RAIRO. Recherche opérationnelle

I. C. Lerman

Les présentations factorielles de la classification. I

RAIRO. Recherche opérationnelle, tome 13, n° 2 (1979), p. 107-128

<http://www.numdam.org/item?id=RO_1979__13_2_107_0>

R.A.I.R.O. Recherche opérationnelle/Operations Research
(vol. 13, n° 2, mai 1979, p. 107 à 128)

LES PRÉSENTATIONS FACTORIELLES DE LA CLASSIFICATION. I (*)

par I. C. LERMAN (1)

Résumé. — *Différentes tentatives de rapprochement entre l'Analyse Factorielle et la Classification ont été proposées. On montre que toutes ces tentatives correspondent en fait à des présentations « factorielles » du problème de la Classification. La tentation 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 travail 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 par des résultats nouveaux à chacune des approches en les situant, dans une synthèse, les unes par rapport aux autres. Ce qui nous permettra de nous rendre compte de l'intérêt relatif de ces différentes tentatives.

Abstract. — *Attempts have been made to connect Factor Analysis and Classification. We show that all these attempts propose, in fact, to interpret the Classification problem in terms of Factor Analysis. This kind of interpretation is rather tempting because of the historically established character of the Principal Component Analysis that has the advantage to provide an optimal solution for the retained variance criterion.*

The purpose of this paper is to analyse each of these attempts in order to find out their real nature and to adapt them to the new situations. Thus the new results will be contributed to each of theses approaches by locating their relative positions. This allows us to realize the relative interest of these attempts.

Différentes tentatives de rapprochement entre l'Analyse Factorielle et la Classification ont été proposées : N. Howard (1969), J. P. Benzécri (1971), M. Gondran (1975), M. Jambu suivant J. P. Benzécri (1976). Toutes ces tentatives correspondent en fait, comme nous le verrons ci-dessous à des présentations « factorielles » du problème de la Classification. La tentation 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 cet article 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

(*) Reçu avril 1978.

(1) Laboratoire de Statistique, I.R.I.S.A., Université de Rennes-I, 35042 Rennes Cedex.

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 part 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 part 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 de l'article, sans doute plus aisée, se fera selon le schéma suivant :

$$A\,11, \quad A\,12 \to B\,1; \qquad A\,2 \to B\,2.$$

A. PRÉSENTATION 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 (mais non encore publié) 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 × 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.

1. INTRODUCTION ET RAPPELS

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

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

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

$$\mathfrak{M}=\sum_{i\in I}\mu_i\|M_i-G\|^2=\sum_{1\leqq h\leqq k}\Big\{\sum_{i\in I_h}\mu_i\|M_i-G_h\|^2\Big\}+\sum_{1\leqq h\leqq k}\nu_h\|G_h-G\|^2.\tag{1}$$

Dans cette formule, dite d'analyse de la variance, de même type que celles attribuées à Huygens; $\{I_h/1\leqq h\leqq k\}$ désigne une partition en k classes de I. A chacune des classes I_k se trouvent associées, d'une part sa masse

$$\nu_h=\sum_{i\in I_h}\mu_i,\tag{2}$$

et d'autre part, son centre de gravité

$$G_h=\frac{1}{\nu_h}\sum_{i\in I_h}\mu_i M_i.\tag{3}$$

On a, bien entendu,

$$\mu=\sum_{1\leqq h\leqq k}\nu_h\qquad\text{et}\qquad G=\frac{1}{\mu}\sum_{1\leqq h\leqq k}\nu_h G_h.\tag{4}$$

Le moment d'inertie de la classe I_h :

$$\mathfrak{M}_h=\sum_{i\in I_h}\mu_i\|M_i-G_h\|^2\tag{5}$$

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

D'autre part,

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

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

$$\mathfrak{M} = \sum_{1 \leqq h \leqq k} (\mathfrak{M}_h + \mathcal{L}_h). \tag{7}$$

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 $\sum \{ \mathfrak{M}_h / 1 \leqq h \leqq k$; c'est-à-dire, en vertu de la relation (7), qui rendent maximale $\sum \{ \mathcal{L}_h / 1 \leqq h \leqq k \}$. 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], chap. 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 $\sum \{ \mathcal{L}_h / 1 \leqq h \leqq k \}$ est la part d'inertie (i. e.; variance) retenue ou expliquée et $\sum \{ \mathfrak{M}_h / 1 \leqq h \leqq k \}$ est la part perdue.

