).

Nous désignerons par I_1 la classe définie par le noeud formé au niveau 1 ; ainsi dans l'exemple précédent $I_3 = \{5, 6, 7\}$, $I_5 = \{1, 2, 3, 4\}$. a_ℓ et b_ℓ indiqueront les deux classes composantes dont la réunion définit I_ℓ ; ainsi $a_3 = \{5, 7\}$ et $b_3 = \{6\}$, $a_5 = \{1, 2, 4\}$ et $b_5 = \{3\}$. On notera par λ_ℓ la valeur de l'indice de proximité entre a_ℓ et b_ℓ conformément à la formule (1) :

$$\lambda_\ell = \max\{s(i, j) / (i, j) \in a_\ell \times b_\ell\}, \quad (2)$$

De par la définition de l'algorithme, on voit sans peine que la suite des valeurs de λ_ℓ est strictement décroissante :

$$\lambda_1 > \lambda_2 > \dots > \lambda_\ell > \dots > \lambda_{(n-1)}, \quad (3)$$

On distinguera enfin par $(C_1^\ell, C_2^\ell, \dots, C_k^\ell, \dots, C_{(n-\ell)}^\ell)$ la suite des classes de la partition de niveau ℓ . Pour fixer les idées et sans restreindre en rien la généralité, on suppose cette suite des classes rangée comme suit :

(a) $C_1^\ell = I_\ell$

(b) $p(C_1^\ell, C_k^\ell)$ est une fonction décroissante de k ; $1 \leq k \leq (n-\ell)$.

L'ensemble des classes $\{C_k^\ell / 1 \leq k \leq (n-\ell)\}$ représente, ce que nous appellerons au paragraphe B.1.1. suivant, l'ensemble des "parties maximales" au niveau $(\ell+1)$.

La proximité ultramétrique entre deux éléments i et j de I , conformément à l'arbre binaire des classifications ci-dessus défini, peut être exprimée de deux façons que nous montrerons bientôt équivalentes. La première est $s_1^*(i,j) = \lambda_\ell$; où ℓ est l'indice du niveau où i et j se retrouvent réunis pour la première fois dans I_ℓ .

La deuxième façon est

$$s_2^*(i,j) = \bigvee_{c \in \mathcal{C}(i,j)} \{s(i_0, i_1) \wedge s(i_1, i_2) \wedge \dots \wedge s(i_r, i_{r+1})\}, \quad (4)$$

où $c = (i_0 = i, i_1, i_2, \dots, i_r, i_{r+1} = j)$ est une chaîne reliant i à j et où $\mathcal{C}(i,j)$ est l'ensemble de toutes les chaînes reliant i à j .

LEMME 1. $(\forall (i,j) \in I \times I) s_1^*(i,j) = s_2^*(i,j)$.

En appelant $s(i,j)$ valeur de l'arc (i,j) ; dans toute chaîne $c(i,j)$, l'arc le plus "tênu" a une valeur nécessairement inférieure ou égale à λ_ℓ , compte tenu de la définition de λ_ℓ (cf. (2)). D'où

$$s_2^*(i,j) \leq \lambda_\ell = s_1^*(i,j).$$

Soient (i_a, i_b) le couple de sommets de $(a_\ell \times b_\ell)$ tel que

$$\lambda_\ell = s(i_a, i_b);$$

il existe nécessairement entre i et i_a (resp. j et i_b) une chaîne dont les sommets appartiennent à a_ℓ (resp. b_ℓ) et dont chaque arc possède une valeur supérieure à λ_ℓ . Une telle chaîne peut être construite pas à pas en partant des deux sommets à relier et en considérant à chaque fois l'arc de valeur maximale joignant les deux composantes d'un même noeud, lesquelles contenant respectivement deux sommets intermédiaires de la chaîne.

Par conséquent, on a nécessairement

$$s_2^*(i,j) = \lambda_\ell \quad \text{C.Q.F.D.}$$

Nous notons désormais $s^*(i,j)$ la proximité ultramétrique entre i et j et par S^* la matrice carrée $I \times I$ des proximités ultramétriques. Dans l'exemple précédent cette matrice est la suivante

	1	2	3	4	5	6	7
1	1						
2	.857	1					
3	.671	.671	1				
4	.714	.714	.671	1			
5	.607	.607	.607	.607	1		
6	.607	.607	.607	.607	.800	1	
7	.607	.607	.607	.607	.814	.800	1

"matrice des proximités ultramétriques S^* "

Lemme 2. Les sommets d'un même noeud sont à égale proximité ultramétrique de tout sommet extérieur à ce noeud.

Soit I_ℓ un noeud de l'arbre et soit j un sommet n'appartenant pas à I . Il s'agit de prouver que $(\forall i, i' \in I_\ell), s^*(i, j) = s^*(i', j)$

Or, on a nécessairement

$$s^*(i, i') > s^*(i, j) \text{ et } s^*(i, i) > s^*(i', j),$$

ce qui permet de se rendre compte du résultat en faisant appel à la forme s_2^* de la proximité ultramétrique (cf. (4)) ; ou bien en faisant appel à la propriété caractéristique d'un espace ultramétrique : "Tout triangle est isocèle, la base étant le plus petit côté". C.Q.F.D.

Ces lemmes seront utilisés aux paragraphes suivants

III - FORME DE L'EQUATION FACTORIELLE

L'intervalle $[0, 1]$ est l'ensemble des valeurs possibles de l'indice "s". Munissons $[0, 1]$ des deux opérations \vee et \wedge (i.e. "max" et "min") qu'on peut noter par l'"addition" et par la "multiplication" \oplus et $*$. Dans ces conditions $([0, 1], \oplus, *)$ définit une structure de semi anneau où l'élément neutre pour l'addition est 0 et celui, pour la multiplication, 1.

Etendons la définition de l'addition \oplus et de la multiplication $*$ à l'ensemble $M_n([0, 1])$ des matrices carrées d'ordre n , dont les éléments appartiennent à l'intervalle $[0, 1]$. Muni de ces deux opérations, $M_n([0, 1])$ admet alors la même structure de semi-anneau où l'élément nul (i.e. neutre pour \oplus) et où l'élément unité (i.e. neutre pour $*$) est défini par la matrice dont les seuls éléments non nuls sont ceux de la diagonale principale qui sont tous égaux à 1.

On indiquera par 0_n la matrice nulle et par 1_n , celle, unité. Dans ces conditions, S (resp. S^*) désignant la matrice des proximités (resp. des proximités ultramétriques) sur I (cf. paragraphe précédent), on a les propriétés suivantes, qu'il est facile de démontrer en utilisant la forme $s_2^*(i, j)$ (cf. formule (4)) de la proximité ultramétrique $s^*(i, j)$ entre deux sommets quelconques i et j .

$$S^* = \sum_{1 \leq k \leq (n-1)} S^k, \quad (5)$$

$$S^* = S * S^* = S^* * S, \quad (6)$$

et

$$S^* = S^* \oplus 1_n = 1_n \oplus S^*, \quad (7)$$

où le signe Σ est, bien entendu, relatif à l'"addition" définie.

