REVUE DE STATISTIQUE APPLIQUÉE

A. ZÁRRAGA
B. GOITISOLO

Méthode factorielle pour l'analyse simultanée de tableaux de contingence

Revue de statistique appliquée, tome 50, n° 2 (2002), p. 47-70
<http://www.numdam.org/item?id=RSA_2002__50_2_47_0>

Rev. Statistique Appliquée, 2002, **L** (2), 47-70

MÉTHODE FACTORIELLE POUR L'ANALYSE SIMULTANÉE DE TABLEAUX DE CONTINGENCE

A. ZÁRRAGA et B. GOITISOLO

Departamento de Economía Aplicada III,
Universidad del País Vasco-Euskal Herriko Unibertsitatea, Bilbao, España

RÉSUMÉ

L'objectif de ce travail est de développer une méthode factorielle pour l'analyse conjointe des tendances générales de plusieurs tableaux de contingence qui permette, de manière similaire à l'analyse des correspondances, l'étude des ressemblances entre l'ensemble des lignes, des colonnes, et les relations entre les deux ensembles, sans permettre que l'un des tableaux ait une prépondérance sur les autres. Avec la méthode proposée, il est possible d'analyser conjointement plusieurs tableaux de contingence, comme par exemple ceux qui correspondent à une même enquête effectuée dans différents pays ou à différentes époques, ou à des enquêtes différentes effectuées sur une même population. La méthode proposée peut s'étendre à l'analyse d'un ensemble d'enquêtes quand les données des différents groupes d'individus sont codifiées en tables disjonctives complètes juxtaposées pour les mêmes modalités.

Mots-clés : *Analyse factorielle, Juxtaposition de Tableaux de Contingence, Analyse des Correspondances.*

ABSTRACT

The aim of this work is to develop a factorial method for the joint analysis of general tendencies in various contingency tables so that the similarity between the set of rows and of columns, and the relationships between them can be studied in a way similar to correspondence analysis, without allowing any one table to be preponderant over the rest. With the method proposed it is possible to analyse several contingency tables in the true sense of the term jointly, along with tables corresponding, for instance, to the same survey carried out in different countries or for different time periods, and those for surveys on different topics carried out on the same population. The method proposed can be used for the analysis of a set of surveys when data from different groups of individuals are coded in complete disjunctive tables and those tables are juxtaposed for the same modalities.

Keywords : *Factor Analysis, Contingency Tables Juxtaposition, Correspondence Analysis.*

1. Introduction

La nécessité d'analyser simultanément plusieurs tableaux de données a engendré de nombreuses méthodes factorielles. Toutefois, la majorité d'entre elles requièrent des tableaux de variables quantitatives et/ou exigent que les différents tableaux de données fassent référence au même ensemble d'individus, (comme par exemple Cazes (1980a), Cazes (1981a)), l'analyse canonique (Hotelling (1936), Carroll (1968)), l'analyse procrustéenne (Tucker (1958), Gower (1975), Gower (1984)), les modèles INDSCAL (Carroll & Chang (1970)) et IDIOSCAL (Tucker (1972)). Parmi les méthodes supportant des variables différentes et l'introduction d'une pondération sur les tableaux, on trouve l'analyse factorielle multiple (Escofier & Pàges (1984)) et la méthode statis : (Structuration des Tableaux À Trois Indices de la Statistique (L'Hermier des Plantes (1976)).

Pour analyser conjointement des tableaux de contingence, quelques-unes des méthodes mentionnées consistent à analyser le tableau somme des tableaux de contingence séparés ou le tableau obtenu comme juxtaposition des tableaux de départ (Cazes (1980b), Cazes (1981b) et Leclerc (1973)). Citons aussi l'analyse non symétrique (Lauro & D'Ambra (1984) et D'Ambra & Lauro (1989)), l'analyse intra (Escofier (1983)) et plus récemment l'analyse factorielle multiple intra-tableaux (Bécue-Bertaut & Pagès (2000)). Pourtant, au § 2 on voit qu'il existe des situations où aucune de ces méthodes ne permet une analyse des ressemblances entre les lignes qui maintienne la similitude existant dans les analyses des tableaux séparés.

L'objectif de ce travail est de développer une méthode factorielle pour l'analyse conjointe de plusieurs tableaux de contingence qui permette, de manière similaire à l'analyse des correspondances, l'étude des ressemblances entre l'ensemble des lignes, des colonnes, et les relations entre les deux ensembles.

La méthode proposée permet de faire l'analyse quelque soit le nombre de tableaux de contingence ou de tableaux de nombres positifs susceptibles d'être analysés par l'analyse des correspondances. Nous développerons, cependant, la méthode en faisant référence aux tableaux de contingence.

2. Problèmes

Dans l'analyse factorielle simultanée de plusieurs tableaux de contingence, une série de problèmes surgissent, comme l'effectif différent des tableaux, des totaux différents de la même ligne dans les différents tableaux qui supposent que les lignes ont des poids différents, et de plus l'analyse conjointe sera influencée par les inerties des différents tableaux sur chaque axe.

Un problème apparaît quand le total de chaque tableau (nombre d'individus) est différent. Cela influence l'analyse de la juxtaposition des tableaux de contingence. Le tableau de plus grand effectif aura une contribution plus grande sur la création des axes.

En analysant la juxtaposition des tableaux on pourrait obtenir que les premiers axes ne montrassent plus que les ressemblances et différences entre les profils du tableau ayant le plus grand effectif.

Donc, la première étape pour équilibrer l'influence des tableaux, avant de passer à l'analyse conjointe, sera de transformer chacun d'eux en fréquences.

Un autre aspect déterminant dans l'analyse conjointe de tableaux de contingence est la différence entre les marginales de deux tableaux. Les relations existant à l'intérieur de chaque tableau se verront altérées si dans l'analyse chaque tableau ne conserve pas ses poids et sa métrique.

Cet inconvénient nous induit à chercher une méthode d'analyse factorielle conjointe de tableaux de contingence dans laquelle on maintienne pour chaque tableau les poids ainsi que la métrique.

Enfin, la différence d'ordre de grandeur de l'inertie projetée sur le premier axe pour chacun des tableaux peut amener à une analyse conjointe dominée par les différences d'un des tableaux.

Ce problème a été déjà considéré par l'analyse factorielle multiple et la méthode statis, quand les tableaux sont de variables continues. Afin d'équilibrer l'influence que chaque tableau exerce dans l'analyse conjointe et de ne pas permettre que celle-ci soit dominée par un tableau particulier, ces méthodes introduisent une pondération sur les tableaux. Une analyse simultanée de plusieurs tableaux dont les premiers axes seraient engendrés par un seul d'entre eux ne présenterait en effet que peu d'intérêt.

Dans l'analyse conjointe des tableaux de contingence on introduira aussi une pondération sur chaque tableau.

3. Analyse des Correspondances d'un tableau de contingence

Soit un tableau de contingence où $\mathcal{I} = \{1, \ldots, i, \ldots, I\}$ et $\mathcal{J} = \{1, \ldots, j, \ldots, J\}$ représentent l'ensemble des lignes et des colonnes du tableau qui correspondent aux modalités de deux variables qualitatives. Soit f_{ij} la fréquence relative d'appartenir aux modalités $i \in \mathcal{I}$ et $j \in \mathcal{J}$, $f_{i.} = \sum_{j \in \mathcal{J}} f_{ij}$, $f_{.j} = \sum_{i \in \mathcal{I}} f_{ij}$ les marges associées.