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

PROPRIÉTÉ : Soit $\{ a_r / 1 \leqq r \leqq p \}$ 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 $\{ I_h / 1 \leqq h \leqq k \}$ 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

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

et

$$C=\sum_{r>p}\left\{\sum_{1\leq h\leq k}\nu_h[q(a_r,\,G_h-G)]^2\right\}.\tag{9}$$

En effet, compte tenu de la décomposition

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

on a, en raison des propriétés barycentriques, la relation

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

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

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

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

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

il reste donc à comparer F et C 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 $\mathscr{N}(I)$ 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.

3 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. $J=\{1, 2, \ldots, m\}$ 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 $(1 \leqq j \leqq m)$, la j-ième 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

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

où $\{e_j/1 \leqq j \leqq m\}$ 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-ième individu. De la sorte, le centre de gravité du nuage est défini par

$$(G-O)=\sum_{1 \leqq j \leqq m} g_j e_j \tag{14}$$

où

$$g_j=\frac{1}{\mu} \sum_{1 \leqq i \leqq n} \mu_i x_{ij} \tag{15}$$

Enfin, l'espace R^m est supposé muni d'une métrique q, diagonale 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^n où la suite des coordonnées du j-ième point N_j (représentant la variable v_j) est définie par

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

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

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

où $\{f_i/1 \leqq i \leqq n\}$ 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-ième axe canonique porté par e_j ; soit

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

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

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

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

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

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

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

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 $\{ b_r/1 \leqq r \leqq p \}$ 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

$$\sum_{1 \leqq r \leqq p} \sum_{i \in I} \mu_i [q(b_r, M_i - G)]^2 = \sum_{1 \leqq r \leqq p} \sum_{i \in I} \mu_i [\sum_{1 \leqq j \leqq m} \beta_{rj}(x_{ij}-g_j) q_{jj}]^2, \tag{23}$$

où on a posé

$$b_r = \sum_{1 \leqq j \leqq m} \beta_{rj} e_j. \tag{24}$$

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 $(r=1, \ldots, p)$ sont assujettis à la condition suivante :*

$$(\forall r=1, \ldots, p), \quad (\exists J_r \neq \emptyset); \qquad \beta_{rj} = \begin{cases} 0 & si \quad j \notin J_r, \\ \gamma(r)/\sqrt{q_{jj}} & si \quad j \in J_r, \end{cases} \tag{25}$$

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 $1/\sqrt{q_{jj}}$, $j=1, 2, \ldots, m$; le coefficient de proportionnalité $\gamma(r)$ étant positif.

Le caractère orthogonal de la base $\{ b_r/1 \leqq r \leqq p \}$ impose nécessairement que l'ensemble des parties $\{ J_r/1 \leqq r \leqq p \}$ 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 :

$$\sum_{j \in J_r} \beta_{rj}^2 q_{jj} = 1, \tag{26}$$

qui donne

$$\gamma(r) = 1/\sqrt{m(r)}, \tag{27}$$

où $m(r) = \text{card}(J_r)$.

Quant au critère optimisé

$$\sum_{1 \leq r \leq p} \sum_{1 \leq i \leq n} \mu_i [q(b_r, M_i - G)]^2, \tag{28}$$

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 :

$$\sum_{1 \leq r \leq p} \sum_{i \leq i \leq n} \mu_i \left[\frac{1}{\sqrt{m(r)}} \sum_{j \in J_r} (x_{ij} - g_j) \sqrt{q_{jj}} \right]^2, \tag{29}$$