La forme de l'équation factorielle nécessaire ici pour interpréter la classification est la même que celle (1) du paragraphe A.1.2. précédent, où l'inconnue est un système de facteurs dont chacun est défini comme une fonction sur I . Ce système de facteurs, attaché à un même niveau ℓ de l'arbre binaire des classifications, comportera $(n - \ell)$ éléments dont chacun in-

diquera une même classe C_k^ℓ , $1 \leq k \leq (n-\ell)$, de ce niveau (cf. §2 ci-dessus)

Avec les notations de ce paragraphe, l'équation factorielle devient ici

$$(\forall i \in I), \sum_{i' \in I} s(i', i) * \phi(i') = \lambda * \phi(i), \tag{8}$$

puisque chaque élément de I est affecté de la masse unité.

Dans la formule (8), la valeur propre est indiquée par λ qui est un nombre compris entre 0 et 1. La valeur que nous considérons bientôt de λ pour interpréter la classification $\{C_k^\ell / 1 \leq k \leq (n-\ell)\}$ est précisément définie par $\lambda_\ell = p(a_\ell, b_\ell)$ (cf. formule (2)).

Commençons par rappeler dans notre cadre les propriétés préliminaires (lemmes 2 et 3 de [2]). Pour cela désignons par Φ_λ l'ensemble des fonctions sur I répondant à l'équation (8).

PROPRIETE 1. Pour toute fonction ϕ de Φ_λ , on a : $(\forall i \in I), \phi(i) \leq \lambda$

En effet, puisque $s(i, i) = 1$, $\phi(i) \leq \lambda * \phi(i)$; ce qui implique nécessairement $\phi(i) \leq \lambda$, pour tout i appartenant à I.

De la sorte, Φ_λ apparaît comme l'ensemble des fonctions ϕ à valeurs dans l'intervalle $[0, \lambda]$, solutions de l'équation

$$(\forall i \in I), \sum_{i' \in I} s(i', i) * \phi(i') = \phi(i), \tag{9}$$

Considérons à présent pour λ , la valeur λ_ℓ que nous venons de mentionner ci-dessus.

PROPRIETE 2. Pour toute fonction ϕ de $\Phi_{\lambda(\ell)}$, les valeurs de ϕ sur une même classe C_k^ℓ , $1 \leq k \leq (n-\ell)$, sont égales.

Compte tenu du critère de formation de la classification $\{C_k^\ell / 1 \leq k \leq (n-\ell)\}$; il existe nécessairement pour chacune des classes une chaîne passant par tous les sommets de la classe et dont chaque arc est de valeurs supérieure ou égale à λ_ℓ .

Soit C_k^ℓ l'une des classes et (i_1, i_2, \dots, i_r) la chaîne en question ; on a

$$(\forall j = 1, 2, \dots, (r-1)), s(i_j, i_{j+1}) > \lambda_\ell.$$

Considérons alors la suite des équations

$$\begin{aligned} \sum_{1 \leq i' \leq n} s(i', i_1) * \phi(i') &= \phi(i_1) \\ \sum_{1 \leq i' \leq n} s(i', i_2) * \phi(i') &= \phi(i_2) \\ &\dots\dots\dots \\ \sum_{1 \leq i' \leq n} s(i', i_r) * \phi(i') &= \phi(i_r) \end{aligned}$$

où ϕ appartient à $\Phi(\lambda(\ell))$.

La j -ème équation implique les deux relations suivantes

$$s(i_{(j-1)}, i_j) * \phi(i_{(j-1)}) \leq \phi(i_j), 2 \leq j \leq r$$

et

$$s(i_{(j+1)}, i_j) * \phi(i_{(j+1)}) \leq \phi(i_j), 1 \leq j \leq (r-1)$$

Comme $s(i_{(j-1)}, i_j) \wedge s(i_{(j+1)}, i_j) \geq \lambda_\ell$ pour tout $j=2, \dots, (r-1)$; les deux relations précédentes deviennent

$$\phi(i_{(j-1)}) \leq \phi(i_j), 2 \leq j \leq r$$

et

$$\phi(i_{(j+1)}) \leq \phi(i_j), 1 \leq j \leq (r-1)$$

La première, relation donne

$$\phi(i_1) \leq \phi(i_2) \leq \dots \leq \phi(i_r) ;$$

alors que la seconde relation donne

$$\phi(i_r) \leq \phi(i_{(r-1)}) \leq \dots \leq \phi(i_1)$$

Donc

$$\phi(i_1) = \phi(i_2) = \dots = \phi(i_r).$$

C.Q.F.D.

Remarquons que la propriété 2 démontrée reste vraie pour toute valeur propre $\lambda \leq \lambda_\ell$; de façon précise, si $\lambda \leq \lambda_\ell$, toute fonction solution de l'équation factorielle (8) est constante sur chacune des classes de la partition $\{C_k^\ell / 1 \leq k \leq (n-1)\}$.

IV - LE SYSTEME DES FACTEURS

Reprenons la suite des classes $(C_1^\ell, C_2^\ell, \dots, C_k^\ell, \dots, C_{(n-\ell)}^\ell)$ de la partition de niveau ℓ de l'arbre binaire des classifications. Cette suite des classes a été définie au paragraphe 2 ci-dessus après l'introduction des paramètres λ_ℓ (cf. formules (2) et (3)), à partir des conditions (a) et (b). Nous désignerons ici par i_k un élément courant de la classe C_k^ℓ .

Le lemme 2 ci-dessus (cf. § 2) nous permet d'assurer que la proximité ultramétrique entre deux sommets appartenant respectivement à C_k^ℓ et à C_h^ℓ ($k \neq h$), est la même quels que soient les deux éléments choisis ; l'un dans C_k^ℓ et l'autre dans C_h^ℓ . Cette proximité ultramétrique pourra donc être notée $s^*(i_k, i_h)$.

Considérons dans ces conditions la suite des fonctions $(\phi_1, \phi_2, \dots, \phi_k, \dots, \phi_{(n-\ell)})$ où la fonction ϕ_k sur I est définie de la manière suivante

$$\begin{aligned} \forall i_k \in C_k^\ell, \phi_k(i_k) = \lambda_\ell \\ (\forall h \neq k), (\forall i_h \in C_h^\ell), \phi_k(i_h) = s^*(i_k, i_h), 1 \leq k \leq (n-\ell) \end{aligned} \quad (10)$$

THEOREME. La suite des fonctions qu'on vient de définir $(\phi_1, \phi_2, \dots, \phi_{(n-l)})$, constitue une base unique de l'espace des solutions $\phi(\lambda_\rho)$ de l'équation factorielle (8) relativement à la valeur propre λ_ρ .