À chaque ligne i du tableau est associé un profil-ligne, $\left\{\frac{f_{ij}}{f_{i}} \mid j \in \mathcal{J}\right\}$, affecté d'un poids $f_{i.}$. L'ensemble des profils-lignes se représentent dans l'espace R^J, muni de la métrique du χ^2. L'analyse des correspondances recherche les directions principales d'allongement, appelées axes factoriels, du nuage $\mathcal{N}(\mathcal{I})$ des profils-lignes, centré sur son centre de gravité. La projection du profil-ligne i sur le s-ième axe factoriel est notée $F_s(i)$ et F_s désigne le s-ième facteur sur l'ensemble $\mathcal{I}$. Étant donné que dans un tableau de contingence les lignes et colonnes jouent des rôles symétriques l'étude des profils-colonnes $\left\{\frac{f_{ij}}{f_{j}} \mid i \in \mathcal{I}\right\}$, avec les poids $f_{.j}$ peut être réalisée de façon analogue. G_s désignera le s-ième facteur sur l'ensemble $\mathcal{J}$.

L'obtention de ces facteurs peut être réalisée de plusieurs façons différentes, largement décrites dans de nombreuses publications, dont en particulier (Benzécri & collaborateurs (1973)). Dans le but d'introduire le lecteur aux expressions employées au § 4, on présente dans la suite une manière possible d'obtention des facteurs ainsi que les relations entre les facteurs sur $\mathcal{I}$ et sur $\mathcal{J}$.

En diagonalisant la matrice $X_*^T X_*$, où le terme général de X_* est :

$$\sqrt{f_{i.}} \left(\frac{f_{ij}}{f_{i.} f_{.j}} - 1 \right) \sqrt{f_{.j}} \qquad \begin{array}{c} i \in \mathcal{I} \\ j \in \mathcal{J} \end{array} \tag{1}$$

on obtient les valeurs propres λ_s^*, $s \in \mathcal{S}$, inertie projetée sur le s-ième axe factoriel, et les vecteurs propres $u_s^* \in R^J$, $s \in \mathcal{S}$, normalisés pour la métrique euclidienne. Et en diagonalisant la matrice $X_* X_*^T$, on obtient les valeurs propres μ_s^* et les vecteurs propres $v_s^* \in R^I$, $s \in \mathcal{S}$, normalisés pour la métrique euclidienne.

Si P_* et M_* désignent les matrices diagonales qui contiennent les poids des profils-lignes et profils-colonnes, $f_{i.}$ et $f_{.j}$, respectivement, on peut obtenir les projections des profils de la manière suivante :

$$\begin{aligned} F_s &= P_*^{-1/2}\, X_*\, u_s^* = P_*^{-1/2}\, X_*\, M_*^{1/2}\, G_s\, \lambda_s^{*\,-1/2} && s \in \mathcal{S} \\ G_s &= \hat{M}_*^{-1/2}\, X_*^T\, P_*^{1/2}\, F_s\, \lambda_s^{*\,-1/2} && s \in \mathcal{S} \end{aligned}$$

puisque la relation entre les valeurs propres et vecteurs directeurs des deux analyses est :

$$\begin{aligned} \lambda_s^* &= \mu_s^* && s \in \mathcal{S} \\ u_s^* &= \lambda_s^{*\,-1/2} X_*^T\, v_s^* && s \in \mathcal{S} \\ v_s^* &= \lambda_s^{*\,-1/2}\, X_*\, u_s^* && s \in \mathcal{S} \end{aligned}$$

4. Traitement simultané de plusieurs tableaux de contingence

Au § 2 nous avons pu vérifier l'importance de conserver dans une analyse factorielle conjointe de tableaux de contingence aussi bien les poids que la métrique de chacun des tableaux et la nécessité d'équilibrer l'influence des tableaux en ajustant l'effectif de chacun et en introduisant une pondération sur chacun d'eux. Dans la suite, nous proposons une analyse conjointe de tableaux de contingence respectant ces conditions.

Soit $\mathcal{G} = \{1, \ldots, g, \ldots, G\}$ l'ensemble des tableaux de contingence à analyser. Chacun d'entre eux classifie les réponses de $k_{..}^g$ individus à deux variables qualitatives codifiées en modalités. Tous les tableaux ont une des variables (dont les modalités, $\mathcal{I} = \{1, \ldots, i, \ldots, I\}$, se retrouvent dans les lignes) en commun. L'autre variable de chaque tableau de contingence peut être la même ou différente, codifiée de la même manière ou d'une manière différente. Les modalités de la seconde variable de chaque tableau g sont $\mathcal{J}_g = \{1, \ldots, j, \ldots, J_g\}$. En juxtaposant tous ces tableaux de contingence, on obtient un ensemble $\mathcal{J} = \{1, \ldots, j, \ldots, J\}$ de colonnes.

L'élément k_{ij}^g, correspond au nombre total d'individus qui choisissent simultanément les modalités $i \in \mathcal{I}$ de la première variable et $j \in \mathcal{J}_g$ de la seconde variable (appartenant au tableau $g \in \mathcal{G}$).

Nous avons vu que l'ajustement des effectifs des tableaux est le premier pas pour équilibrer l'influence des tableaux (§ 2). Pour cela, chacun des tableaux de contingence est transformé, en le divisant par $k_{..}^g$, total du g-ième tableau, $g \in \mathcal{G}$, en un tableau de fréquences relatives.

On notera :

$$\begin{aligned} f_{ij}^g &= \frac{k_{ij}^g}{k^g} & f_{.j}^g &= \sum_{i \in \mathcal{I}} f_{ij}^g & & i \in \mathcal{I} \\ f_{i.}^g &= \sum_{j \in \mathcal{J}_g} f_{ij}^g & f_{..}^g &= \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}_g} f_{ij}^g = 1 & & j \in \mathcal{J} \\ & & & & & g \in \mathcal{G} \end{aligned}$$

Au § 2, nous avons conclu à la nécessité d'introduire une pondération, que nous appelons α^g , $g \in \mathcal{G}$, sur chacun des groupes (ici les tableaux de contingence) avec l'objectif d'équilibrer l'influence des tableaux. Cette pondération dépendra des inerties qui résultent d'une analyse des correspondances simples de chacun des tableaux traités séparément. Ces inerties pour le tableau g, $g \in \mathcal{G}$, seront notées λ_s^g (l'inertie projetée sur l'axe s, $s \in \mathcal{S}$) et λ_{TOT}^g (l'inertie totale). La pondération peut être équivalente à celle de l'analyse factorielle multiple pour variables continues $(1/\lambda_1^g)$, celle équivalente à la méthode statis (L'Hermier des Plantes (1976)), $1/\lambda_{TOT}^g$ pour privilégier des groupes de dispersion minimale, 1 si on ne veut pas de pondération sur les tableaux, etc.

Si l'on considère $\alpha_g = 1/\lambda_1^g$, $g \in \mathcal{G}$, pondération adoptée dans l'exemple d'application (§ 5), la première valeur propre de l'analyse factorielle de l'ensemble des tableaux de contingence sera comprise entre 0 et G. Si on veut rester dans le cadre de l'analyse des correspondances et que la première valeur propre de l'analyse de l'ensemble des tableaux soit comprise entre 0 et 1, il suffira d'adopter α_g/α comme pondération avec $\alpha = \sum_{g \in G} \alpha_g$.