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

$$(1/q_{jj}) = s_j^2 = \sum_{i \in I} \mu_i (x_{ij} - g_j)^2, \tag{30}$$

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

$$\xi_{ij} = (x_{ij} - \overline{x}_j)/s_j, \qquad i = 1, \ldots, n; \tag{31}$$

centrées réduites de la j-ième 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 $\mathcal{N}(I)$:

$$\sum_{i \in I} \mu_i \sum_{j \in J} \xi_{ij}^2 = \sum_{i \in I} \mu_i \sum_{1 \leq r \leq p} \sum_{j \in J_r} (\xi_{ij} - \overline{\xi}_i(r))^2 + \sum_{i \in I} \mu_i \sum_{1 \leq r \leq p} m(r) [\overline{\xi}_i(r)]^2, \tag{32}$$

où on a noté

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

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 équipondérée; 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 montré, qu'au-delà d'un certain seuil, un accroissement de la variance retenue peut ne pas s'accompagner d'une interprétation plus nuancée.

4. ANALYSE FACTORIELLE CONTRAINTE DE $\mathcal{N}(J)$ ET CLASSIFICATION DE L'ENSEMBLE DES OBJETS

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

$$\sum_{1 \leqq r \leqq p} \sum_{j \in J} [q^*(c_r, N_j - O)]^2, \tag{33}$$

où q^* a été précisé en (17).

L'expression (33) se met analytiquement, sous la forme suivante :

$$\sum_{1 \leqq r \leqq p} \sum_{j \in J} [\sum_{i \in I} \mu_i \gamma'_{ri} (x_{ij} - g_j) \sqrt{q_{jj}}]^2, \tag{34}$$

où on a noté

$$c_r = \sum_{1 \leqq i \leqq n} \gamma'_{ri} \mathfrak{f}_i, \tag{35}$$

$\{\mathfrak{f}_i/1 \leqq i \leqq n\}$ désignant, bien entendu, la base canonique de R^n.

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

$$(\forall r = 1, \ldots, p), \quad (\exists I_r \neq \emptyset); \qquad \gamma'_{ri} = \left\{ \begin{array}{ll} 0 & si \quad i \notin I_r, \\ \delta(r) & si \quad i \in I_r, \end{array} \right\} \tag{36}$$

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 $\{ c_r/1 \leqq r \leqq p \}$ implique que l'ensemble des parties $\{ I_r/1 \leqq r \leqq p \}$ définit une partition de I en p classes non vides; la r-ième 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 $\delta(r)$:

$$q^*(c_r, c_r) = \sum_{i \in I_r} (\delta(r))^2 \mu_i = 1, \tag{37}$$

d'où

$$\delta(r) = 1/\sqrt{\nu(r)} \qquad \text{avec} \quad \nu(r) = \sum_{i \in I_r} \mu_i. \tag{38}$$

Dans ces conditions, l'expression (34) du critère maximisé devient

$$\sum_{1 \leqq r \leqq p} \nu(r) \sum_{1 \leqq j \leqq m} \left[\frac{1}{\nu(r)} \sum_{i \in I_r} \mu_i (x_{ij} - g_j) \right]^2 q_{jj}, \tag{39}$$

laquelle, pouvant manifestement se mettre sous la forme

$$\sum_{1 \leqq r \leqq p} \nu(r) \| G_r - G \|^2 \tag{40}$$

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.

5. CAS SYMÉTRIQUE 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 une 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-ième modalité du premier caractère et la j-ième du second. On substitue à ce tableau de nombres entiers, celui des fréquences relatives $\mathfrak{f}_{ij} = (k_{ij}/k_{..})$, 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 $p_{i.} = \sum \{ \mathfrak{f}_{ij}/j \in J \}$, $i = 1, \ldots, n$, une marge ligne qui contiendra celles $p_{.j} = \sum \{ \mathfrak{f}_{ij}/i \in I \}, j = 1, \ldots, m$. Ces fréquences relatives sont dites marginales; $p_{i.}$ (resp. $p_{.j}$) est une mesure de l'« importance » (numérique) de la i-ième (resp. j-ième) 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 :

$$\mathcal{N}(I) = \{ (\mathfrak{f}_J^i, p_{i.})/i \in I \}, \tag{41}$$

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

$$\mathfrak{f}_J^i = (\mathfrak{f}_1^i, \ldots, \mathfrak{f}_j^i, \ldots, \mathfrak{f}_m^i), \tag{42}$$

dont la suite des coordonnées est la suite des proportions conditionnelles $\mathfrak{f}_j^i = (\mathfrak{f}_{ij}/p_{i.})$: proportion, dans la i-ième classe définie par la première variable partition, des individus possédant la j-ième 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 $\mathfrak{f}_J^i$ 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 :

$$q(e_j, e_h) = \left\{ \begin{array}{ll} 0 & \text{si} \quad h \neq j, \\ 1/p_{.j} & \text{si} \quad h = j, \end{array} \right\} \tag{43}$$

