

QUELQUES METHODES D'ANALYSE FACTORIELLE D'UNE SERIE DE TABLEAUX DE DONNEES

P.CAZES
LISE-CEREMADE
UNIVERSITE PARIS-DAUPHINE
Place du M^{al} de LATTRE DE Tassigny
75775, Paris, Cedex 16

SOMMAIRE

I INTRODUCTION

II ANALYSE FACTORIELLE D'UN ENSEMBLE DE TABLEAUX DEFINIS SUR LE MÊME ENSEMBLE D'INDIVIDUS

II 1 NOTATIONS

II 2 PRINCIPE DES METHODES FACTORIELLES D'UN ENSEMBLE DE TABLEAUX DEFINIS SUR LE MÊME ENSEMBLE D'INDIVIDUS

II 3 REPRESENTATIONS DES INDIVIDUS ASSOCIEES A CHAQUE

TABLEAU X_k

II 4 REPRESENTATION DES TABLEAUX X_k

III ANALYSE FACTORIELLE D'UN ENSEMBLE DE TABLEAUX DEFINIS SUR LE MÊME ENSEMBLE DE VARIABLES

III 1 NOTATIONS

III 2 PRINCIPALES METHODES

III 3 REPRESENTATIONS DES VARIABLES ASSOCIEES A CHAQUE TABLEAU X_k

III 4 REPRESENTATION DES TABLEAUX X_k A PARTIR DE STATIS DUAL ET DE L'AFM DUALE

IV ANALYSE FACTORIELLE D'UN ENSEMBLE DE TABLEAUX DEFINIS SUR LE MÊME ENSEMBLE D'INDIVIDUS ET SUR LE MÊME ENSEMBLE DE VARIABLES

V CAS DE L'ANALYSE DES CORRESPONDANCES

V 1 INTRODUCTION-NOTATIONS

V 2 ETUDE DU CAS OU $I_t = I$

V 3 ETUDE DU CAS OU $I_t = I$ et $J_t = J$

VI CAS D'UN MELANGE DE VARIABLES QUANTITATIVES ET QUALITATIVES

RESUME :

On étudie ici les méthodes d'analyse factorielle d'une série de tableaux de données, chaque tableau étant associé à un ensemble de variables quantitatives mesurées sur un ensemble d'individus.

On examine d'abord le cas où les tableaux sont définis sur le même ensemble d'individus, puis le cas dual où c'est l'ensemble des variables qui est le même, et enfin le cas où tous les tableaux sont de même format.

On adapte ensuite les méthodes définies dans le cas quantitatif, au cas de l'analyse des correspondances, avant de considérer un mélange de deux types tableaux, les tableaux du premier type étant associés à des variables quantitatives, et donc analysables par l'ACP, tandis que les tableaux du second type relèvent de l'analyse des correspondances.

MOTS-CLES :

Analyse Factorielle, Analyse en Composantes Principales, Analyse des Correspondances, Analyse Factorielle Multiple, STATIS, Analyse Interclasses, Analyse Intra-classes, Tableaux Cubiques.

ABSTRACT :

We study here the methods of factorial analysis of a set of tables, each table being associated to a set of quantitative variables measured on a set of observations.

We consider at first the case where the tables are defined on the same set of observations, then the case where the set of variables is the same and at last the case where all the tables have the same dimensions.

We adapt the methods defined in the quantitative case, to the case of correspondence analysis before to examine the case of two types of tables, the first type of tables being studied by PCA, and the second type by correspondence analysis

KEY-WORDS :

Factorial Analysis, Principal Component Analysis, Correspondence Analysis, Multiple Factorial Analysis, STATIS, Interclass Analysis, Intraclass Analysis, Cubic Tables.

I INTRODUCTION

On se propose d'étudier les méthodes d'analyse factorielle d'un ensemble de tableaux $X_1, \dots, X_k, \dots, X_r$, chaque tableau X_k étant défini sur le produit $I_k \times J_k$, I_k étant l'ensemble des variables et J_k l'ensemble des individus associés. On envisagera les trois cas suivants qui seront respectivement traités aux §§ II, III, IV :

a) $J_k = J$: tous les tableaux sont définis sur le même ensemble d'individus

b) $I_k = I$: tous les tableaux sont définis sur le même ensemble de variables.

c) $I_k = I, J_k = J$ auquel cas on peut considérer qu'on a un tableau cubique défini sur le produit $I \times J \times K$ (où $K = \{1, 2, \dots, k, \dots, r\}$).

Enfin, au § V, on adapte les méthodes développées au cas de l'analyse des correspondances, tandis qu'au § VI, on envisage le cas où l'on a un mélange de deux types de tableaux, les tableaux du premier type étant des tableaux associés à des variables quantitatives que l'on peut étudier par l'Analyse en Composantes Principales (ACP), tandis que les tableaux du second type sont analysables par l'analyse des correspondances (tableaux de contingence, de fréquences, tableaux disjonctifs complets, etc.).

On ne se place pas ici dans le cadre de la modélisation multi-tableaux, comme c'est le cas dans VIVIEN (2002) et on se restreint à un cadre factoriel et géométrique. D'autres méthodes d'analyse ternaire sont également exposées dans un cadre plus général par D'AMBRA et al. (2001).

Les notations utilisées, qui sont analogues à celles définies dans CAZES (1987) ou CAZES (1999) sont rappelées au fur et à mesure.

Dans toute la suite, on appellera analyse factorielle ou ACP du triplet (X, M, D_p) où X est un tableau croisant p variables avec n individus, M la métrique dans R^p , et D_p la métrique diagonale des poids dans R^n , la recherche des valeurs propres et des vecteurs propres normés des matrices $XD_p X'M$ et $X'MXD_p$ (le prime correspondant, comme c'est classique en statistique au transposé) qui permettent la représentation des colonnes et des lignes du tableau X dans les repères orthonormés correspondants.

Si X est centrée, et M la métrique identité ($M = Id_p$), on obtient l'ACP usuelle sur matrice variance, tandis que si X est centrée réduite, M étant toujours égale à la métrique identité, on obtient l'ACP sur matrice de corrélation ou ACP normée.

II ANALYSE FACTORIELLE D'UN ENSEMBLE DE TABLEAUX DEFINIS SUR LE MÊME ENSEMBLE D'INDIVIDUS

Dans ce cas le tableau X_k est défini sur le produit $I_k \times J$, l'ensemble des individus étant le même pour tous les tableaux.

II 1 NOTATIONS

On désigne par n le cardinal de J et par p_k (qu'on notera aussi p_k) celui de I_k .

On suppose l'espace R^n muni de la métrique diagonale des poids

$D_p = \text{Diag}(p_j \mid j=1, n)$ (p_j étant le poids de l'individu j , ($p_j > 0$,

$\sum \{p_j \mid j=1, n\} = 1$) et l'espace R^{p_k} , de la métrique dont la matrice associée dans la base canonique de R^{p_k} est M_k .

Au tableau X_k , de dimensions $p_k \times n$, on associe :

- dans R^{p_k} , le nuage \mathbf{M}_k , des individus, i.e. le nuage des colonnes du tableau X_k .

- dans R^n , le nuage \mathbf{N}_k , des variables, i.e. le nuage des colonnes du tableau X'_k , transposé de X_k .

-le sous-espace H_k de R^n , engendré par \mathbf{N}_k et le projecteur P_k associé.

-le centre de gravité $\underline{g}_k = X_k D_p \underline{1}_n$ de \mathbf{M}_k , $\underline{1}_n$ étant le vecteur de R^n , dont toutes les composantes sont égales à 1. On supposera (sauf avis contraire) que $\underline{g}_k = \underline{0}$, i.e. que X_k est centré.

-la matrice variance $V_k = X_k D_p X'_k$.

-la matrice de produits scalaires $W_k = X'_k M_k X_k$.

On pose également :

- $p = \sum \{p_k \mid k=1, r\}$: nombre total de variables.

- $X' = (X'_1, \dots, X'_k, \dots, X'_r)$: tableau global, juxtaposition des X'_k .

- $M = \text{Diag}(M_k \mid k=1, r)$, métrique bloc diagonale dans R^p dont le $k^{\text{ème}}$ bloc diagonal est égal à M_k .

- $W = X' M X = \sum \{X'_k M_k X_k \mid k=1, r\} = \sum \{W_k \mid k=1, r\}$.

- \mathbf{M} : Nuage des colonnes du tableau X (nuage global des individus).

- $\mathbf{N} = \mathbf{N}_1 \cup \mathbf{N}_2, \dots, \cup \mathbf{N}_k, \dots, \cup \mathbf{N}_r$: nuage des colonnes du tableau X' (i.e. nuage de toutes les variables).

- $H = L(H_1, \dots, H_k, \dots, H_r)$, le sous-espace vectoriel de R^n , engendré par \mathbf{N} (i.e. par toutes les variables).

II -2 PRINCIPE DES METHODES FACTORIELLES D'UN ENSEMBLE DE TABLEAUX DEFINIS SUR LE MÊME ENSEMBLE D'INDIVIDUS.

II-2-1 INTRODUCTION

Le but de ces méthodes est de construire une base orthonormée adéquate soit du sous-espace H , soit des sous-espaces H_k , soit de l'espace R^p , soit des sous-espaces R^{pk} , pour y avoir une représentation "optimale" des variables et des individus.

La plupart des méthodes d'analyse factorielle d'un ensemble de tableaux

$X_1, \dots, X_k, \dots, X_r$ reviennent du moins au premier pas à faire l'analyse factorielle

(Analyse en Composantes Principales (ACP)) du triplet (X, M, D_p) ou du triplet (Y, M, D_p) , où

$$Y' = (\alpha_1 X'_1, \dots, \alpha_k X'_k, \dots, \alpha_r X'_r) = (Y'_1, \dots, Y'_k, \dots, Y'_r)$$

correspond au tableau X' où chaque tableau X'_k (ou X_k) a été pondéré par un coefficient α_k ($\alpha_k > 0$) et remplacé par le tableau $Y_k = \alpha_k X_k$.

Notons que l'analyse de (Y, M, D_p) est équivalente à celle de (X, M_p, D_p) , M_p étant la métrique bloc diagonale de $k^{\text{ème}}$ bloc diagonal $(\alpha_k)^2 M_k$:

$$M_p = \text{Diag}((\alpha_k)^2 M_k | k=1, r)$$

Les caractéristiques associées au tableau Y seront indicées par Y . Ainsi, \mathbf{M}_Y ,

$W_{Yk} = Y'_k M_k Y_k$, $W_Y = Y' M Y = \sum \{ W_{Yk} | k=1, r \}$ correspondent respectivement à \mathbf{M} , W_k , et W .