Les formules (10) montrent que chacune des fonctions indique une classe de la classification. ϕ_k est attachée à $C_k^\ell, 1 \leq k \leq (n-l)$.

Commençons par montrer que ϕ_k est bien solution de l'équation factorielle (8) qui, nous l'avons vu, se réduit à celle (9). Le premier membre de cette équation

$$\sum_{i' \in I} s(i', i) * \phi_k(i'), \tag{11}$$

est, ne l'oublions pas, la transcription de

$$\bigvee_{1 \leq i' \leq n} (s(i', i) \wedge \phi_k(i')), \tag{11'}$$

Considérons deux cas :

- (a) $i = i_k \in C_k^\ell$
- (b) $i = i_h \in C_h^\ell$, où $h \neq k$

Dans le premier cas, (a), la valeur maximum d'un même argument $s(i', i_k) * \phi_k(i')$, est obtenue pour $i' = i_k$, où elle vaut exactement λ_ρ qui est la valeur de $\phi_k(i_k)$. On a donc bien

$$(\forall i_k \in C_k^\ell), \quad \sum_{i' \in I} s(i', i_k) * \phi_k(i') = \phi_k(i_k)$$

Dans le deuxième cas (b), l'expression (11) peut se mettre sous la forme

$$\sum_{i' \in C_k^\ell} s(i', i_h) * \phi_k(i') \oplus \sum_{j \neq k, i' \in C_j^\ell} s(i', i_h) * \phi_k(i'),$$

Or

$$\phi_k(i') = \lambda_\rho \quad \text{pour } i' \in C_k^\ell \tag{12}$$

et

$$\phi_k(i') = s^*(i_j, i_k) = s^*(i', i_k) \quad \text{pour } i' \in C_j^\ell.$$

pour un quelconque sommet i_k de C_k^ℓ .

D'autre part, par construction, en se donnant un tel sommet i_k ,

$$\lambda_\rho \leq s^*(i', i_k), \quad \text{pour } i' \in C_k^\ell, \tag{13}$$

de sorte que l'expression (12) peut se mettre sous la forme

$$\sum_{1 \leq i' \leq n} s(i_h, i') * s^*(i', i_k), \tag{14}$$

où i_k est un élément quelconque fixé dans C_k^ℓ .

Mais (14) définit le terme de la i_h -ème ligne et de la i_k -ème colonne

de la matrice $S*S^*$. Comme $S*S^*=S^*$ (cf. formule (6) §3 ci-dessus, l'expression (14) est égale à

$$s^*(i_h, i_k) = \phi_k(i_h), \quad (15)$$

pour $h \neq k$ (cf. (10)).

On a donc bien aussi dans le deuxième cas (b), le résultat.

Soit à présent ϕ une fonction quelconque de $\phi(\lambda_\rho)$; on a

$$(\forall k=1, \dots, (n-1)), \phi \geq \phi(i_k) * \phi_k, \quad (16)$$

Cette relation est trivialement vérifiée pour $i_k \in C_k^\rho$. Considérons alors $i_h \in C_h^\rho$, pour $h \neq k$. Il s'agit de prouver

$$(\forall i_h \in C_h^\rho, \phi(i_h) \geq \phi(i_k) * \phi_k(i_h), \quad (16')$$

relation qui se met sous la forme

$$(\forall i_h \in C_h^\rho), \phi(i_h) \geq s^*(i_h, i_k) * \phi(i_k), \quad (17)$$

Or, par définition

$$\phi(i_h) = \sum_{i' \in I} s(i_h, i') * \phi(i'), \quad (18)$$

Soit $(i'_0 = i_h, i'_1, i'_2, \dots, i'_{(j+1)} = i_k)$ la chaîne reliant i_h à i_k qui réalise

$$s^*(i_h, i_k) = s(i'_0, i'_1) * s(i'_1, i'_2) * \dots * s(i'_j, i'_{(j+1)}), \quad (19)$$

La relation (18) entraîne la suite des inégalités

$$\phi(i'_m) \geq s(i'_m, i'_{(m+1)}) \wedge \phi(i'_{(m+1)}), \quad (20)$$

pour $0 \leq m \leq j$.

Ce qui, par transitivité, implique l'inégalité (17) et donc celle, générale, (16).

Cette relation (16) permet d'écrire

$$\phi \geq \sum_{1 \leq k \leq (n-1)} \phi(i_k) * \phi_k, \quad (21)$$

Mais on a

$$\phi(i_r) \leq \phi(i_r) * \phi_r(i_r)$$

puisque

$$\phi(i_r) \leq \lambda(r) = \phi_r(i_r)$$

Par conséquent,

$$(\forall r=1, \dots, (n-1)) (\forall i_r \in C_r^\rho) \phi(i_r) \leq \sum_{1 \leq k \leq (n-1)} \phi(i_k) * \phi_k(i_r) \quad (22)$$

qui est le reflet de l'inégalité contraire de celle (21) ; donc

$$\phi = \sum_{1 \leq k \leq (n-1)} \gamma_k \phi_k, \quad (23)$$

où pour tout $k=1,2,\dots,(n-l), \gamma_k = \phi(i_k)$.

Ainsi, toute fonction solution de l'équation factorielle est combinaison linéaire du système de facteurs $\{\phi_k / 1 \leq k \leq (n-l)\}$. Il reste à montrer qu'il s'agit du seul système libre. Supposons en effet que l'on puisse avoir, pour k fixé,

$$\phi_k = \sum_{1 \leq h \leq m} \beta_h \chi_h, \tag{24}$$

où, pour tout $h, \chi_h \in \phi(\lambda_\ell)$.

Comme, pour $i_k \in C_k^\ell$,

$$\phi_k(i_k) = \lambda_\ell \geq \chi_h(i), \text{ pour tout } i;$$

il existe nécessairement un indice h_0 , pour lequel

$$\chi_{h_0}(i_k) = \lambda_\ell \text{ avec } \beta_{h_0} \geq \lambda_\ell.$$

de sorte que

$$\beta_{h_0} * \chi_{h_0} = \chi_{h_0};$$

la relation (24) montre alors que

$$\chi_{h_0} \leq \phi_k \tag{25}$$

D'autre part, la relation (16) donne

$$\chi_{h_0} \geq \lambda_\ell * \phi_k = \phi_k, \tag{26}$$

D'où nécessairement,

$$\chi_{h_0} = \phi_k$$

C.Q.F.D.