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 la 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 R^m, pour un seul point i de masse $p_{i.} = p_{i'.} + p_{i''.}$. D'autre part, avec une telle métrique (43) 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 $\{\mathfrak{f}_j^{i_1}/j \in J\}$ et $\{\mathfrak{f}_j^{i_2}/j \in J\}$. 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

$$\mathcal{N}(J) = \{(\mathfrak{f}_I^j, p_{.j})/j \in J\}, \tag{44}$$

où $\mathfrak{f}_I^i$ est le point de R^n :

$$(\mathfrak{f}_1^j, \ldots, \mathfrak{f}_i^j, \ldots, \mathfrak{f}_n^j), \tag{45}$$

dont la suite des coordonnées est celle des proportions conditionnelles $\mathfrak{f}_i^j = (\mathfrak{f}_{ij}/p_{.j})$.

Conformément à (43), l'espace ambiant R^n du nuage $\mathcal{N}(J)$, dont la base canonique est notée $\{\mathfrak{f}_i/1 \leq i \leq n\}$, sera muni de la métrique q suivante :

$$q^*(\underline{\mathfrak{f}}_i, \underline{\mathfrak{f}}_{i'}) = \left\{ \begin{array}{ll} 0 & \text{si} \quad i' \neq i, \\ 1/p_{i.} & \text{si} \quad i' = i, \end{array} \right\} \tag{46}$$

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

$$\sum_{j \in J} p_{.j} \| \mathfrak{f}_I^j - g_I \|^2 = \sum_{1 \leq r \leq p} [\sum_{j \in J_r} p_{.j} \| \mathfrak{f}_I^j - g_I^r \|^2] + \sum_{1 \leq r \leq p} p(J_r) \| g_I^r - g_I \|^2, \tag{47}$$

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

$$p(J_r) = \sum_{j \in J_r} p_{.j}, \tag{48}$$

son centre de gravité est le point

$$g_I^r = \frac{1}{p(J_r)} \sum_{j \in J_r} p_{.j} \mathfrak{f}_I^j. \tag{49}$$

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

$$(\mathfrak{f}_1^{J_r}, \ldots, \mathfrak{f}_i^{J_r}, \ldots, \mathfrak{f}_n^{J_r}),$$

où

$$(\forall i \in I), \quad \mathfrak{f}_i^{J_r} = \mathfrak{f}(i, J_r)/p(J_r) = (\sum_{j \in J_r} \mathfrak{f}_{ij})/(\sum_{j \in J_r} p_{.j}), \tag{50}$$

qui représente une proportion relative.

Le centre de gravité g_I du nuage total $\mathcal{N}(J)$ est le point dont la i-ième coordonnée peut se mettre sous la forme (50) :

$$p_{i.} = (\sum_{j \in J} \mathfrak{f}_{ij})/(\sum_{j \in J} p_{.j}).$$

Le critère maximisé, seconde somme du second membre de (47), peut être explicité comme suit :

$$\sum_{1 \leqq r \leqq p} p(J_h) \left\{ \sum_{i \in I} \frac{1}{p_{i.}} (\mathfrak{f}_i^{J_h} - p_{i.})^2 \right\}. \tag{51}$$

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 $\{a_r/1 \leqq r \leqq p\}$, est

$$\sum_{1 \leqq r \leqq p} \{\sum_{i \in I} p_{i.} [q(a_r, \mathfrak{f}_J^i - g_J)]^2\}. \tag{52}$$