Au premier pas, quand on fait l'analyse factorielle du triplet (Y, M, D_p) , on est ramené à chercher la plus grande valeur propre des matrices $Y D_p Y' M$ et $Y' M Y D_p = W_Y D_p$ et à considérer les vecteurs propres normés associés qui correspondent respectivement au premier vecteur axial factoriel et à la première composante principale normée.

II-2-2 PRINCIPALES METHODES

II-2-2-1 Analyse factorielle du triplet (X, M, D_p)

Dans ce cas, $\alpha_1 = \dots = \alpha_k \dots = \alpha_r = 1$.

Si tous les M_k sont égaux à la métrique usuelle Id_{pk} , on obtient l'ACP usuelle sur matrice variance du tableau X .

Si de plus toutes les variables sont réduites ou si pour tout k , $M_k = (\text{Diag}(V_k))^{-1}$, on obtient l'ACP sur matrice de corrélation, ou ACP normée.

Dans ces analyses, on cherche des bases orthonormées de R^p ou plutôt du sous-espace de R^p engendré par \mathbf{M} (les vecteurs axiaux factoriels) et de H (les composantes principales normées).

II-2-2-2 Analyse Canonique Généralisée de CARROLL (ACG)

Il s'agit encore de l'analyse factorielle du triplet (X, M, D_p) , mais avec $M_k = V_k^{-1}$.

Ici les composantes principales sont vecteurs propres de $\Sigma\{P_k \mid k=1, r\}$.

II-2-2-3 Analyse Factorielle Multiple (AFM) d'ESCOFIER-PAGES (1998)

Il s'agit de l'analyse factorielle du triplet (Y, M, D_p) , avec $\alpha_k = (1/\lambda_{1k})^{1/2}$, λ_{1k} désignant la plus grande valeur propre dans l'ACP du triplet (X_k, M_k, D_p) .

II-2-2-4 STATIS (LAVIT, 1988)

Il s'agit toujours de l'analyse factorielle du triplet (Y, M, D_p) , les pondérations α_k étant égales à $\sqrt{u_k}$, u_k étant la $k^{\text{ème}}$ composante du vecteur propre normé \underline{u}

($\underline{u}' \underline{u} = 1$) associé à la plus grande valeur propre de la matrice carrée C d'ordre r , dont le terme général $c_{kk'}$ ($1 \leq k, k' \leq r$) a une des deux expressions suivantes :

$$c_{kk'} = \text{Trace}(W_k D_p W_{k'} D_p) \quad (1)$$

$$c_{kk'} = \text{Trace}(W_k D_p W_{k'} D_p) / (\text{Trace}((W_k D_p)^2) \text{Trace}((W_{k'} D_p)^2))^{1/2} \quad (2)$$

On peut noter que par construction, la matrice C a tous ses termes positifs ; donc d'après le théorème de Frobenius, le vecteur propre de C associé à sa plus grande valeur propre a toutes ses composantes de même signe, signe que l'on peut toujours choisir positif, ce qui légitime l'écriture $\alpha_k = \sqrt{u_k}$.

II-2-2-5 Analyse en Composantes Principales Généralisées (ACPG) de CASIN (1996)

Dans cette analyse qui revient à chaque pas à rechercher le premier facteur d'une ACP, on cherche à construire une base orthonormée de chaque sous-espace H_k (et non plus une base orthonormée de l'espace H engendré par les H_k).

Au premier pas, on recherche le premier facteur dans l'ACP du triplet (X, M, D_p) .

Soit :

ψ_1 la première composante principale de l'ACP précédente (rappelons que ψ_1 n'est pas normée, puisque sa variance est égale à la plus grande valeur propre de WD_p).

$\psi_1^k = P_k \psi_1$ la projection de ψ_1 sur H_k .

$P_{\psi_{1k}}$ le projecteur sur ψ_1^k .

$X_k^{(1)} = (Id_n - P_{\psi_{1k}}) X_k'$, le tableau des résidus des variables de X_k' quand on fait la régression sur ψ_1^k , Id_n étant la matrice identité d'ordre n .

$X^{(1)} = (X_1^{(1)}, \dots, X_k^{(1)}, \dots, X_r^{(1)})$.

Au second pas, on recherche le premier facteur dans l'ACP du triplet $(X^{(1)}, M, D_p)$.

Soit :

ψ_2 la première composante principale de l'ACP précédente.

$\psi_2^k = P_k \psi_2$ la projection de ψ_2 sur H_k .

$P_{\psi_{2k}}$ le projecteur sur ψ_2^k .

$X_k^{(2)} = (Id_n - P_{\psi_{2k}}) X_k^{(1)} = (Id_n - P_{\psi_{1k}} - P_{\psi_{2k}}) X_k'$.

puisque par construction ψ_1^k et ψ_2^k sont orthogonaux, et donc $P_{\psi_{1k}} P_{\psi_{2k}} = P_{\psi_{2k}} P_{\psi_{1k}} = 0$.

$X^{(2)} = (X_1^{(2)}, \dots, X_k^{(2)}, \dots, X_r^{(2)})$

On itère la procédure en recherchant la première composante principale ψ_3 et sa projection $\psi_3^k = P_k \psi_3$ sur H_k dans l'ACP du triplet $(X^{(2)}, M, D_p)$, puis on corrige le tableau $X^{(2)}$ pour obtenir le tableau $X^{(3)}$, avec :

$$X^{(3)} = (X_1^{(3)}, \dots, X_k^{(3)}, \dots, X_r^{(3)})$$

$$X_k^{(3)} = (Id_n - P_{\Psi_{3k}}) X_k^{(2)} = (Id_n - P_{\Psi_{1k}} - P_{\Psi_{2k}} - P_{\Psi_{3k}}) X_k^{(2)},$$

$P_{\Psi_{3k}}$ étant le projecteur sur Ψ_{3k} , et l'on continue le processus en faisant l'ACP du triplet $(X^{(3)}, M, D_p)$, dont on ne considèrera que le premier facteur, etc.

Remarques : a) Comme annoncé, les $\{ \Psi_s^k / \| \Psi_s^k \| \mid s=1, \dim(H_k) \}$ constituent une base orthonormée du sous-espace H_k .

b) CASIN (1996) a montré que les Ψ_s étaient non corrélées, ce qui permet après les avoir normés, d'avoir une représentation globale des variables dans le système orthonormé correspondant.

c) Si on a un seul tableau ($r=1$) l'ACPG est équivalente à l'ACP usuelle du fait que $\Psi_1^1 = \Psi_1$ et que les composantes principales sont non corrélées.

d) Si on adopte les métriques $M_k = V_k^{-1}$, on obtient l'analyse discriminante des tableaux évolutifs de CASIN (1995).

II-2-2-6 Analyse de Co-Inertie Multiple (ACOM) de CHESSEL-HANAFI (1996)

Cette méthode est en quelque sorte duale de la précédente, dans la mesure où l'on recherche ici une base orthonormée adéquate de chaque sous-espace $E_k = R^{pk}$ (et non pas de $E = R^p$).

Au premier pas on retient dans l'ACP du triplet (X, M, D_p) , le premier vecteur axial factoriel $\underline{u} = (\underline{u}_1', \dots, \underline{u}_k', \dots, \underline{u}_r')$, \underline{u}_k étant le sous-vecteur de \underline{u} associé au $k^{\text{ème}}$ groupe de variables, et on pose :

$$\underline{v}_k^1 = \underline{u}_k / \| \underline{u}_k \| \text{ sous-vecteur } \underline{u}_k \text{ de } \underline{u} \text{ normalisé .}$$

$P_{v_{k1}}$: projecteur sur l'axe engendré par \underline{v}_k^1 dans R^{pk} .

$X_k^{(1)} = (Id_{pk} - P_{v_{k1}}) X_k$, le tableau associé à la projection du nuage \mathbf{M}_k des individus sur l'hyperplan de R^{pk} orthogonal à \underline{v}_k^1 .

$$X^{(1)} = (X_1^{(1)}, \dots, X_k^{(1)}, \dots, X_r^{(1)}) .$$

Le second pas est identique au premier en remplaçant X par $X^{(1)}$.

Si donc on considère le premier vecteur axial factoriel \underline{u} dans l'ACP du triplet $(X^{(1)}, M, D_p)$ et si \underline{u}_k désigne le sous-vecteur de \underline{u} associé au $k^{\text{ème}}$ groupe de variables, on pose :

$\underline{v}_k^2 = \underline{u}_k / \|\underline{u}_k\|$ sous-vecteur \underline{u}_k de \underline{u} normalisé .

P_{vk2} : projecteur sur l'axe engendré par \underline{v}_k^2 dans R^{pk} .

$$X_k^{(2)} = (\text{Id}_{pk} - P_{vk2}) X_k^{(1)} = (\text{Id}_{pk} - P_{vk1} - P_{vk2}) X_k$$

$$X'^{(2)} = (X'_{1(2)}, \dots, X'_{k(2)}, \dots, X'_{r(2)}) ,$$

et on continue le processus avec le triplet $(X^{(2)}, M, D_p)$.

Remarques :

a) Le nombre maximal de pas possible est égal à $\text{Max } p_k$. Pour un pas $s > p_k$, les vecteurs \underline{u}_k et \underline{v}_k^s associés seront pris égal à $\underline{0}$.

b) Les $\{\underline{v}_k^s \mid s = 1, p_k\}$ constituent une base orthonormée de R^{pk}

c) Les composantes principales $\underline{\psi}^s = X'^{(s-1)} M \underline{u}$ ($= X' M \underline{u}$, compte tenu des contraintes d'orthogonalité) , \underline{u} étant le premier vecteur axial factoriel issu de l'ACP du triplet $(X^{(s-1)}, M, D_p)$ (avec $X^{(0)} = X$) sont non corrélées , ce qui permet , comme dans le cas de l'ACPG, d'avoir une représentation globale des variables dans le système orthonormé associé aux $\underline{\psi}^s$.

d) On peut introduire des coefficients de pondération α_k pour chaque tableau X_k (cf. CHESSEL et HANAFI (1996)). La procédure étudiée s'adapte immédiatement en remplaçant le tableau X par le tableau Y (ou la métrique M par M_p) .