4.1. Analyse des colonnes

À chacun des g, $g \in \mathcal{G}$, tableaux de contingence est associé un sous-nuage, $\mathcal{N}(\mathcal{J}_g)$, de J_g profils-colonnes centrés :

$$\left\{ \frac{f_{ij}^g}{f_{.j}^g} - f_{i.}^g \mid i \in \mathcal{I} \right\} \qquad \begin{array}{c} j \in \mathcal{J}_g \\ g \in \mathcal{G} \end{array}$$

avec les poids $f_{.j}^g$, $j \in \mathcal{J}_g$, et la métrique de matrice associée la matrice diagonale des $1/f_{i.}^g$, $i \in \mathcal{I}$, $g \in \mathcal{G}$.

Pour analyser les tableaux ensemble, nous sur-pondérons chaque sous-nuage par α^g , $g \in \mathcal{G}$ et comme ils sont tous situés dans le même espace, R^I, nous considérons le nuage global, $\mathcal{N}(\mathcal{J})$, qui englobe les J profils-colonnes. Dans ce nuage, les métriques sont différentes pour chaque sous-nuage de profils du même tableau, ce qui peut sembler empêcher l'analyse conjointe.

On peut transformer les profils-colonnes de chacun des tableaux pour considérer leurs distances euclidiennes. Dans l'analyse conjointe cela signifie considérer les profils-colonnes :

$$\left\{ \frac{\sqrt{\alpha^g}}{\sqrt{f^g_{i.}}} \left(\frac{f^g_{ij}}{f^g_{.j}} - f^g_{i.} \right) \mid i \in \mathcal{I} \right\} \qquad \begin{array}{l} j \in \mathcal{J}_g \subset \mathcal{J} \\ g \in \mathcal{G} \end{array}$$

avec les poids $f^g_{.j}$, $j \in \mathcal{J}$, et la métrique euclidienne usuelle.

Dans cette analyse, les distances euclidiennes entre profils-colonnes du même sous-nuage respectent les distances du χ^2 dans le sous-nuage original.

4.2. Analyse des lignes

Dans chacun des g, $g \in \mathcal{G}$, tableaux de contingence on définit les profils-lignes centrés,

$$\left\{ \frac{f^g_{ij}}{f^g_{i.}} - f^g_{.j} \mid j \in \mathcal{J}_g \right\} \qquad \begin{array}{l} i \in \mathcal{I} \\ g \in \mathcal{G} \end{array} \tag{2}$$

avec les poids $f^g_{i.}$, $i \in \mathcal{I}$, et la métrique de matrice associée la matrice diagonale des $1/f^g_{.j}$, $j \in \mathcal{J}_g$, $g \in \mathcal{G}$.

Puisque l'on cherche à analyser ensemble les g, $g \in \mathcal{G}$, tableaux et que les profils-lignes de chaque tableau sont représentés dans des espaces distincts, chacun dans un espace de dimension J_g, il faut chercher un espace commun dans lequel on puisse effectuer l'analyse. Cet espace commun est R^J, où l'on représente les profils-lignes de chacun des tableaux, appelés profils-lignes partiels. Les coordonnées de ces points correspondent à celles définies en (2), en complétant le reste des coordonnées par 0. Le profil-ligne partiel i^g, $i \in \mathcal{I}$, $g \in \mathcal{G}$ a alors pour coordonnées :

$$i^g = \begin{cases} \dfrac{f^g_{ij}}{f^g_{i.}} - f^g_{.j} & \text{si} \quad j \in \mathcal{J}_g \\ 0 & \text{sinon} \end{cases} \tag{3}$$

L'ensemble des profils-lignes partiels d'un même tableau forment un sous-nuage de points que nous noterons $\mathcal{N}(\mathcal{I}^g)$.

Afin d'effectuer l'analyse conjointe, on cherche pour chaque ligne un représentant, que nous noterons i^c, $i \in \mathcal{I}$, dit «compromis», qui représente le sous-nuage de profils-lignes formé par tous les points de la même ligne dans les différents tableaux. L'ensemble de tous les représentants sur R^J, muni de la métrique de matrice associée la matrice diagonale des $\alpha^g / f^g_{.j}$, $j \in \mathcal{J}_g \subset \mathcal{J}$, forme le nuage $\mathcal{N}(\mathcal{I}^c)$. Le compromis est choisi avec l'objectif que son inertie puisse s'exprimer comme une somme pondérée des inerties des profils-lignes partiels :

$$\text{In}(i^c) = \sum_{g \in \mathcal{G}} \alpha^g \, \text{In}(i^g) \tag{4}$$

et que, par conséquent, l'inertie du nuage compromis s'exprime comme somme pondérée des inerties des nuages partiels. Pour cela on définit le compromis comme moyenne pondérée des profils-lignes partiels i^c avec les poids p_{i^c} , $i \in \mathcal{I}$: [1]

$$i^c = \sum_{g \in \mathcal{G}} \frac{\sqrt{f_{i.}^g}}{\sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g}} \, i^g \qquad\qquad p_{i^c} = \left(\sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g} \right)^2$$

En effet, pour le vérifier, commençons par obtenir la distance, au carré, du compromis à l'origine :

$$\begin{aligned} d^2(i^c, 0) &= \sum_{g \in \mathcal{G}} \sum_{j \in \mathcal{J}_g} \alpha^g \frac{1}{f_{.j}^g} \left(\frac{\sqrt{f_{i.}^g}}{\sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g}} \left(\frac{f_{ij}^g}{f_{i.}^g} - f_{.j}^g \right) \right)^2 = \\ &= \sum_{g \in \mathcal{G}} \alpha^g \frac{f_{i.}^g}{p_{i^c}} \sum_{j \in \mathcal{J}_g} \frac{1}{f_{.j}^g} \left(\frac{f_{ij}^g}{f_{i.}^g} - f_{.j}^g \right)^2 \end{aligned} \tag{5}$$

D'après la définition du profil-ligne partiel vu en (3), la seconde somme correspond à la distance, au carré, d'un profil-ligne partiel au centre de gravité du sous-nuage formé par tous les profils-lignes partiels correspondants au même tableau que nous noterons $d^2(i^g, 0)$. De cette manière, la distance, au carré, peut s'écrire :

$$d^2(i^c, 0) = \sum_{g \in \mathcal{G}} \alpha^g \frac{f_{i.}^g}{p_{i^c}} d^2(i^g, 0) \qquad\qquad i \in \mathcal{I}$$

et en multipliant par le poids du compromis, p_{i^c} , on obtient la relation cherchée (4) entre l'inertie du compromis et des profils-lignes partiels.

4.3. Obtention des facteurs

Afin de chercher la relation entre les analyses des lignes et des colonnes dans les développements suivants, l'analyse factorielle de l'ensemble des tableaux de contingence s'effectue en recherchant les valeurs propres (λ_s) et les vecteurs propres (u_s), $s \in \mathcal{S}$, issus de la diagonalisation de la matrice $A = X^T X$ où le terme général de la matrice X est :

$$x_{ij} = \sqrt{\alpha^g} \sqrt{f_{i.}^g} \left(\frac{f_{ij}^g}{f_{i.}^g f_{.j}^g} - 1 \right) \sqrt{f_{.j}^g} \qquad \begin{array}{c} i \in \mathcal{I} \\ j \in \mathcal{J} \\ g \in \mathcal{G} \end{array} \tag{6}$$

[1] Avec $\sum_{i \in \mathcal{I}} p_{i^c} = p \neq 1$. Si l'on considère $p_{i^c}^* = p_{i^c} / p$ (avec $\sum_{i \in \mathcal{I}} p_{i^c}^* = 1$), les résultats factoriels ne se verront altérés que dans la proportion $1/p$. On a considéré p_{i^c} dans la mesure où cela rend les formules plus claires.