Nous avons ainsi repris de façon plus naturelle la démonstration du théorème de M. Gondran, en nous dégageant de l'accessoire que constitue l'"arbre de longueur minimale" dont l'intérêt est, comme nous l'avons signalé ci-dessus, essentiellement algorithmique.

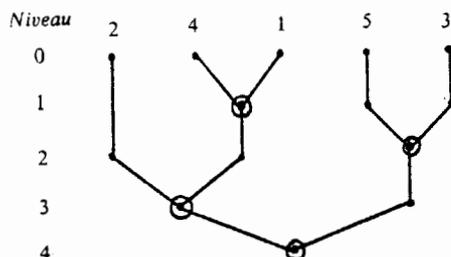
B. PRESENTATION FACTORIELLE DE LA RECHERCHE D'UN ARBRE BINAIRE DES CLASSIFICATIONS.

B.1. Critère de l'inertie expliquée

B.1.1. Recherche d'un système de facteurs et d'axes factoriels associés.

I - DEFINITIONS ET PROPRIETES PRELIMINAIRES LIEES A L'ARBRE DES CLASSIFICATIONS

$I = \{1, 2, \dots, n\}$ désigne un ensemble de cardinal n . Un arbre binaire des classifications sur I dont la suite des niveaux est indexée par $0, 1, \dots, (n-1)$; peut être présenté comme la suite de ses noeuds $(I_1, I_2, \dots, I_j, \dots, I_{(n-1)})$ respectivement formés aux niveaux $1, 2, \dots, (n-1)$. Ainsi, l'arbre binaire suivant sur $I = \{1, 2, 3, 4, 5\}$



sera représenté par la suite des parties

$$(\{1,4\}, \{3,5\}, \{1,2,4\}, \{1,2,3,4,5\})$$

De façon générale, un noeud définira une *partie maximale au niveau j*, dans l'un des deux cas suivants

- a) la partie contient un seul élément extérieur à chacune des parties I_1, I_2, \dots, I_j ; de I .
- b) il s'agit de l'un des sous-ensembles $I_h, 1 \leq h \leq j$; tel qu'aucune autre partie de I , parmi $I_1, I_2, \dots, I_{(j-1)}$, ne l'inclue proprement.

Dans ces conditions, on se rend compte que la suite des noeuds $(I_1, I_2, \dots, I_{(n-1)}=I)$ d'un arbre binaire des classifications est telle que le noeud $I_j (1 \leq j \leq (n-1))$ se déduit de la réunion de deux parties maximales au niveau j .

Réciproquement, soit une suite de parties $(I_1, I_2, \dots, I_{(n-1)}=I)$ de I telle que pour tout $j=1, 2, \dots, (n-1)$, I_j est formée de la réunion de deux parties maximales au niveau j . Une telle suite définit nécessairement un arbre binaire des classifications sur I . Les deux parties maximales dont I_j est la réunion seront respectivement notées $a(j)$ et $b(j)$.

Nous revenons ici dans le cadre de la classification d'un nuage de points $\mathcal{N}(I) = \{(M_i, \mu_i) / i \in I\}$ situé dans un espace euclidien muni d'une métrique q . On a une décomposition du moment total d'inertie du nuage

$$\mathcal{M} = \sum_{i \in I} \mu_i \|M_i - G\|^2, \quad (1)$$

selon les différents noeuds d'un arbre, en l'occurrence binaire des classifications :

$$\mathcal{M} = \sum_{1 \leq j \leq (n-1)} \lambda_j, \quad (2)$$

où λ_j est la différence entre le moment d'inertie de la nouvelle classe I_j et de ses deux sous-classes composantes

$$\lambda_j = \mathcal{M}(I_j) - [\mathcal{M}(a(j)) + \mathcal{M}(b(j))]; \quad (3)$$

il s'agit de la part d'inertie perdue en remplaçant I_j par ses deux classes composantes $a(j)$ et $b(j)$. On a

$$\lambda_j = \frac{v[a(j)]v[b(j)] \|G[a(j)] - G[b(j)]\|^2}{v(I_j)} \tag{4}$$

où $v[a(j)]$, $v[b(j)]$ et $v(I_j) = v[a(j)] + v[b(j)]$

désignent respectivement les masses de $a(j)$, $b(j)$ et $I(j)$; où, d'autre part $G[a(j)]$ et $G[b(j)]$ sont les centres de gravité des classes $a(j)$ et $b(j)$.

Nous supposons à présent que l'arbre binaire des classifications sur I que nous allons devoir bientôt interpréter de façon "factorielle", est bâti de manière ascendante en partant de la partition "discrète" où chaque classe contient exactement un objet, pour aboutir à la partition "grossière" à une classe I . A chaque pas de la construction de l'arbre on réunit la paire de classes, supposée unique, qui minimise le critère (3) à partir de sa forme (4).

En notant la quantité (3) ou (4) sous la forme $\Delta[a(j), b(j)]$; on a

$$\Delta(I_j, I_h) = \frac{1}{v(I_j + v(I_h))} \{ [v(a(j)) + v(I_h)] \Delta[a(j), I_h] + [v(b(j)) + v(I_h)] \Delta[b(j), I_h] - v(I_h) \Delta[a(j), b(j)] \} \tag{5}$$

pour toute partie maximale I_h au niveau j .

Cette formule permet de montrer que la suite des valeurs $(\lambda_1, \lambda_2, \dots, \lambda_{(n-1)})$ de (4) attachée à la suite croissante des niveaux de l'arbre, est *strictement croissante*.

II - BASE ORTHONORMALE DE FONCTIONS SUR I .

Commençons par signaler que la forme de l'équation factorielle retenue pour interpréter l'arbre binaire des classifications sur I , construit en optimisant à chaque pas, la part locale affectée du critère de l'inertie expliquée, est la même que celle du paragraphe A.1.2. ; soit

$$\sum_{i' \in I} \mu_{i'} q(i', i) \phi(i') = \lambda \phi(i) \tag{6}$$

avec les notations adoptées dans ce paragraphe.