II-3 REPRESENTATIONS DES INDIVIDUS ASSOCIEES A CHAQUE TABLEAU X_k

Les analyses présentées au paragraphe précédent permettent d'avoir une représentation globale des individus. Si on veut une représentation des individus associée à chaque tableau k, on peut opérer de la façon suivante :

On considère le tableau T suivant de dimensions $p \times nr$. T est le tableau bloc diagonal dont le $k^{\text{ème}}$ bloc diagonal est égal à Y_k :

$$T =$$

	1	2		k		r
1	Y_1	0		0		0
2	0	Y_2		0		0
k	0	0		Y_k		0
r	0	0		0		Y_r

et on projette chaque colonne de T sur les axes factoriels déterminés en II-2. On obtient ainsi pour chaque individu j, r représentations $j_1, j_2, \dots, j_k, \dots, j_r$. Pour que la représentation globale de j soit au centre de gravité des représentations partielles j_k ($1 \leq k \leq r$) il est classique dans certaines analyses (par exemple dans l'AFM) de considérer non pas le tableau Y, mais le tableau Y/r . En fait d'un point de vue pratique dans l'AFM, on fait l'ACP du triplet (Y, M, D_p) , mais on fait subir une homothétie de rapport $1/r$ à la représentation associée des individus, pour obtenir le principe du centre de gravité, très utile au niveau de l'interprétation.

II-4 REPRESENTATION DES TABLEAUX X_k

Cette représentation, beaucoup moins naturelle que dans le cas où l'on a des tableaux définis sur le même ensemble de variables (cf. § III) et pas toujours facile à interpréter, a été particulièrement développée dans STATIS et l'AFM.

Pour effectuer cette représentation, on se place dans l'espace $G = \mathbb{R}^N$ (avec $N=n^2$) des opérateurs de la forme $A = WD_p$, cet espace étant muni de la métrique associée au produit scalaire défini par :

$$\langle A, B \rangle = \text{Trace}(AB)$$

II-4-1 CAS DE STATIS

Dans ce cas, on considère dans l'espace précédent le nuage des $\{A_k \mid k=1, r\}$ avec $A_k = W_k D_p$.

Pour représenter les A_k , on effectue alors l'analyse factorielle (non centrée) du nuage précédent, chaque A_k étant muni de la masse unité, ce qui revient à diagonaliser la matrice C définie par (1).

Si \underline{u}^s désigne le $s^{\text{ème}}$ vecteur propre normé de C ($\underline{u}^s \cdot \underline{u}^s = 1$) associé à la valeur propre λ_s , les λ_s étant rangées par valeurs décroissantes, la coordonnée de A_k sur l'axe factoriel s est donnée par $u_k^s \sqrt{\lambda_s}$, u_k^s étant la $k^{\text{ème}}$ composante de \underline{u}^s .

Remarques :

a) l'opérateur $\sum \{u_k^1 A_k \mid k=1, r\}$ est l'opérateur qui maximise $\|B\|^2$ parmi tous les opérateurs de la forme $B = \sum \{u_k A_k \mid k=1, r\}$ avec $\sum \{(u_k)^2 \mid k=1, r\} = 1$.

Rappelons que le vecteur $\underline{\alpha}$ des pondérations adoptées dans STATIS est le vecteur de composantes $\alpha_k = \sqrt{u_k^1}$. L'opérateur précédent est classiquement appelé compromis et la représentation de l'ensemble $K = \{1, 2, \dots, r\}$ des r tableaux correspond avec la terminologie de STATIS à l'interstructure. De

même la diagonalisation du compromis qui correspond à l'ACP du triplet (Y, M, D_p) s'appelle l'intrastructure.

b) On peut aussi considérer le nuage des opérateurs normés $A_k = W_k D_p / \|W_k D_p\|$ ce qui revient à diagonaliser la matrice C définie par (2).

c) Si on munit les A_k de masses m_k , et si D_m désigne la matrice diagonale des m_k , on est amené à diagonaliser la matrice $(D_m)^{-1} C$, les vecteurs propres \underline{u}_s étant alors orthonormés pour D_m .

d) Comme on l'a vu au § II-2-4, le vecteur propre de C associé à sa plus grande valeur propre a toutes ses composantes de même signe. Il en résulte un effet taille qui peut être important avec un taux d'inertie du premier axe qui peut être très élevé (supérieur à 0.9), ce qui limite l'intérêt de la représentation effectuée, comme on l'a déjà dit.

e) Si l'on cherche l'opérateur A tel que $\sum \{ \|A_k - A\|^2 \mid k=1, r \}$ soit minimum, on trouve l'opérateur $A = \sum \{ A_k \mid k=1, r \} / r$ dont la diagonalisation correspond au facteur $1 / \sqrt{r}$ près à l'ACP du triplet (X, M, D_p) .

II-4-2CAS DE L'AFM

Dans ce cas, si $\underline{\psi}_s$ est la $s^{\text{ème}}$ composante principale issue de l'ACP du triplet (Y, M, D_p) , on considère dans l'espace G , les opérateurs de rang 1,

$$B_s = \underline{\psi}_s \underline{\psi}_s' D_p / \lambda_s, \lambda_s \text{ étant la valeur propre associée à } \underline{z}_s.$$

Par construction, les B_s sont orthonormés puisque :

$$\langle B_s, B_t \rangle = \text{Trace} (\underline{\psi}_s \underline{\psi}_s' D_p \underline{\psi}_t \underline{\psi}_t' D_p) / (\lambda_s \lambda_t) \\ = (\underline{\psi}_s' D_p \underline{\psi}_t)^2 / (\lambda_s \lambda_t) = \delta_s^t = 1 \text{ si } s = t \text{ et } 0 \text{ sinon,}$$

puisque les composantes principales sont non corrélées et de variance la valeur propre.

On se sert du système orthonormé des B_s pour représenter les $A_{Yk} = W_{Yk}$, avec l'inconvénient que les B_s ne forment pas une base de G , ni du sous-espace de G engendré par les A_{Yk} . La coordonnée η_k^s de A_{Yk} sur B_s s'écrit alors :

$$\eta_k^s = \text{TRACE} (W_{Yk} D_p \underline{\psi}_s \underline{\psi}_s' D_p / \lambda_s) = \underline{\psi}_s' D_p W_{Yk} D_p \underline{\psi}_s / \lambda_s$$

Notons que : $\sum \{ \eta_k^s \mid k=1, r \} = \underline{\psi}_s' D_p W_Y D_p \underline{\psi}_s / \lambda_s = \lambda_s$

η_k^s peut donc être considérée comme la contribution du tableau Y_k à l'inertie λ_s du $s^{\text{ème}}$ facteur, dans l'ACP du triplet (Y, M, D_p) .

Remarque : On a : $W_Y D_p = \sum \{ \psi_s \psi_s' D_p \mid s=1, t \}$, t étant le rang de W_Y i.e. le nombre de facteurs non triviaux dans l'ACP du tableau Y .

III ANALYSE FACTORIELLE D'UN ENSEMBLE DE TABLEAUX DEFINIS SUR LE MÊME ENSEMBLE DE VARIABLES.

Dans ce cas le tableau X_k est défini sur le produit $I \times J_k$, l'ensemble des variables étant maintenant identique pour tous les tableaux.

III-1 NOTATIONS

On désigne par n_k le cardinal de J_k (qu'on notera aussi n_k) et par p celui de I (i.e. le nombre de variables)

On suppose que chaque l'individu j de J_k est muni de la masse p_j ($p_j > 0$, $\sum \{ p_j \mid j \in J_k \} = 1$) et on désignera par D_{pk} la matrice diagonale des poids associée.

On suppose également qu'une pondération α_k ($\sum \{ \alpha_k \mid k=1, r \} = 1$) est attribuée à chaque tableau X_k

Au tableau X_k , de dimensions $p \times n_k$, on associe :

- dans R^p , le nuage \mathbf{M}_k , des individus, i.e. le nuage des colonnes du tableau X_k .

- dans R^{n_k} , le nuage \mathbf{N}_k , des variables, i.e. le nuage des colonnes du tableau X_k' , transposé de X_k .

-le sous-espace E_k de R^p , engendré par \mathbf{M}_k et le projecteur P_k associé.

-le centre de gravité $\underline{g}_k = X_k D_{pk} \underline{1}_{n_k}$ de \mathbf{M}_k , $\underline{1}_{n_k}$ étant le vecteur de R^{n_k} , dont les n_k composantes sont égales à 1.

-la matrice variance V_k et la matrice de corrélation R_k (si $\underline{g}_k = \underline{0}$, $V_k = X_k D_p X_k'$).

- X_k^- le tableau centré associé à X_k ($X_k^- = X_k$, si $\underline{g}_k = \underline{0}$).

- Z_k le tableau centré réduit associé à X_k .

On pose également :

- $n = \sum \{ n_k \mid k=1, r \}$: nombre total d'individus.

- $X = (X_1, \dots, X_k, \dots, X_r)$: tableau global, juxtaposition des X_k .

- $G = (g_1, \dots, g_k, \dots, g_r)$: tableau $p \times r$ des centres de gravité.

- $D_p = \text{Diag}(\alpha_k D_{p_k} \mid k = 1, r) = \text{Diag}(\alpha_k p_j \mid j \in J_k \mid k = 1, r)$:
 métrique des poids dans R^n qui revient à attribuer le poids $\alpha_k p_j$ à l'individu j de J_k .

- $\underline{g} = G D_\alpha \underline{1}_r = \sum \{ \alpha_k g_k \mid k=1, r \}$, centre de gravité global qu'on
 supposera à l'origine ($\underline{g} = \underline{0}$). Dans l'expression précédente, D_α désigne la
 matrice diagonale des α_k .

- $V = X D_p X'$, matrice variance totale.

- $X^- = (X_1^-, \dots, X_k^-, \dots, X_r^-)$: tableau, juxtaposition des tableaux centrés par
 blocs X_k^- .

- $Z = (Z_1, \dots, Z_k, \dots, Z_r)$: tableau, juxtaposition des tableaux centrés réduits
 par blocs Z_k .

On peut noter que le tableau Z est centré réduit, mais ce n'est pas le tableau
 centré réduit associé à X (sauf si tous les g_k sont nuls auquel cas $X^- = X$) mais
 à X^- .

- $\mathbf{M} = \mathbf{M}_1 \cup \mathbf{M}_2, \dots, \cup \mathbf{M}_k, \dots, \cup \mathbf{M}_r$: nuage des colonnes du tableau X
 (i.e. nuage de tous les individus).

- \mathbf{N} : Nuage des colonnes du tableau X' (nuage global des variables).

III-2 PRINCIPALES METHODES

On peut considérer les analyses suivantes où pour les trois premières, les
 coefficients α_k sont généralement égaux à n_k/n , ce qui n'est pas le cas des deux
 suivantes.

III-2-1 ANALYSE FACTORIELLE DU TRIPLET (X, M, D_p) :

Si M est la métrique usuelle Id_p on obtient l'ACP usuelle du tableau X .