De plus, on utilisera les matrices diagonales R d'ordre $I \times I$ et N d'ordre $J \times J$, de termes généraux respectivement :

$$r_{ii} = p_{i^c} \qquad n_{jj} = f^g_{.j} \qquad \begin{array}{l} i \in \mathcal{I} \\ j \in \mathcal{J}_g \subset \mathcal{J} \\ g \in \mathcal{G} \end{array}$$

On définit aussi la matrice Y, diagonale par blocs, où chaque bloc de la diagonale est la matrice X_g d'ordre $(I \times J_g)$ qui se compose des coordonnées de la matrice X pour l'ensemble des colonnes J_g du tableau $g,\ g \in \mathcal{G}$, et on définit la matrice Q, diagonale par blocs aussi, où chaque bloc est, de même, une matrice diagonale d'ordre $(I \times I)$ et de terme général $f^g_{i.},\ i \in \mathcal{I},\ g \in \mathcal{G}$.

4.4. Relation entre les vecteurs propres des deux analyses

Puisque dans l'analyse des lignes on diagonalise la matrice $X^T X$ et dans celle des colonnes la matrice $X X^T$ la relation entre les valeurs propres et vecteurs directeurs des analyses des lignes (λ_s et $u_s \in R^J$) et colonnes (μ_s et $v_s \in R^I$) est :

$$\lambda_s = \mu_s \qquad s \in \mathcal{S}$$

$$u_s = {\lambda_s}^{-1/2}\, X^T\, v_s \qquad s \in \mathcal{S} \tag{7}$$

$$v_s = {\lambda_s}^{-1/2}\, X\, u_s \qquad s \in \mathcal{S} \tag{8}$$

4.5. Projections des profils

On calcule les projections sur l'axe $s,\ s \in \mathcal{S}$, des profils-lignes partiels $F_s(i^g)$, des compromis $F_s(i^c),\ i \in \mathcal{I}$, et des profils-colonnes $G_s(j),\ j \in \mathcal{J}$. Les projections de tous les profils-lignes partiels et compromis sont respectivement $F_s(\mathcal{I}^{\mathcal{G}})$ avec $\mathcal{I}^{\mathcal{G}} = \{\mathcal{I}^g / g \in \mathcal{G}\}$ et $F_s(\mathcal{I}^c)$ et de toutes les colonnes $G_s(\mathcal{J})$. Les projections sur l'axe $s,\ s \in \mathcal{S}$, des lignes et colonnes sont calculées en sachant qu'il est nécessaire d'éliminer l'effet de l'introduction des poids et les profils-lignes partiels sont projetés comme supplémentaires en éliminant aussi l'effet de l'introduction du poids :

$$F_s(\mathcal{I}^c) = R^{-1/2}\, X\, u_s = {\lambda_s}^{1/2}\, R^{-1/2}\, v_s \tag{9}$$

$$G_s(\mathcal{J}) = N^{-1/2}\, X^T\, v_s = {\lambda_s}^{1/2}\, N^{-1/2}\, u_s \tag{10}$$

$$F_s(\mathcal{I}^{\mathcal{G}}) = Q^{-1/2}\, Y\, u_s \tag{11}$$

4.6. Propriétés des facteurs

Les projections des profils-colonnes ainsi que celles des profils-lignes partiels sont centrées, dans chaque tableau partiel, pour leurs poids.

- À partir de la relation (10) on obtient les projections sur l'axe s, $s \in \mathcal{S}$, d'une colonne déterminée j, $j \in \mathcal{J}$, $G_s(j)$:

$$G_s(j) = \sum_{i \in \mathcal{I}} \sqrt{\alpha^g} \sqrt{f_{i.}^g} \left(\frac{f_{ij}^g}{f_{i.}^g f_{.j}^g} - 1 \right) v_{si} \qquad \begin{matrix} i \in \mathcal{I} \\ g \in \mathcal{G} \end{matrix}$$

où v_{si} est la i - ième coordonnée du vecteur directeur v_s, $s \in \mathcal{S}$.

Le barycentre de ces projections pour le tableau partiel g est :

$$\sum_{j \in \mathcal{J}_g} f_{.j}^g G_s(j) = \sum_{j \in \mathcal{J}_g} \sqrt{\alpha^g} \left(f_{.j}^g \sum_{i \in \mathcal{I}} \frac{1}{\sqrt{f_{i.}^g}} \frac{f_{ij}^g}{f_{.j}^g} v_{si} - f_{.j}^g \sum_{i \in \mathcal{I}} \sqrt{f_{i.}^g} v_{si} \right) =$$

$$= \sum_{i \in \mathcal{I}} \frac{\sqrt{\alpha^g}}{\sqrt{f_{i.}^g}} f_{i.}^g v_{si} - \sum_{i \in \mathcal{I}} \sqrt{\alpha^g} \sqrt{f_{i.}^g} v_{si} = 0 \tag{12}$$

c'est-à-dire, les projections des colonnes sont centrées parce que les colonnes de chaque tableau, et donc de l'ensemble, ont été centrées.

- La projection sur l'axe s, $s \in \mathcal{S}$, pour un profil-ligne partiel est, d'après la relation (11) :

$$F_s(i^g) = \sum_{j \in \mathcal{J}_g} \sqrt{\alpha^g} \left(\frac{f_{ij}^g}{f_{i.}^g f_{.j}^g} - 1 \right) \sqrt{f_{.j}^g} u_{sj} \qquad \begin{matrix} i \in \mathcal{I} \\ g \in \mathcal{G} \end{matrix} \tag{13}$$

avec u_{sj} la j - ième coordonnée du vecteur directeur u_s, $s \in \mathcal{S}$.

D'où, le barycentre des projections des profils-lignes partiels correspondant au groupe g, $g \in \mathcal{G}$, est :

$$\sum_{i \in \mathcal{I}} f_{i.}^g F_s(i^g) = \sqrt{\alpha^g} \sum_{i \in \mathcal{I}} \left(f_{i.}^g \frac{1}{f_{i.}^g} \sum_{j \in \mathcal{J}_g} \frac{f_{ij}^g}{\sqrt{f_{.j}^g}} u_{sj} \right) -$$

$$- \sqrt{\alpha^g} \sum_{i \in \mathcal{I}} \left(f_{i.}^g \sum_{j \in \mathcal{J}_g} \sqrt{f_{.j}^g} u_{sj} \right) =$$

$$= \sqrt{\alpha^g} \sum_{j \in \mathcal{J}_g} \sqrt{f_{.j}^g} u_{sj} - \sqrt{\alpha^g} \sum_{j \in \mathcal{J}_g} \sqrt{f_{.j}^g} u_{sj} = 0$$