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

$$(\forall r = 1, \ldots, p), \quad a_r = \sum_{1 \leqq j \leqq m} \alpha_{rj} e_j, \tag{53}$$

la condition de normalisation s'écrit :

$$(\forall r=1, \ldots, p), \quad \sum_{1 \leqq j \leqq m} (\alpha_{rj}^2/p_{.j})=1 \quad , \tag{54}$$

et on a

$$q(a_r, f_J^i - g_J) = \sum_{1 \leqq j \leqq m} \frac{\alpha_{rj}}{p_{.j}} (f_j^i - p_{.j}), \tag{55}$$

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

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

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

La condition de normalisation (54) permet de préciser $\delta(r)$:

$$(\forall r=1, \ldots, p), \quad \delta(r)=1/\sqrt{p(J_r)}, \tag{56}$$

de sorte que le critère maximisé (52) se met sous la forme

$$\sum_{1 \leqq n \leqq p} \sum_{i \in I} p_{i.} \left[\sum_{j \in J_r} \frac{(f_j^i - p_{.j})}{\sqrt{p(J_r)}} \right]^2 = \sum_{1 \leqq n \leqq p} \sum_{i \in I} \left[\frac{f(i, J_r) - p_{i.}\, p(J_r)}{\sqrt{p_{i.}\, p(J_r)}} \right]^2 \tag{57}$$

et l'expression (51) peut aisément se réduire à cette dernière forme; d'où le théorème :

THÉORÈME : *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 :

$$\sum_{i' \in I} \mu_{i'}\, q(i', i)\, \varphi(i') = \lambda \varphi(i); \tag{1}$$

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

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

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

$$(\forall r = 1, \ldots, p), \quad \nu(r) = \sum_{i \in I_r} \mu_i,$$

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

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

$$(I_1, I_1 UI_2, I_1 UI_2 UI_3, \ldots, I_1 UI_2 U \ldots UIp), \tag{5}$$

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

$$(\forall r = 1, \ldots, p), \quad N(r) = \sum_{s \leq r} \nu(s)$$

$$H_r = \sum_{i \in I_1 U \ldots UI_r} \frac{\mu_i}{N(r)} M_i = \sum_{s \leq r} \frac{\nu(s)}{N(r)} G_s.$$

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

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

qui représente la part d'inertie perdue dans la réunion de I_{r+1} à $I_1 UI_2 U \ldots UI_r$; $r=1, \ldots, p-1$.

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 $\{1, 2, \ldots, r, \ldots, p\}$ muni de la mesure $\{\nu(r)/1 \leqq r \leqq p\}$; P aura pour rôle dans la suite de coder les classes de la partition $\{I_r/1 \leqq r \leqq p\}$.

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

$$\langle \varphi, \psi \rangle = \sum_{1 \leqq r \leqq p} \nu(r)\,\varphi(r)\,\psi(r), \tag{8}$$

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

L'espace des fonctions de moyenne nulle sur P :

$$\{\varphi \in R^p / \sum_{1 \leqq r \leqq p} \nu(r)\,\varphi(r) = 0\}, \tag{9}$$

est de dimension $(p-1)$. Une base orthonormale peut en être fournie par le système de fonctions de Helmert (*cf.* [4]) qu'on notera $(\varphi^0, \varphi^1, \ldots, \varphi^{(p-1)})$, 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 $\varphi^1(2) > 0$;

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

. . .

φ^r prend des valeurs non nulles sur $\{1, 2, \ldots, r+1\}$ et $\varphi^r(r+1) > 0$;

. . .

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

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

$$\varphi^r(s) = -\sqrt{\nu(r+1)}/\sqrt{N(r)\,N(r+1)} \qquad \text{pour} \quad 1 \leqq s \leqq r,$$
$$\varphi^r(r+1) = \sqrt{N(r)}/\sqrt{\nu(r+1)\,N(r+1)},$$
$$\varphi^r(s) = 0 \qquad \text{pour} \quad s > r+1; \tag{10}$$

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

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

Étendons naturellement la définition de chacune des fonctions de la suite $(\varphi^r/1 \leqq r \leqq p)$ à I en posant