Si de plus toutes les variables sont réduites (globalement) (auquel cas V est
 égale à la matrice de corrélation globale), ou si $M = \text{Diag}(V^{-1})$, on obtient
 l'ACP sur matrice de corrélation, ou ACP normée.

III-2-2 ANALYSE INTERCLASSES :

Il s'agit de l'analyse factorielle du triplet (G, M, D_α) .

Cette analyse est équivalente à celle du triplet $(X - X^-, M, D_p)$, $X - X^-$ correspondant au tableau où chaque individu j de J_k a été remplacé par le centre de gravité de sa classe, à savoir \underline{g}_k .

Si $M = V^{-1}$, on obtient en particulier l'analyse discriminante.

III-2-3 ANALYSE INTRA-CLASSES :

Il s'agit de l'analyse factorielle du triplet (X^-, M, D_p) c'est à dire de l'ACP sur matrice variance ou de l'ACP sur matrice de corrélation de X^- , suivant le choix de la métrique M .

On peut également faire l'analyse factorielle du triplet (Z, M, D_p) qui est aussi une analyse intra classes normée.

Notons que dans l'analyse de (X^-, M, D_p) , si l'on choisit la métrique V^{-1} , on obtient encore l'analyse discriminante.

III-2-4 L'AFM DUALE

Il s'agit toujours de l'analyse factorielle du triplet (X, M, D_p) (ou de (X^-, M, D_p) dans le cas d'une analyse intra classes) où les coefficients α_k sont donnés par :

$$\alpha_k = (1 / \lambda_{k1}) / (\sum \{ (1 / \lambda_{m1}) \mid m = 1, r \})$$

λ_{k1} étant la plus grande valeur propre dans l'ACP de (X_k, M, D_p) (ou de (X_k^-, M, D_p) dans le cas de l'analyse intra classes)

III-2-5 STATIS DUAL

Il s'agit encore de l'analyse factorielle du triplet (X, M, D_p) où les coefficients α_k sont obtenus à partir du vecteur propre normé \underline{u} associé à la plus grande valeur propre de la matrice carrée C d'ordre r , dont le terme général $c_{kk'}$ ($1 \leq k, k' \leq r$) a une des deux expressions suivantes :

$$c_{kk'} = \text{Trace} (A_k A_{k'}) \quad (3)$$

ou

$$c_{kk'} = \text{Trace} (A_k A_{k'}) / (\text{Trace} ((A_k)^2) \text{Trace} ((A_{k'})^2))^{1/2} \quad (4)$$

avec : $A_k = X_k D_{pk} X_k' M$.

Comme dans le cas de STATIS, on peut toujours choisir les composantes u_k de \underline{u} positives, et on a alors :

$$\alpha_k = u_k / (\sum \{ u_m \mid m=1, r \})$$

En remplaçant X_k par X_k^- et X par X^- , on obtient une analyse STATIS duale intra classes. Il faut alors remplacer l'opérateur A_k par $A_k^- = X_k^- D_{pk} X_k^- ' M = V_k M$ dans les formules (3) et (4) et l'opérateur compromis s'écrit alors

$\Sigma \{ \alpha_k A_k^- | k= 1, r \} = (\Sigma \{ \alpha_k V_k | k= 1, r \})M = WM$, W étant la matrice variance intra classes .Si donc $M = V^{-1}$, on obtient une analyse discriminante où les masses des classes sont déterminées par la méthodologie STATIS.

III-2-6 ADAPTATION DE L'ACPG ET DE L'ANALYSE DE CO-INERTIE .

Dans l'espace R^p , le sous-espace vectoriel E_k engendré par M_k est en général identique à R^p , car, en pratique, le nombre d'individus n_k de J_k est supérieur ou très supérieur à p . Quand on fait l'analyse factorielle du triplet (X, M, D_p) , on obtient une base orthonormée de R^p et donc de chaque E_k . Il semble donc inutile d'adapter l'ACPG de CASIN.

Par contre dans l'espace R^n , l'analyse factorielle du triplet (X, M, D_p) , fournit une base orthonormée de R^n , ou plutôt du sous-espace engendré par N et il peut être intéressant si on s'intéresse à J_k et au nuage N_k , d'avoir une base orthonormée de R^{n_k} . On opérera de façon analogue à ce qui a été fait au § II-2-2-6 en intervertissant les rôles de X et X' , de p et de n , et en considérant les composantes principales normées au lieu des vecteurs axiaux factoriels.

III-2-7 REPRESENTATION CONJOINTE DES J_k ET DE K

Dans toutes les analyses précédentes, on peut représenter conjointement l'ensemble total des individus $JK = U \{ J_k | k= 1, r \}$ i.e. le nuage M et l'ensemble K . Il suffit de placer côte à côte les tableaux X (ou X') et G et de projeter en éléments supplémentaires les colonnes du tableau non actif (à savoir G , sauf dans l'analyse interclasses où c'est X qui est passif).

III-3 REPRESENTATIONS DES VARIABLES ASSOCIEES A CHAQUE TABLEAU X_k

Pour obtenir cette représentation, il suffit de considérer le tableau T défini au § II-3, mais où l'on remplace chaque Y_k par X_k (ou X_k^- , dans le cas d'une analyse intra- classes), tableau qui est ici de dimensions $pr \times n$. On projette alors chaque colonne du tableau T' sur les composantes principales normées issues de l'ACP du triplet (X, M, D_p) (ou (X^-, M, D_p) pour une analyse intra classes), ce qui revient à considérer T comme un tableau supplémentaire dans l'analyse précédente.

Dans le cas de l'analyse interclasses, cette façon de procéder revient à rajouter en supplémentaire au tableau G le tableau bloc diagonal de dimensions $pr \times r$ dont la $k^{ième}$ colonne est formée de zéros, sauf le $k^{ième}$ bloc qui est égal à g_k . La représentation associée n'a dans ce cas guère d'intérêt pratique, comme il est facile de le voir.

Si on veut que la représentation globale de chaque variable i soit au centre de gravité de ses représentations partielles i_k pour le tableau k ($1 \leq k \leq r$), plusieurs façons de faire sont envisageables :

On peut opérer comme au § II-3 en effectuant une homothétie de rapport $1/r$ de la représentation globale de chaque variable. Si on veut que i soit au centre des i_k affectés des masses α_k , on peut aussi dans le tableau T du § II-3 remplacer chaque Y_k par X_k / α_k .

Remarque : Si on représente les variables par leurs corrélations avec les composantes principales, i ne peut pas être en général au centre de gravité de ses représentations partielles.

En particulier, si l'on considère l'ACP du triplet (Z, Id_p, D_p) et si r_i désigne la corrélation entre la variable x_i et une composante principale, et r_{ik} la corrélation entre la variable partielle i_k (i.e. variable égale à x_i sur J_k et 0 ailleurs) et cette composante principale, on a :

$$r_i = \frac{\sum \{ (\alpha_k)^{1/2} r_{ik} \mid k = 1, \dots, r \}}{\sum \{ \alpha_k r_{ik} \mid k = 1, \dots, r \}}$$

III-4 REPRESENTATION DES TABLEAUX X_k A PARTIR DE STATIS DUAL ET DE L'AFM DUALE.

On a déjà vu la représentation tout à fait naturelle des tableaux X_k à partir des centres de gravité associés g_k .

On peut aussi envisager les représentations déduites de STATIS DUAL et de l'AFM DUALE, en transposant ce qui a été vu au § II-4.

III-4-1 CAS DE STATIS DUAL

On se place dans l'espace R^P avec $P = p^2$ des opérateurs de la forme $A_k = X_k D_{pk} X_k' M$ ($= V_k M$ si X_k est centré), le produit de 2 opérateurs étant toujours défini par la trace de leur produit.

La représentation du nuage des A_k munis de masses unité s'obtient comme au § II-4-1 à partir des vecteurs propres rangés par valeurs propres décroissantes de la matrice C définie par (3) (ou par (4) si on raisonne sur les opérateurs normés $A_k = X_k D_{pk} X_k' M / || X_k D_{pk} X_k' M ||$), la coordonnée de A_k sur le $s^{\text{ème}}$ axe factoriel étant toujours égale à $u_k^s \sqrt{\lambda_s}$, u_k^s étant la $k^{\text{ème}}$ composante du $s^{\text{ème}}$ vecteur propre normé \underline{u}^s ($\underline{u}^s \underline{u}^s = 1$) de C relatif à la valeur propre λ_s .

III-4-2 CAS DE L'AFM DUALE

Dans ce cas, si \underline{u}_s est le $s^{\text{ème}}$ vecteur axial factoriel issu de l'ACP du triplet

(X, M, D_p) , associé à la valeur propre λ_s , on se sert des opérateurs de rang 1 $B_s = \underline{u}_s \underline{u}_s' M$, qui sont orthonormés, pour représenter les A_k , la coordonnée de A_k sur B_s étant égale à :

$$\eta_k^s = \text{TRACE}(X_k D_{pk} X_k' M \underline{u}_s \underline{u}_s' M) = \underline{u}_s' M X_k D_{pk} X_k' M \underline{u}_s$$

On a la même interprétation de η_k^s qu'au § II-4-2, puisque

$$\sum \{ \alpha_k \eta_k^s \mid k=1, r \} = \underline{u}_s' M X D_p X' M \underline{u}_s = \underline{u}_s' M V M \underline{u}_s = \lambda_s$$

η_k^s peut donc être considéré comme la contribution du tableau X_k (muni de la masse α_k) à l'inertie λ_s du $s^{\text{ème}}$ facteur, dans l'ACP du triplet (X, M, D_p) .

Remarques : 1) On a : $VM = \sum \{ \lambda_s \underline{u}_s \underline{u}_s' M \mid k=1, t \}$, t étant le rang de V i.e. le nombre de facteurs non triviaux dans l'ACP du tableau X .

2) Les techniques précédentes s'appliquent aussi si on considère les X_k^- (auquel cas $A_k = V_k M$) et donc le tableau X^- ou les Z_k (auquel cas $A_k = R_k M$, R_k étant, rappelons le, la matrice de corrélation associée à X_k), et donc le tableau Z .

IV ANALYSE FACTORIELLE D'UN ENSEMBLE DE TABLEAUX DEFINIS SUR LE MÊME ENSEMBLE D'INDIVIDUS ET SUR LE MÊME ENSEMBLE DE VARIABLES.

Dans ce cas, $I_k = I$, $J_k = J$ ($k=1, r$) et on a un tableau ternaire (ou cubique) :

$$X_{IJK} = \{ X_{Ijk} \mid k=1, r \} = \{ X_k \mid k=1, r \}.$$

On garde des notations analogues à ce qui a été vu précédemment. Ainsi, on désigne par :

n , le nombre total d'individus ($n = \text{CARDJ}$).

p , le nombre total de variables ($p = \text{CARDI}$).