- La projection sur l'axe s, $s \in \mathcal{S}$, d'un compromis vient de (9) :

$$F_s(i^c) = \sum_{g \in \mathcal{G}} \sum_{j \in \mathcal{J}_g} \frac{\sqrt{f_{i.}^g}}{\sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g}} \sqrt{\alpha^g} \left(\frac{f_{ij}^g}{f_{i.}^g f_{.j}^g} - 1 \right) \sqrt{f_{.j}^g} u_{sj} = \tag{14}$$

$$= \frac{1}{\sum_{g\in\mathcal{G}} \sqrt{f_{i.}^g}} \sum_{g\in\mathcal{G}} \left(\sqrt{\alpha^g} \sum_{j\in\mathcal{J}_g} \frac{f_{ij}^g}{\sqrt{f_{i.}^g}} \frac{1}{\sqrt{f_{.j}^g}} u_{sj} \right) -$$

$$- \frac{1}{\sum_{g\in\mathcal{G}} \sqrt{f_{i.}^g}} \sum_{g\in\mathcal{G}} \left(\sqrt{\alpha^g} \sqrt{f_{i.}^g} \sum_{j\in\mathcal{J}_g} \sqrt{f_{.j}^g}\, u_{sj} \right) \qquad i \in \mathcal{I}$$

En conséquence, les projections des compromis ne sont pas centrées pour leurs poids, p_{i^c} .

- Les profils de chaque tableau (profils-lignes partiels et profils-colonnes) ont été centrés pour conserver les relations à l'intérieur de chaque tableau et par conséquent leurs projections sont aussi centrées. Les compromis sont une moyenne pondérée des profils-lignes partiels avec des pondérations différentes des poids des profils-lignes partiels et donc ne sont pas centrées, et leurs projections non plus.

4.7. Relations entre $F_s(i^g)$ *,* $F_s(i^c)$ *et* $G_s(j)$

De même que dans l'analyse des correspondances simples, il y a des relations de transition entre les projections des lignes (compromis) et colonnes. On peut de plus les mettre en relation avec les projections des profils-lignes partiels.

1. Relation entre $F_s(i^c)$ et $F_s(i^g)$

   La projection d'un compromis obtenue en (14) peut être exprimée à partir de l'expression (13) :

$$F_s(i^c) = \sum_{g\in\mathcal{G}} \frac{\sqrt{f_{i.}^g}}{\sum_{g\in\mathcal{G}} \sqrt{f_{i.}^g}} F_s(i^g) \qquad i \in \mathcal{I} \tag{15}$$

   Dans la projection, de même que dans l'espace, le compromis est une moyenne pondérée des projections des profils-lignes partiels.

2. Relation entre $F_s(i^g)$ et $G_s(j)$

   En pré-multipliant la relation (7) par $Q^{-1/2}\, Y$ et en exprimant le résultat en fonction des projections sur l'axe s, $s \in \mathcal{S}$, on obtient compte tenu de (10) et (11) :

$$F_s(\mathcal{I}^{\mathcal{G}}) = Q^{-1/2}\, Y\, N^{1/2}\, G_s(\mathcal{J})\, \lambda_s{}^{-1/2}$$

La projection sur l'axe s, $s \in \mathcal{S}$, pour un profil-ligne partiel est, en fonction des projections des colonnes :

$$F_s(i^g) = \sum_{j \in \mathcal{J}_g} \sqrt{\alpha^g}\, f^g_{.j} \left(\frac{f^g_{ij}}{f^g_{i.} f^g_{.j}} - 1 \right) G_s(j)\, \lambda_s{}^{-1/2} =$$

$$= \sqrt{\alpha^g}\, \lambda_s{}^{-1/2} \left(\sum_{j \in \mathcal{J}_g} \frac{f^g_{ij}}{f^g_{i.}} G_s(j) - \sum_{j \in \mathcal{J}_g} f^g_{.j}\, G_s(j) \right) \qquad \begin{matrix} i \in \mathcal{I} \\ g \in \mathcal{G} \end{matrix}$$

Comme la seconde somme est, d'après (12), nulle :

$$F_s(i^g) = \sqrt{\alpha^g}\, \lambda_s{}^{-1/2} \sum_{j \in \mathcal{J}_g} \frac{f^g_{ij}}{f^g_{i.}} G_s(j) \qquad \begin{matrix} i \in \mathcal{I} \\ g \in \mathcal{G} \end{matrix} \tag{16}$$

À part la pondération du groupe $\sqrt{\alpha^g}$ et le coefficient $\lambda_s{}^{-1/2}$, la projection d'un profil-ligne partiel sur l'axe s, $s \in \mathcal{S}$, est, comme dans l'analyse des correspondances classique d'un tableau, le barycentre des projections des profils-colonnes correspondant au tableau auquel il appartient.

3. Relation entre $F_s(i^c)$ et $G_s(j)$

   En exprimant la projection sur l'axe s, $s \in \mathcal{S}$, du compromis en fonction des projections sur le même axe des profils partiels (15) et ceux-ci en fonction des projections des colonnes (16), on obtient :

$$F_s(i^c) = \sum_{g \in \mathcal{G}} \sqrt{\alpha^g}\, \frac{\sqrt{f^g_{i.}}}{\sum_{g \in \mathcal{G}} \sqrt{f^g_{i.}}} \left(\lambda_s{}^{-1/2} \sum_{j \in \mathcal{J}_g} \frac{f^g_{ij}}{f^g_{i.}} G_s(j) \right) \qquad i \in \mathcal{I}$$

   La projection du compromis est donc, aux coefficients $\sqrt{\alpha^g}$ et $\lambda_s{}^{-1/2}$ près, la moyenne pondérée des projections des barycentres des profils-colonnes correspondant à chaque tableau.

4. Relation entre $G_s(j)$ et $F_s(i^c)$ ou $F_s(i^g)$

   En pré-multipliant la relation (8) par $N^{-1/2}\, X^T$, et compte tenu de (9) et (10), nous obtenons les projections sur l'axe s, $s \in \mathcal{S}$, des colonnes en fonction des compromis :

$$N^{-1/2}\, X^T\, v_s = N^{-1/2}\, X^T\, R^{1/2}\, R^{-1/2}\, X\, u_s\, \lambda_s{}^{-1/2}$$
$$G_s(\mathcal{J}) = N^{-1/2}\, X^T\, R^{1/2}\, F_s(\mathcal{I}^c)\, \lambda_s{}^{-1/2}$$

La projection sur l'axe s, $s \in \mathcal{S}$, du profil-colonne j du groupe g, $j \in \mathcal{J}_g$, $g \in \mathcal{G}$, peut donc être exprimée de la manière suivante :

$$G_s(j) = \sqrt{\alpha^g}\, \lambda_s^{-1/2} \left(\sum_{i \in \mathcal{I}} \left(\sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g} \right) \sqrt{f_{i.}^g} \left(\frac{f_{ij}^g - f_{i.}^g f_{.j}^g}{f_{i.}^g f_{.j}^g} \right) F_s(i^c) \right) \qquad \forall j \in \mathcal{J}_g$$

Une colonne se situe du côté des compromis auxquels elle s'associe plus qu'elle ne le ferait dans l'hypothèse d'indépendance et du côté opposé de ceux auxquels elle s'associe moins que dans l'hypothèse d'indépendance.

La même projection en fonction des profils-lignes partiels :

$$G_s(j) = \sqrt{\alpha^g}\, \lambda_s^{-1/2} \left(\sum_{i \in \mathcal{I}} \sqrt{f_{i.}^g} \left(\frac{f_{ij}^g - f_{i.}^g f_{.j}^g}{f_{i.}^g f_{.j}^g} \right) \left(\sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g} F_s(i^g) \right) \right)$$

4.8. Aides à l'interprétation

Dans l'analyse proposée, on peut obtenir les mêmes aides à l'interprétation que dans les analyses factorielles habituelles.