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

c'est-à-dire,

$$(\forall s = 1, \ldots, p), \quad (\forall i \in Is), \qquad \varphi^r(i) = \varphi^r(s). \tag{11}$$

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

3. *Présentation factorielle de la classification*

Relativement à la suite des fonctions $(\varphi^1, \varphi^2, \ldots, \varphi^{(p-1)})$ 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

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

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

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

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

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

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

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

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

$$Q\left[\sum_{1 \leqq r \leqq p} \beta_r (G_r - G)\right]=\sum_{1 \leqq r \leqq p} \beta_r Q(G_r - G). \tag{17}$$

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

$$(M-G)=\sum_{i \in I} \alpha_i (M_i - G),$$

où pour tout i, α_i est un nombre réel,

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

où

$$\beta_r=\sum_{i \in I_r} \alpha_i. \tag{18}$$

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

$$\left(\frac{(G_{r+1} - H_r)}{\| G_{r+1} - H_r \|}\right) \qquad [r=1, \ldots, (p-1)],$$

où la norme $\| \, . \, \|$ *reste relative à q.*

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

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

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

$$\sum_{i' \in I} \mu_{i'} \Big[\sum_{1 \leqq r \leqq (p-1)} \varphi^r(i) \varphi^r(i') \Big] \varphi^s(i'). \tag{20}$$

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

$$\sum_{1 \leqq r \leqq (p-1)} \lambda(r) \varphi^r(i) \Big[\sum_{i' \in I} \mu_{i'} \varphi^r(i') \varphi^s(i') \Big]. \tag{21}$$

Compte tenu du caractère orthonormal de la famille de fonctions $\{ \varphi^r / 1 \leqq r \leqq (p-1) \}$; 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 $\lambda(s) \varphi^s(i)$; 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

$$\sum_{i' \in I} \mu_{i'} Q(i', i) \varphi^s(i'), \tag{22}$$

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

$$\sum_{1 \leqq r \leqq p} \Big\{ \sum_{i \in I_r} \mu_{i'} Q(i', i) \Big\} \varphi^s(r), \tag{23}$$

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

$$\sum_{1 \leqq r \leqq p} \nu(r) Q(G_r, i) \varphi^s(r) \tag{24}$$

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

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

où on a noté

$$M(s, s+1) = N(s) \nu(s+1) / N(s+1), \tag{26}$$

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

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

on a, en effet

$$\varphi^s=\frac{1}{\sqrt{\lambda(s)}}Q(A_{s+1}), \tag{28}$$

et on a du même coup démontré les relations

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

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

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

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

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

pour $j<h$ et

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

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

C.Q.F.D.

BIBLIOGRAPHIE

1. J. B. BENZÉCRI, *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 et G. J. S. ROSS, *Minimum Spanning Trees and Single Linkage Cluster Analysis*, Appl. Stat., vol. 18, n° 1, p. 54-64.
4. F. R. HELMERT, *Die Genauigkeit der formel von Peters zur berechung des wahrscheinlichen beobachtungs fehlers director beobachtungen glicher genauigkeit*, Astron. Nachr., vol. 88, p. 112-131.
5. N. HOWARD, *Least Squares Classification and Principal Component Analysis: a Comparison*, in Quantitative Ecological Analysis in the Social Sciences, M. DOGAN et S. ROKKAN, ed., 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, n° 1, 1976, Paris.
7. N. JARDINE et R. SIBSON, *Mathematical Taxonomy*, Wiley, New York, 1971.
8. S. C. JONHSON, *Hierarchical Clustering Shemes*, Psychometrika, vol. 32, n° 3, 1967.
9. J. B. KRUSKAL, *On the Shortest Spanning Subtree of a Graph and the Travelling 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, *Reconnaissance et classification des structures finies en Analyse des données*; rapport n° 70, I.R.I.S.A., Rennes, 1977.
12. M. H. NICOLAU, *Analyse d'un algorithme de classification*, Thèse de 3e cycle, Université Paris-VI, I.S.U.P., novembre 1972.
13. M. ROUX, *Un algorithme pour construire une hiérarchie particulière*, Thèse de 3e cycle, Université Paris-VI, L.S.M., 1968.