Nous allons, de la même manière, que dans [1], construire dans un cadre ici plus général, une base orthonormale $(\phi^1, \phi^2, \dots, \phi^r, \dots, \phi^{(n-1)})$ dans l'espace des fonctions à moyenne nulle sur I . La fonction ϕ^r associée au noeud I_r de la suite des noeuds $(I_1, \dots, I_r, \dots, I_{(n-1)})$ représentant l'arbre binaire, est définie comme suit

$$\phi^r(i) = \begin{cases} \alpha_r & \text{si } i \in a(r) \\ \beta_r & \text{si } i \in b(r) \\ 0 & \text{si } i \notin I_r \end{cases} \tag{7}$$

$$\text{où} \quad \begin{cases} \alpha_r \vee [a_r] + \beta_r \vee [b_r] = 0 \\ \alpha_r^2 \vee [a_r] + \beta_r^2 \vee [b_r] = 1 \end{cases}, \quad (8)$$

En convenant de $\alpha_r > 0$ et $\beta_r < 0$, on obtient

$$\begin{cases} \alpha_r = [\vee(b_r) / \vee(a_r) \vee(I_r)]^{1/2} \\ \beta_r = -[\vee(a_r) / \vee(b_r) \vee(I_r)]^{1/2} \end{cases}, \quad (9)$$

Il est alors facile de vérifier que le système des fonctions ϕ^r , $1 \leq r \leq (n-1)$, est bien orthonormal ; c'est-à-dire,

$$(\forall 1 \leq r, s \leq (n-1), \sum_{i \in I} \mu_i \phi^r(i) \phi^s(i) = \delta_{rs}), \quad (10)$$

où $\delta_{rs} = 1$ (resp. 0) selon que $r=s$ (resp. $r \neq s$).

L'espace des fonctions de moyenne nulle sur I étant de dimension $(n-1)$; $(\phi^1, \phi^2, \dots, \phi^{(n-1)})$ constitue une base orthonormale de cet espace.

III - PRESENTATION FACTORIELLE DE L'ARBRE DES CLASSIFICATIONS.

3.1. Métrique sur I associée à la base $(\phi^1, \dots, \phi^{(n-1)})$.

Relativement à la suite des fonctions $(\phi^1, \dots, \phi^{(n-1)})$ que nous venons de définir ci-dessus (formules (7) et (9)), considérons la matrice des proximités suivantes entre éléments de I

$$\{Q(i, i') / (i, i') \in I \times I\}, \quad (11)$$

où

$$\forall (i, i') \in I^2, Q(i, i') = \sum_{1 \leq r \leq (n-1)} \lambda(r) \phi^r(i) \phi^r(i'), \quad (12)$$

l'étendue de la somme précédente peut en fait se réduire à $n(i, i') \leq r \leq (n-1)$

où $n(i, i')$ est le plus petit indice du noeud incluant les deux sommets i et i' .

A partir de l'extension linéaire de ϕ^r , $1 \leq r \leq (n-1)$, à l'espace engendré par $\{(M_i - G) / i \in I\}$, on peut étendre linéairement Q par rapport à l'un de ses arguments. Rappelons ici que pour alléger les formules précédentes ((6), (10), (11) et 12)) on a noté i pour $(M_i - G)$.

L'extension linéaire de ϕ^r , $1 \leq r \leq (n-1)$, sera définie par la formule

$$\phi^r \left(\sum_{i \in I} p_i i \right) = \sum_{i \in I_r} p_i \phi^r(i) = p(a_r) \alpha_r + p(b_r) \beta_r, \quad (13)$$

où $\{p_i / i \in I\}$ est une pondération sur I à coefficients réels, où

$$p(a_r) = \sum_{i \in a_r} p_i, \quad p(b_r) = \sum_{i \in b_r} p_i ;$$

où enfin α_r et β_r ont été définis dans les formules (9).

Avec la formule (13), en considérant une masse totale 1, on vérifie :
 $\phi^r(\sum_{i \in I} \mu_i i) = \phi^r(G) = 0$, pour tout $r=1, \dots, (n-1)$.

Avec une telle extension, on a

$$Q(\sum_{i \in I} p_i i, i') = \sum_{1 \leq r \leq (n-1)} \lambda(r) p(a_r) \alpha_r \phi^r(i') + \sum_{1 \leq r \leq (n-1)} \lambda(r) p(b_r) \beta_r \phi^r(i') =$$

$$= \sum_{i \in I} p_i Q(i, i'), \tag{14}$$

3.2. THEOREME. La suite des fonctions $(\phi^1, \phi^2, \dots, \phi^{(n-1)})$ définit la suite des facteurs normalisés de moyenne nulle sur I, de la matrice des proximités sur I définie par (11). La valeur propre associée à ϕ^r est $\lambda(r)$, $r=1, \dots, (n-1)$. L'axe factoriel unitaire associé à ϕ^r au sens de la métrique Q, est défini par $[G(a_r) - G(b_r)] / \|G(a_r) - G(b_r)\|$, $r=1, \dots, (n-1)$, où la norme $\|\cdot\|$ reste relative à q.

L'équation factorielle associée à Q se met sous la forme

$$\sum_{i \in I} \mu_i Q(i, i') \phi(i') = \lambda \phi(i), \tag{15}$$

pour tout $i \in I$.

En remplaçant $Q(i, i')$ par le second membre de (12) et en considérant pour ϕ la fonction ϕ^s , le premier membre de (15) devient

$$\sum_{i' \in I} \mu_{i'} [\sum_{1 \leq r \leq (n-1)} \lambda(r) \phi^r(i) \phi^r(i')] \phi^s(i'), \tag{16}$$

En intervertissant à présent les deux signes sommes, l'expression précédente devient

$$\sum_{1 \leq r \leq (n-1)} \lambda(r) \phi^r(i) [\sum_{i' \in I} \mu_{i'} \phi^r(i') \phi^s(i')], \tag{17}$$

Compte tenu à présent du caractère orthonormal de la famille de fonctions $\{\phi^r / 1 \leq r \leq (n-1)\}$, la quantité entre crochets n'est différente de 0 que si $s=r$ où elle vaut 1. L'expression (17) se réduit par conséquent à $\lambda(s) \phi^s(i)$; ce qui prouve que la fonction ϕ^s est facteur, solution de (15), relativement à la valeur propre $\lambda(s)$. $s=1, 2, \dots, (n-1)$.

Reprenons le premier membre de (15) où ϕ est remplacé par ϕ^s ; soit

$$\sum_{i' \in I} \mu_{i'} Q(i', i) \phi^s(i'), \tag{18}$$

compte tenu de la définition de ϕ^s (cf. (7), (8) et (9)), l'expression précédente se réduit à

$$\alpha_s \sum_{i' \in a_s} \mu_{i'} Q(i', i) + \beta_s \sum_{i' \in b_s} \mu_{i'} Q(i', i), \quad (19)$$

en tenant compte à présent de l'extension linéaire de Q par rapport à l'un de ses arguments (cf. § 3.1 précédent) et des valeurs de α_s et de β_s , on obtient pour (19)

$$\left[\frac{v(a_s)v(b_s)}{v(I_s)} \right]^{1/2} Q[G(a_s) - G(b_s), i] \quad (20)$$

Considérons à présent le second membre de l'équation factorielle (15) avec $\phi = \phi^s$ et $\lambda = \lambda_s$ donné par la formule (4) ci-dessus. On obtient alors la formule classique en analyse factorielle.

$$\phi^s = \frac{1}{\sqrt{\lambda_s}} Q(A_s), \quad (21)$$

où

$$A_s = [G(a_s) - G(b_s)] / \|G(a_s) - G(b_s)\|, \quad (22)$$

est l'axe factoriel unitaire au sens de q , associé au facteur ϕ^s au sens de Q .