4.8.1. Contributions des points à la formation des axes

Les contributions des points à la formation de l'axe s, $s \in \mathcal{S}$, sont calculées de la manière habituelle en divisant l'inertie projetée d'un point (poids par coordonnée au carré) par la somme des inerties de tous les points du nuage sur l'axe s, c'est-à-dire :

$$\mathrm{cta}_s(i^c) = \frac{p_{i^c}\ F_s^2(i^c)}{\lambda_s} \qquad \mathrm{cta}_s(j) = \frac{f_{.j}^g\ G_s^2(j)}{\lambda_s} \qquad \begin{array}{l} i \in \mathcal{I} \\ j \in \mathcal{J}_g \subset \mathcal{J} \\ g \in \mathcal{G} \end{array}$$

puisque :

$$\lambda_s = \sum_{j \in \mathcal{J}} f_{.j}^g\ G_s^2(j) = \sum_{i \in \mathcal{I}} p_{i^c}\ F_s^2(i^c) \qquad \begin{array}{l} g \in \mathcal{G} \\ s \in \mathcal{S} \end{array}$$

4.8.2. Qualité de représentation des points sur les axes

La qualité de représentation d'un point sur l'axe s, $s \in \mathcal{S}$, se mesure par les contributions relatives. On les calcule par le quotient de l'inertie projetée du point sur l'axe s, $s \in \mathcal{S}$, sur l'inertie totale du point ou par le carré de la projection sur l'axe

$s,\ s \in \mathcal{S}$, sur la distance, au carré, du point à l'origine :

$$\mathrm{ctr}_s(i^c) = \frac{F_s^2(i^c)}{d^2(i^c,0)} \qquad \mathrm{ctr}_s(j) = \frac{G_s^2(j)}{d^2(j,0)} \qquad \begin{array}{c} i \in \mathcal{I} \\ j \in \mathcal{J} \\ g \in \mathcal{G} \end{array}$$

avec pour distance, au carré, de la colonne $j,\ j \in \mathcal{J}$ à l'origine :

$$d^2(j,0) = \sum_{i \in \mathcal{I}} \alpha^g \left(\frac{f_{ij}^g}{f_{.j}^g} - f_{i.}^g \right)^2 \frac{1}{f_{i.}^g} = \sum_s G_s^2(j) \qquad g \in \mathcal{G}$$

et pour la distance, au carré, du compromis à l'origine celle trouvée en (5).

4.9. Cas particulier : $f_{i.}^g$ ne dépend pas du tableau g

Un cas particulier de la présente analyse se présente quand $f_{i.}^g,\ i \in \mathcal{I},\ g \in \mathcal{G}$, coïncide dans les g tableaux, $f_{i.}^g = f_{i.}$, seul cas envisagé pour l'analyse factorielle multiple par Escofier & Pagès (1998).

Alors l'analyse des correspondances du tableau pondéré :

$$\delta_{IJ} = \left(\beta_1 f_{IJ_1}^1, \ldots, \beta_g f_{IJ_g}^g, \ldots \beta_G f_{IJ_G}^G \right)$$

où le tableau de fréquence $f_{IJ_g}^g,\ g \in \mathcal{G}$ (correspondant au tableau $g \in \mathcal{G}$ et dont le terme général est $f_{ij}^g,\ i \in \mathcal{I},\ j \in \mathcal{J}_g \subset \mathcal{J},\ g \in \mathcal{G}$) a été pondéré par le coefficient $\beta_g,\ g \in \mathcal{G}$, est similaire à l'analyse proposée dans cet article si on a $\frac{\beta_g}{\beta} = \alpha_g$, avec $\beta = \sum_{g \in \mathcal{G}} \beta_g$, ce que nous supposerons dans la suite de ce paragraphe.

L'étude théorique des relations entre les deux analyses est facilitée si l'on considére l'analyse des correspondances dans la forme spécifiée au § 3, la matrice X_* définie à partir de (1) pour le tableau δ_{IJ} étant identique à la matrice X donnée par (6).

Dans l'analyse que nous proposons, le compromis est ici $i^c = \frac{1}{G} \sum_{g \in \mathcal{G}} i^g$, $i \in \mathcal{I}$, avec le poids [2] $p_{i^c} = G^2 f_{i.}$, tandis que dans l'analyse des correspondances du tableau δ_{IJ}, ce compromis qu'on notera simplement i, correspond à la $i-$ème ligne (en fait à son profil) de δ_{IJ} avec le poids $f_{i.}$.

Le profil partiel i^g de i dans l'analyse des correspondances de δ_{IJ} correspond à la ligne i de δ_{IJ} dont on ne conserve que les termes $\{\delta_{ij} \mid j \in \mathcal{J}_g\}$, les autres termes étant annulés, ligne dont on projette le profil en supplémentaire.

Notons $F'_s(i)$, $F'_s(i^g)$ et $G'_s(j)$ les projections de i, i^g et j ($j \in \mathcal{J}_g$) sur le s-ième axe factoriel issu de l'analyse des correspondances de δ_{IJ}.

[2] voir note au bas de la page 53.

On a, en conservant les notations du § 4.5 :

$$
\begin{aligned}
F'_s(i) &= GF_s(i^c) && i \in \mathcal{I} \\
G'_s(j) &= G_s(j)/\sqrt{\alpha^g} && j \in \mathcal{J}_g,\ g \in \mathcal{G} \\
F'_s(i^g) &= F_s(i^g)/\ \alpha^g && i \in \mathcal{I},\ g \in \mathcal{G} \\
F'_s(i) &= \sum_{g \in \mathcal{G}} \alpha^g\ F'_s(i^g) \\
F_s(i^c) &= \frac{1}{G} \sum_{g \in \mathcal{G}} F_s(i^g)
\end{aligned}
$$

Dans l'analyse des correspondances de δ_{IJ}, i est le barycentre des i^g pondérés par les α^g (puisque $\sum_{g \in \mathcal{G}} \alpha^g = 1$) tandis que dans notre analyse i^c est le centre de gravité des i^g affectés des masses $1/G$.

Les différences entre notre analyse et l'analyse des correspondances de δ_{IJ} proviennent essentiellement du fait que dans cette dernière l'introduction de la pondération β_g maintient les poids du profils-lignes mais modifie les poids des profils-colonnes (dans le tableau δ_{IJ} par rapport aux tableaux $f^g_{IJ_g}$).

5. Exemple d'Application

Pour appliquer la méthode proposée, dans ce travail, méthode que nous appellerons analyse simultanée, et la comparer à certaines des méthodes existantes, notamment dans ce cas l'analyse de la juxtaposition et l'analyse intra, on a sélectionné les données de l'exemple Shoplifting 1977/1978 d'Israëls (1987).

Dans cet exemple on considère les deux tableaux de contingence, un pour les hommes et un autre pour les femmes, dérivés de la table ternaire qui répartit une population de suspects de voler suivant les critères du sexe, de l'âge des personnes et du type d'article volé.