M , la métrique dans R^p .

D_p , la métrique diagonale des poids dans R^n .

\mathbf{M}_k , le nuage des colonnes de X_k dans R^p .

\mathbf{N}_k , le nuage des colonnes de X_k' dans R^n .

g_k , le centre de gravité de \mathbf{M}_k .

$G = (g_1, \dots, g_k, \dots, g_r)$ le tableau $p \times r$ des centres de gravité.

$X_C = (X_1, \dots, X_k, \dots, X_r)$: tableau juxtaposition des X_k .

$X'_L = (X'_1, \dots, X'_k, \dots, X'_r)$: X'_L est le tableau superposition des X_k .

On peut alors appliquer les procédures développées aux §§ II (où X est remplacé par X_L) et III (où X est remplacé par X_C), mais on a intérêt à développer une méthode faisant jouer un rôle symétrique aux tableaux X_k , du moins quand le troisième ensemble K est de nature différente des deux premiers I et J , par exemple s'il correspond au temps.

On définit alors le tableau moyen H :

$$H = (1/r) \sum \{ X_k \mid k=1, r \}.$$

Le cas échéant, si des poids α_k de somme 1 sont attribués à chaque tableau X_k , on prendra $H = \sum \{ \alpha_k X_k \mid k=1, r \}$.

On fait alors l'analyse factorielle du triplet (H, M, D_p) et l'on projette dans R^p sur les vecteurs axiaux factoriels associés :

- toutes les colonnes de H .
- toutes les colonnes de X_k (i.e. tous les éléments du nuage \mathbf{M}_k) ($1 \leq k \leq r$).
- toutes les colonnes de G (i.e. tous les centres de gravité g_k).

De même, dans R^n on projette sur les composantes principales normées associées :

- toutes les colonnes de H' .
- toutes les colonnes de X'_k (i.e. tous les éléments du nuage \mathbf{N}_k) ($1 \leq k \leq r$)

Cette façon de procéder revient à effectuer l'ACP du triplet (H, M, D_p) et à mettre les tableaux X_C, X_L, G en tableaux supplémentaires dans l'analyse précédente suivant le schéma de la figure 1.

On obtient ainsi à partir des axes factoriels et des composantes principales normées associées issues de l'ACP du tableau moyen H :

- une représentation moyenne de chaque individu j .
- une représentation partielle j_k de chaque j associé au tableau X_k , la représentation moyenne de j étant au centre de gravité des j_k ($1 \leq k \leq r$).
- une représentation moyenne de chaque tableau k , cette représentation étant au centre de gravité des j_k ($1 \leq j \leq n$).
- une représentation moyenne de chaque variable i .
- une représentation i_k de chaque variable i associé au tableau X_k , la représentation moyenne de i étant au centre de gravité des i_k ($1 \leq k \leq r$).

H	X_1	X_2	X_r	G
X_1				
X_2				
X_r				

Analyse factorielle de H , tous les autres tableaux étant placés en supplémentaires.

Figure 1

V CAS DE L'ANALYSE DES CORRESPONDANCES

V-1 INTRODUCTION - NOTATIONS

Dans ce cas, le $t^{\text{ième}}$ tableau X_t ($1 \leq t \leq r$) est noté, suivant l'usage en analyse des correspondances $k_{I_t J_t}$ ou plus simplement k_t , et on désigne par $T = \{1, 2, \dots, k, \dots, r\}$, l'ensemble des r tableaux (et non plus K , pour éviter toute confusion entre l'indice k et un tableau k_t).

Les ensembles I_t , J_t , jouant des rôles symétriques en analyse des correspondances, le cas $J_t = J$, I_t quelconque se déduit immédiatement du cas $I_t = I$, J_t quelconque.

On ne traitera donc que les deux cas suivants :

a) $I_t = I$, J_t quelconque

b) $I_t = I$, $J_t = J$

Toutes les méthodes décrites aux §§ II à IV peuvent s'adapter de façon plus ou moins évidente, du fait que les métriques en analyse des correspondances et en particulier les métriques des poids sont fonction des marges du tableau étudié, ce qui pose des problèmes quand les tableaux $k_{I_t J_t}$ ont, quand $I_t = I$, des marges sur I différentes ou non proportionnelles (ceci étant encore le cas si les marges sur J diffèrent, ou ne sont pas proportionnelles quand $J_t = J$).

On va traiter ici des méthodes les plus usuelles pour analyser simultanément les tableaux $k_{I_t J_t}$, méthodes obtenues en faisant l'analyse des correspondances des tableaux juxtaposés (cf. § V-2-1) ou superposés (cf. § V-3) ou en effectuant des analyses interclasses (cf. § V-2-2 et § V-3) ou intra-classes (cf. § V-2-3 et § V-3). On considérera aussi des analyses pondérées permettant de généraliser STATIS, et l'AFM (cf. § V-2-4).

On pose :

$k_{I_t} = \{k(i, \cdot, t) \mid i \in I_t\}$: marge sur I_t de $k_{I_t J_t}$:

$k(i, \cdot, t) = \sum \{k(i, j, t) \mid j \in J_t\}$.

$k_{J_t} = \{k(\cdot, j, t) \mid j \in J_t\}$: marge sur J_t de $k_{I_t J_t}$:

$k(\cdot, j, t) = \sum \{k(i, j, t) \mid i \in I_t\}$.

$k(\cdot, \cdot, t) = \sum \{k(i, j, t) \mid i \in I_t, j \in J_t\}$: total du tableau $k_{I_t J_t}$.

$k(\cdot, \cdot, \cdot) = \sum \{k(\cdot, \cdot, t) \mid t \in T\}$: total général.

$IT = \cup \{I_t \mid t \in T\}$: union disjointe des I_t (Si $I_t = I$, IT est l'ensemble I dupliqué r fois, et on a $IT = I \times T$).

$JT = \cup \{J_t \mid t \in T\}$: union disjointe des J_t (Si $J_t = J$, JT est l'ensemble J dupliqué r fois et on a $JT = J \times T$).

On posera encore :

Si $I_t = I$: $k_{I \times T} = \{k_{I_t} \mid t \in T\} = \{k(i, \cdot, t) \mid i \in I, t \in T\}$: ensemble des marges sur I des tableaux $k_{I_t J_t}$.

$k_I = \{k(i, \cdot, \cdot) \mid i \in I\}$: marge sur I du tableau $k_{I \times JT}$, croisant I avec JT :

$$k(i, ., .) = \Sigma \{ \Sigma \{ k(i, j, t) \mid j \in J_t \} \mid t \in T \} = \{ \Sigma \{ k(i, j, t) \mid (j, t) \in JT \} \}$$

Si $J_t = J : k_{J,T} = \{ k_{J_t} \mid t \in T \} = \{ k(., j, t) \mid j \in J, t \in T \}$: ensemble des marges sur J des tableaux k_{IJ_t} .

$k_J = \{ k(., j, .) \mid j \in J \}$: marge sur J du tableau $k_{J \times IT}$, croisant J avec IT :

$$k(., j, .) = \Sigma \{ \Sigma \{ k(i, j, t) \mid i \in I_t \} \mid t \in T \} = \{ \Sigma \{ k(i, j, t) \mid (i, t) \in IT \} \}$$

Remarques : a) Si $I_t = I$ (cf. §§ V-2, V-3) (resp. $J_t = J$, cf. § V-3) on désignera toujours par k_{I_t} (resp. k_{J_t}), la marge sur I (resp. J) du tableau k_{IJ_t} (resp. k_{IJ_t}) pour signifier qu'il s'agit de la marge sur I (resp. J) du $t^{\text{ième}}$ tableau k_t , et ne pas la confondre avec la marge d'ordre 1 k_I (resp. k_J) du tableau $k_{I \times JT}$ (resp. $k_{J \times IT}$).

b) Quand $I_t = I$ (resp. $J_t = J$), le tableau $k_{I \times JT}$ (resp. $k_{J \times IT}$, ou plutôt son transposé $k_{IT \times J}$) est la juxtaposition (horizontale) (resp. superposition (verticale)) des tableaux k_{IJ_t} .

Pour terminer ce paragraphe, rappelons une propriété classique de l'analyse des correspondances, qui nous servira plus loin :

Additionner un ensemble de lignes (resp. de colonnes) en analyse des correspondances revient, du fait des pondérations adoptées dans cette analyse, à remplacer cet ensemble par son centre de gravité.

V-2 ETUDE DU CAS OU $I_t = I$

Dans ce cas, on a une série de tableaux k_{IJ_t} (puisque $I_t = I$), et on peut adopter les méthodes du § III pour représenter I et $JT = \cup \{ J_t \mid t \in T \}$ ainsi que T, et les éléments i associés à chaque tableau k_t , éléments que l'on notera (i, t) ce qui revient à représenter IT.

Outre la juxtaposition des tableaux k_{IJ_t} qui permet de faire les représentations précédentes de façon classique (cf. §V-2-1), on considérera aussi des analyses interclasses (cf. §V-2-2), intra-classes (cf. §V-2-3), ainsi que des analyse des correspondances pondérées, ce qui permettra d'adapter STATIS et l'AFM (cf. § V-2-4).

V-2-1 L'ANALYSE DES CORRESPONDANCES DU TABLEAU $k_{I \times JT}$

Les représentations usuelles dont on vient de parler et qui sont obtenues en adaptant les méthodes des §§ III-2-1, III-2-7 et III-3 reviennent (cf. figure 2) à :

a) faire l'analyse des correspondances du tableau $k_{I \times JT}$ croisant I avec JT.

b) rajouter en éléments supplémentaires dans l'analyse précédente :

b1) le tableau k_{IT} ce qui permet de représenter T en colonnes supplémentaires. On notera t_c la représentation associée d'un élément t de T.

b2) le tableau bloc diagonal $k_{IT \times JT}$ dont le $t^{\text{ème}}$ bloc diagonal est égal à k_t , ce qui permet de représenter IT.

b3) le tableau bloc diagonal $k_{T \times JT}$ dont le $t^{\text{ème}}$ bloc diagonal est égal à k_{It} , ce qui permet de représenter T en lignes supplémentaires. On notera t_i la représentation associée d'un élément t de T .

La représentation t_c de t est alors (d'après la propriété rappelée à la fin du § V-1) au centre de gravité des (j, t) ($j \in J_t$); de même la représentation de i est au centre de gravité des (i, t) ($t \in T$).