On a ainsi démontré les relations

$$Q(A_s, i) = \begin{cases} [v(b_s)/v(I_s)] \cdot \|G(a_s) - G(b_s)\| & \text{pour } i \in a_s, \\ -[v(a_s)/v(I_s)] \cdot \|G(a_s) - G(b_s)\| & \text{pour } i \in b_s, \\ 0 & \text{pour } i \notin I_s. \end{cases} \quad (23)$$

A partir de là, on peut vérifier la Q -orthonormalité du système d'axes $\{A_s / 1 \leq s \leq (n-1)\}$.

C.Q.F.D.

IV - RECONSTITUTION DU TABLEAU DES DONNEES

4.1. Cas général.

Considérons la situation du paragraphe 3 de A.1.1. où le nuage $\hat{\mathcal{N}}(I)$ est associé à la description d'un ensemble fini I d'individus au moyen d'un ensemble fini V de variables descriptives numériques. La reconstitution de la suite des composantes $(x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{im})$ du point M_i sera définie par la suite des composantes du point \hat{M}_i défini par

$$\hat{M}_i = G + \sum_{1 \leq r \leq (n-1)} Q(M_i - G, A_r) A_r, \quad (24)$$

Le nuage des points

$$\hat{\mathcal{N}}(I) = \{(\hat{M}_i, \mu_i) / i \in I\}, \quad (25)$$

a bien pour centre de gravité le point G , centre de gravité du nuage $\hat{\mathcal{N}}(I)$. D'autre part, on peut vérifier que par rapport à la métrique Q , le système

des facteurs de $\mathcal{N}(I)$ est bien défini par la suite de fonctions $(\phi^1, \phi^2, \dots, \phi^{(n-1)})$ précisé ci-dessus (cf. formules (7) et (9)).

On reconstitue ainsi la j -ème composante orthogonale x_{ij} du point M_i au moyen de la j -ème composante orthogonale \hat{x}_{ij} de \hat{M}_i , définie par

$$\hat{x}_{ij} = g_j + \sum_{1 \leq r \leq (n-1)} \sqrt{\lambda_r} \phi^r(i) \sigma(A_r, e_j), \quad (26)$$

où g_j (resp. $\sigma(A_r, e_j)$) désigne la j -ème composante du centre de gravité G (resp. de l'axe factoriel A_r).

4.2. Cas d'un tableau de contingence.

Dans le cas de l'analyse d'un nuage $\mathcal{N}(I)$ associé à l'ensemble des lignes d'un tableau de contingence $I \times J$ (cf. § 5 de A.1.1.), la formule (26) précédente devient

$$\hat{f}_j^i = p_{.j} + \sum_{1 \leq r \leq (n-1)} \sqrt{\lambda_r} \phi^r(i) \sigma(A_r, e_j), \quad (27)$$

où \hat{f}_j^i reconstitue f_j^i et où

$$A_r = (g_J^{a_r} - g_J^{b_r}) / \|g_J^{a_r} - g_J^{b_r}\|$$

avec

$$g_J^{a_r} = \frac{1}{p(a_r)} \sum_{i \in a_r} p_i \cdot f_J^i, \quad (28)$$

(revoir les notations du paragraphe 5 de A.1.1.)

A partir de la formule (27) on obtient

$$\hat{f}_{ij} = p_i \cdot p_{.j} \{ 1 + \sum_{1 \leq r \leq (n-1)} \sqrt{\lambda_r} \phi^r(i) q(A_r, e_j) \}, \quad (29)$$

où q est la métrique dont on munit l'espace R^J (cf. formule (43) § 5 de A.1.1)

Puisque le point $g_j = (p_{.1}, \dots, p_{.j}, \dots, p_{.m})$ est centre de gravité du nuage

$$\hat{\mathcal{N}}(I) = \{ (\hat{f}_J^i, p_i) / i \in I \},$$

on a bien

$$\sum_{i \in I} \hat{f}_{ij} = p_{.j}, \quad (30)$$

mais on vérifie également à partir de la formule (29) qu'on a

$$\sum_{j \in J} \hat{f}_{ij} = p_i.$$

4.3. Cas d'une correspondance symétrique.

Une correspondance symétrique sur $I \times I$ est définie par un système de nombres positifs de somme 1 :

$$\{f_{ii'} / (i, i') \in I \times I\}, \quad (31)$$

tel que

$$(\forall (i, i') \in I \times I), f_{ii'} = f_{i'i}, \quad (32)$$

Il est très rare qu'un tel système (31) de nombres représente celui des fréquences défini par une vraie table de contingence. Le plus souvent, le tableau carré (31) des nombres $f_{ii'}$, se trouve défini à partir des formules

$$(\forall (i, i') \in I \times I), f_{ii'} = \sum_{j \in J} (f_{ij} f_{i'j} / p_{.j}), \quad (33)$$

où

$$f_{I \times J} = \{f_{ij} / (i, j) \in I \times J\}, \quad (34)$$

est la distribution empirique de probabilité sur $I \times J$ associée à un véritable tableau de contingence $I \times J$ (cf. §5 de A.1.1).

Or l'analyse factorielle du tableau (34), orientée vers la définition des facteurs, est équivalente à celle du tableau symétrique (31). Pour reconstituer les nombres de ce dernier conformément aux facteurs

$\phi^1, \phi^2, \dots, \phi^{(n-1)}$, précédemment définis et associés à l'arbre binaire des classifications sur I (cf. § 1 ci-dessus) ; on "oubliera" la représentation vectorielle de I à travers J et on utilisera directement la formule de reconstitution

$$\hat{f}_{ii'} = p_{i.} p_{.i'} \left\{ 1 + \sum_{1 \leq r \leq (n-1)} \sqrt{\lambda_r} \phi^r(i) \phi^r(i') \right\}, \quad (35)$$

pour tout couple $(i, i') \in I \times I$.

On peut alors vérifier que la correspondance (35) admet comme facteurs relativement aux valeurs propres $\lambda_1, \lambda_2, \dots, \lambda_{(n-1)}$, la suite des fonctions $\phi^1, \phi^2, \dots, \phi^{(n-1)}$; pour une représentation de I à travers I dans R^I , muni cette fois-ci de la métrique habituelle du χ^2 .