Les catégories d'âge considérées pour les deux sexes sont : <12 (moins de 12 ans), 12-14, 15-17, 18-20, 21-29, 30-39, 40-49, 50-64, 65+ (65 ans ou plus). Les types d'articles volés sont identifiés comme : CLOT : articles d'habillement, CLAC : accessoires d'habillement, TOBA : tabac, WRIT : matériel d'écriture, BOOK : livres, RECO : disques-musique, HOUS : appareils ménagers, SWEE : bonbons, TOYS : jouets, JEWE : bijoux, PERF : parfum, HOBB : hobbies, et OTHE : autres.

Tableau concernant les hommes

	<12	12-14	15-17	18-20	21-29	30-39	40-49	50-64	65+	tot
clot	81	138	304	384	942	359	178	137	45	2568
clac	66	204	193	149	297	109	53	68	28	1167
toba	150	340	229	151	313	136	121	171	145	1756
writ	667	1409	527	84	92	36	36	37	17	2905
book	67	259	258	146	251	96	48	56	41	1222
reco	24	272	368	141	167	67	29	27	7	1102
hous	47	117	98	61	193	75	50	55	29	725
swee	430	637	246	40	30	11	5	17	28	1444
toys	743	684	116	13	16	16	6	3	8	1605
jewe	132	408	298	71	130	31	14	11	10	1105
perf	32	57	61	52	111	54	41	50	28	486
hobb	197	547	402	138	280	200	152	211	111	2238
othe	209	550	454	252	624	195	88	90	34	2496
tot	2845	5622	3554	1682	3446	1385	821	933	531	20819

Tableau concernant les femmes

	<12	12-14	15-17	18-20	21-29	30-39	40-49	50-64	65+	tot
clot	71	241	477	436	1180	1009	517	488	173	4592
clac	19	98	114	108	207	165	102	127	64	1004
toba	59	111	58	76	132	121	93	214	215	1079
writ	224	346	91	18	30	27	23	27	13	799
book	19	60	50	32	61	43	31	57	44	397
reco	7	32	27	12	12	9	7	13	0	119
hous	22	29	41	32	65	74	51	79	39	432
swee	137	240	80	12	16	14	10	23	42	574
toys	113	98	14	10	21	31	8	17	6	318
jewe	162	548	303	74	100	48	22	26	12	1295
perf	70	178	141	70	104	81	46	69	41	800
hobb	15	29	9	14	30	36	24	35	11	203
othe	24	58	72	67	157	107	66	64	55	670
tot	942	2068	1477	961	2115	1765	1000	1239	715	12282

5.1. Analyses séparées de deux tableaux : type d'article volé par âge

Afin d'étudier les relations existantes entre les deux variables nominales, type d'article volé et âge, selon le sexe, on a appliqué l'analyse de correspondances à chaque tableau pris séparément. Sur la figure 1, on montre les plans factoriels formés par les deux premiers axes de l'analyse du tableau des hommes et de l'analyse du tableau des femmes. Les deux plans représentent 88.55% et 91.15% des inerties, 0.3952 et 0.4244, des tableaux respectifs, la première valeur propre de la première analyse valant 0.3047, tandis que celle de la seconde vaut 0.3231, l'inverse de ces valeurs propres servant de pondération dans l'analyse simultanée présentée au § 5.3. En guise de brève interprétation, on peut dire que dans les deux plans on observe que les catégories d'âge se projettent suivant une trajectoire curviligne.

Le premier facteur, dans les deux analyses, montre comment les personnes les plus jeunes, jusqu'à 14 ans, sont attirées, fondamentalement, par le vol de jouets (TOYS), de bonbons (SWEET) et de matériel d'écriture (WRIT) par opposition aux personnes d'âge intermédiaire, qui se sentent plus attirées, au moment de voler, par des articles d'habillement (CLOT).

Le second facteur, dans le cas du tableau des hommes, montre fondamentalement que l'intérêt pour le vol de bijoux (JEWE) et de disques (RECO) est faible parmi les enfants de moins de 12 ans et les adultes de plus de 65 ans et élevé chez les adolescents de 15 à 17 ans.

Dans le cas de l'analyse du tableau des femmes, le second facteur met en évidence la forte attirance que le vol de tabac (TOBA) exerce parmi les femmes de plus de 65 ans, alors que celles-ci, à leur tour, ne sont nullement attirées par le vol d'articles d'habillement (CLOT) et de bijoux (JEWE) plus propre aux femmes ayant entre 15 et 17 ans.

5.2. Analyse de la juxtaposition des deux tableaux et de leur somme

À ce point nous renvoyons le lecteur à l'ouvrage cité d'Israëls (1987) pour l'interprétation de l'analyse du tableau 13 x 18 résultant de la juxtaposition des deux tableaux. Nous mentionnerons brièvement que le premier facteur discrimine entre les catégories d'âge, mettant en opposition les mineurs de moins de 14 ans et les personnes d'âge intermédiaire et le second entre les modalités de la variable sexe, associant aux femmes les vols d'habillement et de bijoux et montrant que les vols de hobbies sont fondamentalement commis par les hommes.

Il est bien connu que, dans cette analyse, l'inertie totale du tableau juxtaposé peut se décomposer en une inertie inter-niveaux d'âge et intra-niveaux d'âge, ainsi qu'en une inertie inter-tableaux et intra-tableaux.

Bien que dans l'exemple analysé chaque axe corresponde fondamentalement à un type d'inertie (intra-tableaux sur l'axe 1 et inter-tableaux sur l'axe 2), il est courant de trouver des axes qui prennent en compte les deux types d'inertie, en compliquant leur interprétation. L'analyse de l'inertie inter-tableaux peut être réalisée par l'analyse du tableau somme. En ce cas, il correspond à un tableau 13 (articles volés) x 2 (hommes et femmes) qui ne présente pas d'intérêt.

L'analyse du tableau 13x9, somme des deux tableaux des hommes et des femmes, avec ces deux derniers tableaux comme supplémentaires, ne répond pas à l'objectif que recherche la méthode proposée dans ce travail. Toutefois, sachant qu'il s'agit d'une méthode classique dans la comparaison de tableaux binaires, nous nous livrerons à un bref commentaire la concernant. Cette analyse permet d'étudier l'inertie inter-catégories d'âge étant donnée que l'on analyse le nuage moyen des profils-colonnes correspondant à la même catégorie d'âge.

L'interprétation des deux premiers axes est très similaire à l'interprétation réalisée dans l'analyse du tableau des hommes (et coïncide avec l'interprétation de l'analyse du tableau des femmes sur le premier axe). Ces résultats montrent comment le tableau des hommes, dont l'effectif est environ le double de celui correspondant au tableau des femmes, exerce une plus grande influence dans l'analyse du tableau somme, situation que l'on essaie d'éviter avec la méthode proposée.

La projection, comme éléments supplémentaires, des profils-lignes et des profils-colonnes des tableaux des hommes et des femmes met en évidence des grandes différences entre les projections de profils homologues. Les différences entre les tableaux ne sont pas nettement mises en évidence par cette technique.

5.3. Analyse Intra et Analyse Simultanée

Dans le but d'étudier la partie intra de l'inertie du tableau juxtaposé, on voit surgir l'analyse intra (Escofier (1983)). Dans cette analyse, la partie inter de l'inertie s'élimine pour considérer chaque sous-nuage de profils-colonne du même tableau (sous-nuage de colonnes des hommes, d'une part, et sous-nuage des femmes, d'autre part) centré sur son propre centre de gravité.