Enfin l'ensemble des i ($i \in I$) et l'ensemble des t_i ($t \in T$) ont le même centre de gravité.

V-2-2 ANALYSE INTERCLASSES (ANALYSE DES CORRESPONDANCES DU TABLEAU k_{IT})

Cette analyse (cf. § III-2-2) revient d'après la propriété rappelée à la fin du § V.1 à faire l'analyse des correspondances du tableau k_{IT} , auquel on adjoindra en élément supplémentaire le tableau $k_{I \times JT}$ pour pouvoir représenter JT ; t est toujours au centre de gravité des (j, t) ($j \in J_t$), mais t est ici actif, alors que les (j, t) sont passifs contrairement à ce qui se passait dans l'analyse précédente.

Remarque : D'après le principe d'équivalence distributionnelle, l'analyse des correspondances de k_{IT} revient à faire l'analyse des correspondances du tableau $h_{I \times JT}$ de terme général :

$$h(i, j, t) = k(i, \dots, t)k(\dots, j, t) / k(\dots, \dots, t)$$

k_{IJ1}	k_{IJ2}	k_{IJr}	k_{IT}
k_{IJ1}	0	0	
0	k_{IJ2}	0	
0	0	k_{IJr}	
k_{J1}	0	0	
	k_{J2}	0	
0	0	k_{Jr}	

Analyse des correspondances de $k_{I \times JT}$ avec les tableaux k_{IT} , $k_{IT \times JT}$ et $k_{T \times JT}$ en supplémentaire. Si de plus $J_t = J$, on peut rajouter le tableau marginal d'ordre 2 k_{IJ} en supplémentaire.

On peut aussi analyser k_{IT} avec $k_{I \times JT}$ en supplémentaire (analyse interclasses).

Figure 2

V-2-3 ANALYSE INTRA-CLASSES (ANALYSE DES CORRESPONDANCES DU TABLEAU $l_{I \times J \times T}$)

Cette analyse revient (cf. CAZES-MOREAU (2000)) à faire l'analyse des correspondances du tableau $l_{I \times J \times T}$ dont le terme général est donné par :

$$\begin{aligned} l(i, j, t) &= k(i, j, t) - h(i, j, t) + k(i, \dots)k(\dots, j, t) / k(\dots, \dots) \\ &= k(i, j, t) - k(i, \dots, t)k(\dots, j, t) / k(\dots, \dots, t) + k(i, \dots)k(\dots, j, t) / k(\dots, \dots) \end{aligned}$$

Remarque:

Il est immédiat de vérifier que les tableaux $k_{I \times J \times T}$ et $l_{I \times J \times T}$ ont mêmes marges sur I et JT respectivement.. De plus, la marge sur I du bloc $l_{I \times J \times t}$ est proportionnelle à la marge sur I de $l_{I \times J \times T}$ et donc de $k_{I \times J \times T}$. En effet le terme général de la marge sur I de $l_{I \times J \times t}$ s'écrit :

$$l(i, \dots, t) = \sum \{ l(i, j, t) \mid j \in J_t \} = k(i, \dots)k(\dots, \dots, t) / k(\dots, \dots)$$

V-2-4 ADAPTATION DE STATIS ET DE L'AFM

V-2-4-1 Introduction

Pour faire cette adaptation, on se ramènera aux équivalences existant entre l'analyse des correspondances et l'ACP, puisque STATIS ET l'AFM ont été définies dans le cadre de l'ACP.

Cette adaptation n'est pas immédiate, car

a) Si on multiplie le tableau $k_{I \times J \times t}$ par une constante, l'analyse des correspondances de ce tableau reste inchangée, et en particulier les valeurs propres restent les mêmes, ce qui pose problème au niveau des pondérations dans l'AFM.

b) Dans STATIS comme dans l'AFM on suppose que les métriques associées à l'ensemble commun, ici l'ensemble I, sont invariantes quand on change de sous-tableaux, ce qui n'est généralement pas le cas en analyse des correspondances.

En effet considérant I comme l'ensemble des individus, la métrique des poids dans l'analyse des correspondances du tableau complet $k_{I \times J \times T}$ est associée à la matrice diagonale des $k(i, \dots) / k(\dots, \dots)$, alors que pour le tableau $k_{I \times J \times t}$, elle est associée à la matrice diagonale des $k(i, \dots, t) / k(\dots, \dots, t)$, ces deux métriques n'étant identiques que si $k(i, \dots, t) = k(i, \dots) k(\dots, \dots, t) / k(\dots, \dots)$ = $a_t k(i, \dots)$, (avec $a_t = k(\dots, \dots, t) / k(\dots, \dots)$, ce qui entraîne que $\sum \{ a_t \mid t \in T \} = 1$), ce qui implique la proportionnalité entre $k_{I \times J \times t}$ et $k_{I \times J \times T}$ ($k_{I \times J \times t} = a_t k_{I \times J \times T}$).

Notons que dans l'analyse intra-classes, la propriété précédente est vérifiée pour le tableau $l_{I \times J \times T}$, ce qui permettra dans ce cas d'adapter STATIS, et l'AFM sans difficultés.

V-2-4-2 Rappels sur l'équivalence entre l'analyse des correspondances et l'ACP.

L'analyse des correspondances du tableau $k_{I \times J \times T}$ est équivalente à l'ACP du tableau $X_{I \times J \times T}$ de terme général :

$$X(i, j, t) = [k(i, j, t) - k(i, \cdot, \cdot) k(\cdot, j, t) / k(\cdot, \cdot, \cdot)] / [k(i, \cdot, \cdot) (k(\cdot, j, t) / k(\cdot, \cdot, \cdot))^{1/2}]$$

I correspondant à l'ensemble des individus associé à la métrique des poids $D_p = \text{Diag}(k(i, \cdot, \cdot) / k(\cdot, \cdot, \cdot))$ et JT à l'ensemble des variables associé à la métrique identité.

De même, l'analyse des correspondances du tableau $k_{I \times J \times t}$ est équivalente à l'ACP du tableau $Z_{I \times J \times t}$ de terme général :

$$Z(i, j, t) = [k(i, j, t) - k(i, \cdot, t) k(\cdot, j, t) / k(\cdot, \cdot, t)] / [k(i, \cdot, t) (k(\cdot, j, t) / k(\cdot, \cdot, t))^{1/2}]$$

avec la métrique des poids $D_{pt} = \text{Diag}(k(i, \cdot, t) / k(\cdot, \cdot, t))$ et la métrique usuelle pour l'ensemble J_t des variables .

Dans le cas où $k_{It} = a_t k_I$, on obtient la même métrique des poids ($D_{pt} = D_p$) que dans l'ACP de X, et il est facile de voir que, $X_{I \times J \times t}$ étant le t^{ème} bloc de $X_{I \times J \times T}$, on a :

$$X_{I \times J \times t} = (a_t)^{1/2} Z_{I \times J \times t} \quad (5)$$

Nous nous placerons dans les §§ V-2-4-3 à V-2-4-5 dans ce cadre, auquel cas, dans toutes les ACP considérées, la métrique des poids sera la métrique D_p définie ci-dessus, alors que les métriques associées aux variables seront toujours les métriques usuelles ; puis, nous traiterons au § V-2-4-6 le cas le plus général où les marges sur I de $k_{I \times J \times t}$ ne sont pas proportionnelles à la marge sur I de $k_{I \times J \times T}$, auquel cas la relation (5) n'est plus vérifiée.

V-2-4-3 ACP pondérée et analyse des correspondances pondérée

Dans le cas de STATIS comme de l'AFM, on est amené à faire l'ACP d'un tableau $Y_{I \times J \times T}$ obtenu par pondération de chaque bloc de $X_{I \times J \times T}$.

Si $Y_{I \times J \times t}$ désigne le le t^{ème} bloc de $Y_{I \times J \times T}$ et α_t le coefficient de pondération associé, on a donc :

$$Y_{I \times J \times t} = \alpha_t X_{I \times J \times t}$$

Nous allons voir que l'ACP du tableau $Y_{I \times J \times T}$ est équivalente (à une homothétie près) à l'analyse des correspondances d'un tableau $m_{I \times J \times T}$ déduit de $k_{I \times J \times T}$ par une pondération adéquate des blocs $k_{I \times J \times t}$.

En effet, posons $m_{I \times J \times t} = b_t k_{I \times J \times t}$, et effectuons l'analyse des correspondances du tableau $m_{I \times J \times T}$ juxtaposition des blocs $m_{I \times J \times t}$. Compte tenu (avec des notations évidentes) que $m(i, j, t) = b_t k(i, j, t)$, $m(\cdot, j, t) = b_t k(\cdot, j, t)$,

$$m(i, \cdot, t) = b_t k(i, \cdot, t) = b_t a_t k(i, \cdot, \cdot), \quad m(i, \cdot, \cdot) = (\sum \{ b_t a_t \mid t \in T \}) k(i, \cdot, \cdot),$$

$m(., ., .) = (\sum \{ b_t a_t \mid t \in T \}) k(., ., .)$, l'analyse des correspondances du tableau $m_{I \times J \times T}$ est équivalente à l'ACP du tableau $U_{I \times J \times T}$ (obtenu en remplaçant dans l'expression de X , le tableau $k_{I \times J \times T}$ par $m_{I \times J \times T}$) dont le $t^{\text{ème}}$ bloc est donné par :

$U_{Ijt} = (b_t / (\sum \{ a_t b_t \mid t \in T \}))^{1/2} X_{Ijt} = (a_t b_t / (\sum \{ a_t b_t \mid t \in T \}))^{1/2} Z_{Ijt}$
 Il suffit donc de poser : $b_t = \alpha_t^2$
 pour que l'analyse des correspondances du tableau $m_{I \times J \times T}$ soit équivalente à l'ACP de $Y_{I \times J \times T}$, au facteur $(\sum \{ a_t b_t \mid t \in T \})^{-1/2} = (\sum \{ a_t \alpha_t^2 \mid t \in T \})^{-1/2}$ près.