C'est cette formule (35), directement proposée dans [1] pour une correspondance symétrique (cf. (31) et (32)) qui a été le point de départ de la recherche des sections A.1.2. et B.1.1. Par rapport à cette formule, nous avons commencé par nous situer dans le cadre général d'un nuage de points dans un espace euclidien. Nous avons considéré l'arbre binaire des classifications (cf. §1 ci-dessus) (resp. la classification (cf. § 1 de A.1.2.)) qui s'impose dans cette présentation factorielle ; c'est-à-dire, dont la construction est basée sur le critère de l'inertie expliquée. Nous avons en conséquence précisé les valeurs propres (cf. formules (4) du paragraphe 1 ci-dessus (resp. formules (7) du paragraphe 1 de A.1.2.)) et donné la correspondance entre facteur et axe factoriel associé. Nous avons enfin établi une reconstitution conforme du tableau des données.

B.2. Critère du "saut minimal" ; recherche d'un système de facteurs.

I - INTRODUCTION

Le cadre est ici le même que celui du paragraphe A.2.1. Dans ce dernier on a interprété factoriellement l'ensemble des classes de la partition définie à un niveau donné de l'arbre binaire des classifications bâti pas à pas sur le critère du "saut minimal". Cette forme de l'interprétation nous paraît plus naturellement adaptée pour la présentation factorielle de la suite des noeuds $(I_1, I_2, \dots, I_1, \dots, I_{(n-1)})$ de l'arbre binaire des classifications (cf. §1 de B.1.1.). On attachera à chaque noeud I_ℓ , $1 \leq \ell \leq (n-1)$, un facteur Ψ_ℓ qui l'indiquera, relativement à la valeur propre λ_ℓ (cf. formule (2) § 2 de A.2).

Commençons par rappeler la forme de l'équation factorielle

$$(\forall i \in I), \sum_{i' \in I} s(i', i) * \Psi(i') = \lambda \Psi(i), \quad (1)$$

La propriété 1 du paragraphe 3 de A.2 reste trivialement vraie ; de sorte que l'ensemble $\bar{\Psi}(\lambda)$ des solutions de l'équation (1), apparaît comme l'ensemble des fonctions sur I à valeurs dans l'intervalle $[0, \lambda]$, répondant à l'équation

$$(\forall i \in I), \sum_{i' \in I} s(i', i) * \Psi(i') = \Psi(i), \quad (2)$$

D'autre part, la propriété 2 (§ 3 de A.2) est vraie sur $C_1^\ell = I_\ell$; elle est donc vraie sur chaque partie I_ℓ de I, définie par un noeud de l'arbre ; $1 \leq \ell \leq (n-1)$.

II - LE SYSTEME DE FACTEURS

A chaque partie I_ℓ de I définie par un noeud de l'arbre, associons la fonction Ψ_ℓ sur I définie comme suit

$$\begin{aligned} (\forall i_1 \in I_\ell), \Psi_\ell(i_1) &= \lambda_\ell \text{ (cf. formule (2), § 2 de A.2).} \\ (\forall i \in I_\ell), \Psi_\ell(i) &= s^*(i, i_1), \end{aligned} \quad (3)$$

où $s^*(i, i_1)$ est la proximité ultramétrique entre le point i et l'un des sommets i_1 de I_1 (voir ici les lemmes 1 et 2, §2 de A.2).

L'ensemble des parties maximales au niveau $(\ell+1)$ est précisément l'ensemble des classes $\{C_1^\ell, C_2^\ell, \dots, C_{(n-\ell)}^\ell\}$ de la partition de niveau k définie dans la section A2 (cf. §2), où on avait posé $I_\ell = C_1^\ell$. Si $i \in C_j^\ell$, pour $j > 1$, $s^*(i, i_1) = s^*(i_j, i_1)$ quel que soit le couple de sommets (i_j, i_1) de $C_j^\ell \times C_1^\ell$.

THEOREME. La suite des fonctions qu'on vient de définir $(\Psi_1, \Psi_2, \dots, \Psi_\ell, \dots, \Psi_{(n-1)})$ constitue un système unique de facteurs indépendants où le facteur Ψ_ℓ est relatif à la valeur propre λ_ℓ .

La démonstration est analogue à celle du théorème correspondant du paragraphe 4 de la section A.2. On peut remarquer que la fonction Ψ_ℓ est identique à celle ϕ_ℓ associée au niveau ℓ de l'arbre (cf. formule (10) § 4 de A.2.). Par conséquent Ψ_ℓ est facteur relativement à la valeur propre λ_ℓ ; $1 \leq \ell \leq (n-1)$.

D'autre part, pour la raison qu'on vient d'invoquer, si Ψ est facteur relativement à la valeur propre λ_ℓ , on a

$$\Psi \geq \gamma_\ell * \Psi_\ell, \quad (4)$$

où γ_ℓ est la valeur commune de Ψ sur I_ℓ . Cette relation est le reflet de la relation (16) (§ 4 de A.2) pour $k=1$. D'autre part,

$$(\forall i_1 \in I_\ell), \Psi(i_1) \leq \gamma_\ell * \Psi_\ell(i_1) \text{ puisque } \Psi_\ell(i_1) = \lambda_\ell \geq \Psi(i_1) = \gamma_\ell ;$$

et

$$(\forall i \notin I_\ell), \Psi(i) \leq \gamma_\ell * \Psi_\ell(i) = \Psi(i_1) * s^*(i, i_1),$$

où i_1 est un point quelconque de I_ℓ et où $s^*(i, i_1)$ est la proximité ultramétrique de i à i_1 ; en vertu de la formule (2) ci-dessus. Donc

$$\Psi \leq \gamma_\ell * \Psi_\ell, \quad (5)$$

finalement

$$\Psi = \gamma_\ell * \Psi_\ell, \quad (6)$$

ainsi Ψ est "colinéaire" au facteur Ψ_ℓ

Les valeurs λ_ℓ ($1 \leq \ell \leq (n-1)$) étant fixées ; il nous reste à prouver qu'il s'agit du seul système libre de facteurs. Supposons en effet qu'il existe un autre système libre de facteurs respectivement associés au λ_ℓ ,

$$\{\omega_k / 1 \leq k \leq (n-1)\}, \quad (7)$$

et que l'on puisse avoir

$$\Psi_\ell = \sum_{1 \leq k \leq (n-1)} \alpha_k \omega_k, \quad (8)$$

Puisque $\Psi_\ell(i) < \lambda_\ell$, pour tout i ; la relation (8) se réduit nécessairement à

$$\Psi_\ell = \sum_{1 \leq k \leq (n-1)} \beta_k \omega_k, \quad (9)$$

Soit $i_1 \in I_\ell$; comme $\Psi_\ell(i_1) = \lambda_\ell \geq \omega_k(i_1)$, pour tout i de I et tout k compris entre ℓ et $(n-\ell)$; il existe nécessairement un indice k_0 tel que

$$\omega_{k_0}(i_1) = \lambda_\ell \quad \text{avec } \beta_{k_0} \geq \lambda_\ell$$