V-2-4-4 Application à l'AFM

Dans le cas de l'AFM, on est amené à considérer l'ACP du bloc X_{Ijt} ainsi que celle de Z_{Ijt} (puisque cette dernière est équivalente à l'analyse des correspondances du tableau k_{Ijt}). Alors, compte tenu de (5), si λ_{1t} (resp. μ_{1t}) désigne la plus grande valeur propre issue de l'ACP de X_{Ijt} (resp. Z_{Ijt} , i.e. de l'analyse des correspondances de k_{Ijt}), on a :

$$\lambda_{1t} = a_t \mu_{1t} \tag{6}$$

L'AFM du tableau $X_{I \times J \times T}$ revient donc, en posant $\alpha_t = (1 / \lambda_{1t})^{1/2}$, à faire l'ACP du tableau $Y_{I \times J \times T}$ dont le $t^{\text{ème}}$ sous-tableau Y_{Ijt} est donné, compte tenu de (5) et de (6) par :

$$Y_{Ijt} = \alpha_t X_{Ijt} = X_{Ijt} / (a_t \mu_{1t})^{1/2} = Z_{Ijt} / (\mu_{1t})^{1/2}$$

Cette AFM est équivalente d'après les propriétés vues au paragraphe précédent, dont on conservera les notations, et en posant :

$$b_t = \alpha_t^2 = 1/\lambda_{1t} = 1 / (a_t \mu_{1t}) ,$$

à l'analyse des correspondances du tableau pondérée $m_{I \times J \times T}$, au facteur $(\sum \{ a_t b_t \mid t \in T \})^{1/2} = (\sum \{ 1/\mu_{1t} \mid t \in T \})^{1/2}$ près.

V-2-4-5 Application à STATIS

La méthode STATIS (cf. §§ II-2-2-4 et II-4-1) s'adapte immédiatement. Il suffit, si on désigne par X_t le tableau X_{Ijt} et si on pose $W_t = X X_t' D_p$ de diagonaliser la matrice C dont le terme général est donné par (1) ou (2). La représentation de T se fait alors comme indiqué au § II-4-1, tandis que la représentation de JT se fait à partir de l'ACP du tableau pondéré $Y_{I \times J \times T}$, le coefficient de pondération α_t étant égal à $\sqrt{u_t}$, u_t étant la $t^{\text{ème}}$ composante (qu'on peut toujours choisir positive) du vecteur propre normé associé à la plus grande valeur propre de C . Rappelons que l'ACP de $Y_{I \times J \times T}$ est équivalente à l'analyse des correspondances du tableau pondérée $m_{I \times J \times T}$, avec les pondérations $b_t = u_t$.

Remarque :

On aurait pu aussi considérer la variante suivante de STATIS, en considérant au lieu de X_t le tableau Z_{Ijt} , qu'on notera aussi Z_t .

Si $W_{Zt} = Z_t Z_t' D_p$, on a compte tenu de (5), $W_t = a_t W_{Zt}$ et la matrice C_Z obtenue à partir de W_{Zt} (de la même façon que C est déduit des W_t) est telle que son terme général $C_Z(t, t')$ se déduit de $C(t, t')$ par la formule :

$$C_Z(t, t') = C(t, t') / (a_t a_{t'}) \quad (1')$$

dans le cas de la formule (1), et :

$$C_Z(t, t') = C(t, t') \quad (2')$$

dans le cas de la formule (2),

ce qui donne dans le premier cas, une représentation différente de l'ensemble T , et ce qui fournit dans les deux cas, une représentation différente de l'ensemble JT , puisque les pondérations s'appliquent aux tableaux Z_t et non pas X_t .

On peut noter que si les coefficients a_t sont égaux ($a_t = (1/r)^{1/2}$), on a d'après (5),

$X_t = (1/r)^{1/2} Z_t$, et il est indifférent de raisonner sur les X_t ou sur les Z_t .

V-2-4-6 Etude du cas où les marges sur I des tableaux k_{Ijt} ne sont pas proportionnelles

Dans ce cas, les marges k_{It} sont différentes de $a_t k_I$, et on ne peut plus utiliser directement les techniques précédentes pour adapter STATIS et l'AFM.

Le plus simple et le plus naturel pour généraliser ces deux méthodes est d'effectuer des analyses intra-classes, i.e. de raisonner sur le tableau l_{IXJT} défini au §V-2-3 pour se ramener à des marges proportionnelles avec le même coefficient de proportionnalité $a_t = k(., ., t) / k(., ., .)$ que précédemment entre la marge sur I du $t^{\text{ème}}$ bloc de l_{IXJT} et la marge globale de l_{IXJT} (qui est égale à k_I).

Tous les résultats obtenus précédemment restent alors valables, à condition de remplacer le tableau k_{IXJT} par l_{IXJT} .

C'est cette méthode, qui se ramène à l'analyse des correspondances du tableau l_{IXJT} pondéré, qui nous semble la plus adéquate, et qui au niveau de l'AFM a été développée et utilisée par BECUE-BERTAUT, PAGES (2001, 2003) sous le nom de MFACT (Multiple Factorial Analysis of Contingency Tables), avec une façon de voir un peu différente de celle proposée ici.

Remarque : Quand les marges sur I des tableaux k_{Ijt} sont proportionnelles i.e., quand $k_{It} = a_t k_I$, auquel cas $k(i, ., t) / k(., ., t) = k(i, ., .) / k(., ., .)$, il est immédiat de vérifier que le tableau k_{IXJT} est identique à l_{IXJT} .

D'autres possibilités ont été proposées, ou peuvent être envisagées, pour étudier une série de tableaux de contingence en effectuant une analyse pondérée :

- a) On peut raisonner sur le tableau $X_{I \times J \times T}$, et sur les blocs X_{Ijt} , en adoptant la métrique des poids D_p associée au tableau global $k_{I \times J \times T}$, la métrique associée à l'ensemble JT des variables étant toujours la métrique usuelle. Dans ce cas, les sous-tableaux X_{Ijt} ne sont plus centrés, et l'AFM (ou STATIS) n'est plus équivalente (à une constante près) à une analyse des correspondances pondérées.
- b) On peut aussi considérer le tableau $Z_{I \times J \times T}$ juxtaposition des blocs Z_{Ijt} , en adoptant toujours la même métrique des poids D_p que précédemment. Là encore, les tableaux Z_{Ijt} ne sont plus centrés, et l'analyse du tableau $Z_{I \times J \times T}$ après pondération ne se ramène pas à une analyse des correspondances pondérées
- c) On peut également travailler sur le tableau $V_{I \times J \times T}$ et les blocs V_{Ijt} associés, le terme général de $V_{I \times J \times T}$ (ou de V_{Ijt}) s'écrivant :

$$V(i, j, t) = [k(i, j, t) - k(i, \cdot, t) k(\cdot, j, t) / k(\cdot, \cdot, t)] / [k(i, \cdot, t) k(\cdot, j, t)]^{1/2}$$

et effectuer avec des métriques identités, l'ACP du tableau $V_{I \times J \times T}$ après pondération, ce qui revient à diagonaliser les matrices SS' et $S'S$, S désignant le tableau $V_{I \times J \times T}$ après pondération.

C'est ainsi que procèdent ZARRAGA et GOITISOLO (2002), qui à partir des vecteurs propres normés des matrices précédentes définissent par transformation linéaire des facteurs sur I et sur JT . Les facteurs sur JT sont normés pour la métrique $\text{Diag}(k(\cdot, j, t) / k(\cdot, \cdot, t) \mid j \in J, t \in T)$ définie à partir de l'analyse des correspondances de $k_{I \times J \times T}$, tandis que les facteurs sur I sont normés pour une métrique $\text{Diag}(p_i \mid i \in I)$ différente de la métrique $D_p = \text{Diag}(k(i, \cdot, \cdot) / k(\cdot, \cdot, \cdot) \mid i \in I)$, p_i étant égal à $(\sum \{ [k(i, \cdot, t) / k(\cdot, \cdot, t)]^{1/2} \mid t \in T \})^2$.

Remarques :

- 1) L'ACP du bloc V_{Ijt} avec les métriques identités, est équivalente à l'analyse des correspondances du tableau k_{Ijt} .
 - 2) Les tableaux V_{Ijt} ne sont pas centrés (pour la métrique des poids identité adoptée) et l'ACP du tableau $V_{I \times J \times T}$ après pondération n'est pas équivalente à une analyse des correspondances pondérée.
- d) On aurait pu aussi raisonner sur le tableau $W_{I \times J \times T}$ de terme général :

$$W(i, j, t) = [k(i, j, t) - k(i, \cdot, \cdot) k(\cdot, j, t) / k(\cdot, \cdot, \cdot)] / [k(i, \cdot, \cdot) k(\cdot, j, t)]^{1/2}$$

tableau dont l'ACP, la métrique des poids étant la métrique identité (et la métrique sur JT également) est équivalente à l'analyse des correspondances du tableau $k_{I \times J \times T}$.

Là encore, les blocs W_{Ijt} de $W_{I \times J \times T}$ ne sont pas centrés et l'ACP du tableau $W_{I \times J \times T}$ après pondération n'est pas équivalente à une analyse des correspondances pondérée.

Remarque : Si $k_{I_t} = a_t k_I$, toutes les analyses précédentes sont équivalentes entre elles et équivalentes aux analyses pondérées définies au §V-2-4-3, la pondération dépendant de la méthode retenue.

V-3 ETUDE DU CAS OU $I_t = I$ et $J_t = J$

Dans ce cas, on a un tableau cubique k_{IJT} et on peut appliquer les méthodes développées au paragraphe précédent ainsi que les méthodes symétriques (ou duales) obtenues en intervertissant les rôles de I_t et J_t , ainsi que ceux de I et J . Ces méthodes duales reviennent à faire :

- soit l'analyse des correspondances du tableau $k_{IT \times J}$ avec k_{TJ} (transposé de k_{JT}), $k_{IT \times JT}$ ainsi que le tableau bloc-diagonal $k_{IT \times T}$ (dont la $t^{\text{ème}}$ colonne est formée de zéros, sauf son $t^{\text{ème}}$ bloc qui est égal à k_{I_t}) en supplémentaire.
- soit l'analyse des correspondances du tableau k_{TJ} avec $k_{IT \times J}$ en supplémentaire (i.e. l'analyse interclasses).
- soit l'analyse intra-classes et les analyses pondérées associées (STATIS, l'AFM), l'analyse intra-classes revenant ici à l'analyse des correspondances du tableau de terme général :

$$k(i, j, t) - k(i, \cdot, t)k(\cdot, j, t) / k(\cdot, \cdot, t) + k(i, \cdot, t)k(\cdot, j, \cdot) / k(\cdot, \cdot, \cdot)$$

Notons que dans toutes les analyses précédentes (analyses du § V-2, et analyses symétriques, que l'on vient de détailler), on peut rajouter le tableau marginal d'ordre 2 k_{IJ} en supplémentaire.