De sorte que

$$\beta_{k_0} * \omega_{k_0} = \omega_{k_0}$$

D'ailleurs, k_0 est nécessairement égal à ℓ ; car pour $k > 1$, on a

$$(\forall i \in I), \omega_k(i) < \lambda_\ell$$

Ainsi, en vertu de (9),

$$\omega_\ell \leq \Psi_\ell, \quad (10)$$

Mais, nous avons montré ci-dessus (cf. formule (4), la relation

$$\omega_\ell \geq \omega_\ell(i_1) * \Psi_\ell = \Psi_\ell, \quad (11)$$

car $\Psi_\ell(i) \leq \lambda_\ell = \omega_\ell(i_1)$, pour tout i de I .

Donc finalement,

$$\omega_\ell = \Psi_\ell, \quad (12)$$

pour tout $\ell = 1, \dots, (n-1)$.

C.Q.F.D.

C. CONCLUSION

Au terme de notre étude, nous allons chercher à répondre brièvement à la question : "Quel est l'intérêt de ces différents types de rapprochement entre la Classification et l'Analyse Factorielle"?

Nous avons montré avec insistance que ces rapprochements correspondent tous à des présentations factorielles du résultat de la Classification "non hiérarchique" (section A) ou bien "hiérarchique" (section B).

L'intérêt concret de telles présentations est de "comprendre" la nature de l'interprétation du résultat d'une classification par rapport à celle d'une analyse factorielle et d'avoir ainsi un élément de réponse relativement à la structure la plus adaptée pour représenter synthétiquement la table de données ; s'agit-il d'une structure en classes de proximité? ou bien d'une structure par axes en composantes? (cf. § 1.1.1, surtout).

Ces présentations ont d'autre part le mérite d'"expliquer" aux praticiens des méthodes des deux formes d'approche de l'analyse des données (i.e. Analyse Factorielle et Classification) pourquoi il arrive qu'une région du plan factoriel soit dominée par une classe (cf. § A.1.2.) ; pourquoi une opposition entre deux parties du plan factoriel, reflète une opposition entre deux noeuds de l'arbre des classifications (cf. § B.1. et [6]). Ces analogies ont d'autant plus de sens que le même critère est en jeu (cf. § A.1 et B.1). C'est au paragraphe A.1.1 que la liaison est faite de la façon la plus intime puisque les deux côtés du tableau de données sont le plus explicitement présents dans le rapprochement entre les deux types de structure de synthèse.

Un intérêt majeur de ces présentations factorielles de la classification est théorique. Ces dernières permettent des rapprochements formels insoupçonnés (cf. § A.2 et § B.2, surtout) entre les deux types de techniques en analyse des données. Ces interprétations factorielles séduiront certai-

nement les tenants de l'approche géométrique dans l'analyse des données. Toutefois, ils doivent se garder de croire que de cette façon ils saisiront le propre de la Classification qui reste essentiellement algébrique, combinatoire et statistique.

En effet, alors que pour l'analyse factorielle, l'équation a un rôle *opérationnel* puisqu'elle conduit à la solution de la représentation condensée ; ici, l'équation factorielle ne fait qu'"habiller" une classification ou un arbre des classifications obtenus par un algorithme tout à fait indépendant qui optimise un critère de forme donnée. D'ailleurs, deux critères seulement ont été ici envisagés ; ce sont les plus classiques. Mais, il en existe d'autres, d'importance. On peut signaler par exemple celui basé sur la quantité d'information mutuelle pour l'analyse d'un tableau de contingence. Indiquons en ce qui nous concerne, que l'indice de proximité entre classes, permettant de construire l'arbre "détaillé" des classifications de notre méthode de classification hiérarchique, ne se réduit à aucun des critères présentés. Il s'agit en effet de la vraisemblance d'un lien qui a la même forme que (1) (§ 2 de A.2) ; mais où les similarités $s(i,j)$ se reflètent à une échelle $[0,1]$ de probabilité où la notion de "ressemblance" est mesurée par une notion de "vraisemblance".

Pour terminer, nous dirons, compte tenu de tout ce qui précède : "Une équation factorielle ne permettra jamais de construire une classification".

BIBLIOGRAPHIE

- [1] J.B. BENZECRI ; *"Analyse de correspondance et hiérarchie indicée"* dans *"L'analyse des données"* chap. II B n°11, Dunod, Paris, 1973.
- [2] M. GONDRAN, *"Valeurs propres et vecteurs propres en classification hiérarchique"*, R.A.I.R.O. série R.
- [3] J.C. GOWER and G.J.S. ROSS, *"Minimu Spanning trees and single linkage cluster analysis"*, Appl. Stat. 18, n°1, 54-64.
- [4] F.R. HELMERT, *"Die Genauigkeit der formel von Peters zur berechnung des wahrscheinlichen beobachtungs fehlers director beobachtungen glicher genauigkeit"*. Astron. Nachr. Vol. 88, 112-131.
- [5] N.HOWARD, *"Least squares classification and principal component analysis : a comparison"* in *"Quantitative ecological analysis in teh social sciences"* (M. DOGAN & S.ROKKAN, eds) Cambridge : M.I.T. Press, (1969).
- [6] M. JAMBU, *"Programme de calcul des contributions mutuelles entre classes d'une hiérarchie et facteurs d'une correspondance"*, Les cahiers de l'analyse des données, Vol. 1-1976-n°1, Paris.
- [7] N. JARDINE and R.SIBSON, *"Mathematical taxonomy"*, Wiley, New York, (1971).
- [8] S.C. JONHSON, *"Hiérarchical clustering shemes"*, Psychometrika 32, n°3, (1967).
- [9] J.B. KRUSKAL, *"On the shortest spanning subtree of a graph and the traveling salesman problem"*, Proc. Amer. Math. Soc. n°7, (1956).

- [10] I.C. LERMAN, "*Les bases de la classification automatique*", Gauthier-Villars, Paris, (1970).
- [11] I.C. LERMAN, " *Les présentations factorielles de la classification*"; R.A.I.R.O. vol. 13 n°2 et n°3, 1979.
- [12] M.H. NICOLAU, "*Analyse d'un algorithme de classification*", thèse de 3ème cycle, Univ. Paris VI, I.S.U.P., Nov. 1972.
- [13] M. ROUX, "*Un algorithme pour construire une hiérarchie particulière*", thèse de 3ème cycle, Univ. Paris VI, L.S.M., 1968.