Si les ensembles I, J, T jouent des rôles symétriques, outre les analyses fondées sur les tableaux croisant I avec JT , et J avec IT , on peut aussi considérer celles basées sur le tableau croisant T avec $IJ = I \times J$. On obtient ainsi 3 types d'analyses possibles, suivant le sous-ensemble, I, J , ou T qui est croisé avec le produit des deux autres. Si on veut garder la symétrie entre I, J et T , on aura intérêt à faire l'analyse des correspondances du tableau de Burt B_{LL} (avec $L=I \cup J \cup T$) associé à k_{IJT} . Rappelons que les blocs non diagonaux de B_{LL} correspondent aux marges binaires de k_{IJT} tandis que les blocs diagonaux notés k_{II}, k_{JJ}, k_{TT} , n'ont de la masse que sur la diagonale, masse associée à la marge d'ordre 1 associée de k_{IJT} .

Dans cette façon de procéder, qui fait donc jouer des rôles symétriques à I, J, T , on ne tient pas compte des interactions supérieures à 2, i.e. de l'interaction d'ordre 3 $I \times J \times T$.

k_{II}	k_{IJ}	k_{IT}
k_{JI}	k_{JJ}	k_{JT}
k_{TI}	k_{TJ}	k_{TT}

Tableau de Burt B_{LL} associé à k_{IJT} .

Ce tableau est formé des marges d'ordre 2 et des marges d'ordre 1 (ces dernières étant dans les diagonales des blocs diagonaux, ces blocs étant remplis de zéros, hors diagonale)

Figure 3

Si les ensembles I et J jouent des rôles symétriques, et que le troisième ensemble T est de nature différente, (par exemple s'il correspond au temps, comme déjà dit au § IV), on a intérêt à faire jouer aux tableaux k_t (ainsi qu'à I et J) des rôles symétriques.

On adapte donc la procédure développée au § IV (où les tableaux moyens considérés sont remplacés par les tableaux marginaux associés, compte tenu de la propriété rappelée à la fin du § V-1).

Si k_{IJ} désigne le tableau marginal d'ordre 2, défini par :

$$k_{IJ} = \{ k(i, j, \cdot) \mid i \in I, j \in J \}$$

$$k(i, j, \cdot) = \sum \{ k(i, j, t) \mid t \in T \}$$

on fera l'analyse des correspondances du tableau k_{IJ} (qui joue le rôle du tableau H du § IV) avec k_{IXJT} et k_{IT} en colonnes supplémentaires, et $k_{IT \times J}$ et k_{TJ} en lignes supplémentaires.

On obtient alors deux représentations pour chaque t de T :

Une représentation t_c en tant que colonne supplémentaire.

Une représentation t_l en tant que ligne supplémentaire.

On a alors les propriétés barycentriques suivantes très utiles pour interpréter l'analyse :

i est au centre de gravité des (i, t) ($t \in T$)

j est au centre de gravité des (j, t) ($t \in T$)

t_c est au centre de gravité des (j, t) ($j \in J$)

t_l est au centre de gravité des (i, t) ($i \in I$)

k_{IJ}	k_{IJ1}	k_{IJ2}	k_{IJr}	k_{IT}
k_{IJ1}				
k_{IJ2}				
k_{IJr}				
k_{TJ}				

Analyse de k_{IJ} avec k_{IXJT} et k_{IT} en colonnes supplémentaires, et k_{ITxJ} et k_{TJ} en lignes supplémentaires.

Figure 4

VI CAS D'UN MELANGE DE VARIABLES QUANTITATIVES ET QUALITATIVES

Dans ce cas, on a deux séries de tableaux $\{X_s, s \in S\}$ et $\{k_t, t \in T\}$, la première série de tableaux étant associée à des variables quantitatives sur lesquels on peut faire l'ACP, tandis que la seconde correspond à des tableaux sur lesquels on peut faire l'analyse des correspondances (tableaux de contingence, de fréquences tableaux disjonctifs complets, etc.).

On suppose ces deux séries de tableaux définis sur le même ensemble de lignes ($I_t = I$), qu'on suppose être l'ensemble des individus (comme cela a été fait au §V-2, et contrairement aux notations du début de l'article où I correspondait à l'ensemble des variables).

On suppose les tableaux X_s centrés, pour la métrique des poids commune à ces tableaux, métrique que l'on notera D_{pX} et on désigne toujours par X la juxtaposition des X_s et par M la métrique bloc-diagonale des métriques M_s associés aux blocs X_s . Comme au §V, on note k_{It} le $t^{\text{ème}}$ tableau k_t et on désigne par k_{IXJT} leur juxtaposition.

Au lieu de raisonner sur le tableau k_{IXJT} , on considérera le tableau l_{IXJT} défini au §V-2-3 ainsi que le tableau Y_{IXJT} déduit de l_{IXJT} de la même façon que le tableau X_{IXJT} du § V-2-4-2 était déduit du tableau k_{IXJT} . Le terme général de Y_{IXJT} s'écrit alors avec des notations évidentes :

$$Y(i, j, t) = \left[\frac{l(i, j, t) - l(i, \cdot, \cdot) l(\cdot, j, t) / l(\cdot, \cdot, \cdot)}{[l(i, \cdot, \cdot) (l(\cdot, j, t) / l(\cdot, \cdot, \cdot))]^{1/2}} \right]$$

$$= \left[\frac{k(i, j, t) - k(i, \cdot, \cdot) k(\cdot, j, t) / k(\cdot, \cdot, \cdot)}{[k(i, \cdot, \cdot) (k(\cdot, j, t) / k(\cdot, \cdot, \cdot))]^{1/2}} \right]$$

Les blocs Y_{Ijt} sont alors centrés pour la métrique des poids $D_{pY} = \text{Diag}(k(i, \cdot, \cdot)/k(\cdot, \cdot, \cdot) \mid i \in I)$ associée à l'analyse des correspondances du tableau k_{IXJT} .

On posera alors, en notant par Y_t le $t^{\text{ème}}$ bloc Y_{Ijt} de Y_{IXJT} , et en désignant par q et r les cardinaux de S et T respectivement :

$$Z = (X_1, \dots, X_s, \dots, X_q, Y_1, \dots, Y_t, \dots, Y_r) = (X, Y),$$

et on fera l'ACP pondérée du tableau Z (AFM ou STATIS), en prenant comme métrique, sur l'ensemble des colonnes, la métrique bloc-diagonale $\text{Diag}(M_1, \dots, M_q, Id_1, \dots, Id_r) = \text{Diag}(M, Id_{JT})$, Id_t et Id_{JT} désignant les métriques identité sur J_t et JT respectivement.

En ce qui concerne la métrique des poids D_p , si $D_{pX} = D_{pY}$, on prendra D_p égale à ces deux métriques. Ce cas qui est en particulier réalisé si $D_{pX} = (1/n) Id_n$ (avec $n = \text{Card}I$, Id_n étant la matrice identité d'ordre n) et si les tableaux k_t sont des tableaux disjonctifs complets a donné dans le cas de l'AFM des résultats intéressants dans les applications.

Si $D_{pX} \neq D_{pY}$, on pourra prendre comme métrique des poids D_p , soit D_{pX} (auquel cas les tableaux Y_t ne sont plus centrés), soit D_{pY} (auquel cas les tableaux X_s ne sont plus centrés). En général, on a plutôt tendance à choisir $D_p = D_{pY}$ métrique des poids dérivée de l'analyse des correspondances du tableau k_{IXJT} . C'est cette façon de faire qui a été réalisée dans le cadre de l'AFM par ABDESSEMED, ESCOFIER (1996) pour l'étude simultanée de deux tableaux, un tableau de mesures et un tableau de fréquences (auquel cas $q=r=1$, $X=X_1$ et $Y=Y_1$).

BIBLIOGRAPHIE

ABDESSEMED, L., ESCOFIER, B. (1996) : Analyse factorielle multiple de tableaux de fréquences ; comparaison avec l'analyse canonique des correspondances, JSSP, tome 137, n°2, pp.3-17.

BECUE-BERTAUT, M., PAGES, J. (2001) : Analyse simultanée de questions ouvertes, et de questions fermées. Méthodologie, Exemple, Journal de la S.F.D.S., tome 142, n° 4, pp.91, 104.

BECUE-BERTAUT, M., PAGES, J. (2003) : A principal axes method for comparing multiple contingency tables : MFACT, à paraître dans Computing, Statistics, and Data Analysis.

CASIN, Ph. (1995) : L'Analyse Discriminante de Tableaux Evolutifs, Revue de Statistique Appliquée, Vol. 43, n° 3, pp. 73, 91.

CASIN, Ph. (1996) : L'Analyse en Composantes Principales Généralisées, Revue de Statistique Appliquée, Vol. 44, n° 3, pp.63, 81.

CAZES, P. (1982) : Note sur les éléments supplémentaires en analyse des correspondances. II Tableaux multiples (EL. SUPP. II), Cahiers de l'analyse des données, Vol. VII, n° 2, pp. 133, 144.

CAZES, P. (1987) : Cours d'analyse des données, polycopié ISUP, MD, 2^{ème} et 3^{ème} cycle, 326 pages.

CAZES, P. (1990) : Quelques considérations sur l'utilisation des éléments supplémentaires en analyse factorielle, Revue de MODULAD n° 5, pp.29-44.

CAZES, P. (1999) : Méthodes de régression, polycopié DESS MD, 204 pages.

CAZES P., MOREAU J. (2000) : Analyse des correspondances d'un tableau de contingence dont les lignes et les colonnes sont munies d'une structure de graphe bistochastique ; in Collection Mathématiques et Applications, Vol. 32, pp.87-103, Springer Verlag.

CHEssel, D., HANAFI, M. (1996) : Analyse de la co-inertie de K nuages de points , Revue de Statistique Appliquée, Vol. 44, n° 2 , pp.35, 60 .

D'AMBRA, L., SABATIER, R., AMENTA, P. (2001) : Analisi fattoriale delle matrici a tre vie : sintesi e nuovi approcci, Statistica Applicata, Vol.13, n° 2, pp.101-117.

DAZY , F., LE BARZIC, J.F. (1996) : L'analyse des données évolutives. Méthodes et applications, Editions Technip, 228 pages.

ESCOFIER, B., PAGES, J. (1998) : Analyses factorielles simples et multiples. Objectifs, méthodes et interprétation, 3^{ème} édition, Dunod, 284 pages.

LAVIT, Ch. (1988) : Analyse conjointe de tableaux quantitatifs, Masson, 260 pages.

SAPORTA, G. (1990) : Probabilités, Analyse des données et Statistique, Editions Technip, 494 pages.

VIVIEN, M. (2002) : Approches linéaires et non linéaires pour la modélisation de multi-tableaux : théorie et applications, thèse, Université de Montpellier I.

ZARRAGA, A., GOITISOLO, B. (2002) : Méthode factorielle pour l'analyse simultanée de tableaux de contingence, Revue de Statistique Appliquée, Vol. 50, n° 2, pp.47, 70.