Le nuage ainsi formé maintient les distances intra entre colonnes. Il faut noter, toutefois, que cette distance est calculée à partir de la métrique définie par la marginale du tableau juxtaposé, et ne coïncide pas avec les distances calculées dans l'analyse de chaque tableau pris séparément. De plus, les poids assignés à ces profils ne coïncident pas non plus dans l'analyse intra (égaux aux poids utilisés dans l'analyse du tableau juxtaposé) avec ceux assignés dans l'analyse de chaque tableau pris séparément. La conséquence en est que l'identité de chaque tableau se perd si l'on réalise l'analyse sur les marginales du tableau juxtaposé.

Dans l'analyse simultanée, en dehors du fait que les sous-nuages sont centrés par rapport à leur propre centre de gravité, chaque tableau participe avec ses propres poids et sa propre métrique. Cette méthode prend en compte les effectifs différents de chaque tableau et les marginales différentes que les types d'articles volés ont pour les hommes et les femmes.

La figure 1 montre aussi les premiers plans factoriels des deux analyses (intra et simultanée) dans lesquels les pourcentages d'inertie sont respectivement 82.55% et 83.79%. La ligne continue représente la trajectoire d'âge des femmes et la ligne discontinue celle des hommes. On observe que dans le plan factoriel de l'analyse simultanée se maintiennent avec une fidélité non négligeable les relations internes à chacun des tableaux.

Les différences qui s'apprécient entre les deux plans sont dues à l'utilisation, ainsi qu'il a été dit, de métriques et de poids différents qui influent entre autres choses dans le calcul des distances et, par conséquent, des inerties des points.

Dans l'exemple, les différences s'apprécient sur le second facteur. Dans l'analyse intra, ce facteur met en évidence l'attirance des femmes de 21 à 39 ans, pour le vol d'habillement (CLOT). Le second facteur de l'analyse simultanée met en évidence l'inclination au vol de tabac (TOBA) de la part des femmes de plus de 65 ans et non vers celui des bijoux (JEWE) et disques-musique (RECO), plus propre des femmes de 12 à 17 ans, ce dernier fait étant observé dans l'analyse séparée du tableau relatif aux femmes.

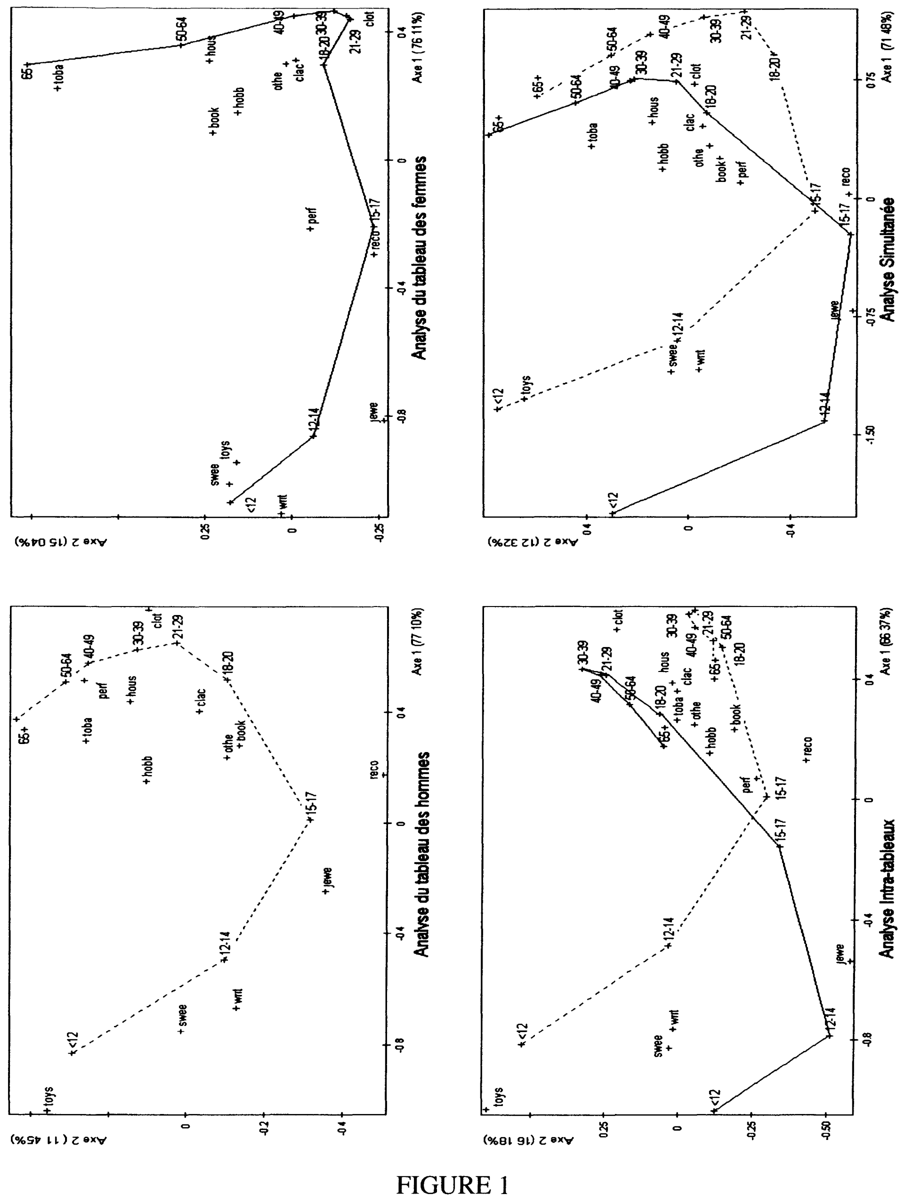

FIGURE 1
Plans factoriels 1-2 issus des analyses

En effet, dans l'analyse du tableau des femmes et dans l'analyse simultanée, les catégories d'âge des femmes 21-29 et 30-39 (associées au vol d'habillement) participent avec moins d'inertie que la catégorie 65+ (associée au vol de tabac). Dans l'analyse intra, on assiste au phénomène inverse.

Et, en conséquence, c'est ce que reflètent les axes factoriels.

Quant à la pondération qui équilibre l'influence des tableaux dans l'analyse simultanée, nous signalerons que, étant donné que les valeurs propres associées au premier facteur sont similaires dans les deux tableaux ($\lambda_1 = 0.3047$ pour les hommes et $\lambda_1 = 0.3231$ pour les femmes, et donc $\alpha_1 = 3.2822$ et $\alpha_2 = 3.0954$), aucun tableau n'aura d'influence prépondérante dans la création des axes factoriels et les résultats ne différeront pas, dans ce cas, de ceux de l'analyse simultanée sans l'introduction de la pondération.

Les différences que l'on observe entre l'analyse intra et l'analyse simultanée proposée sont donc dues, dans cet exemple, quasi exclusivement aux différences de métriques et de poids.

6. Conclusion

- La méthode proposée permet l'analyse simultanée de plusieurs tableaux de contingence en conservant pour chacun d'eux leur structure et leur spécificité.
- Elle coïncide dans le cas où on l'applique à un seul tableau de contingence, avec l'analyse des correspondances simples. En l'appliquant à un tableau disjonctif complet on obtient les mêmes resultats que l'analyse des correspondances multiples classique. Et quand le tableau à analyser est disjonctif incomplet, la méthode donne les resultats de l'analyse des correspondances avec marginale imposée (Zárraga & Goitisolo (2000)).

Remerciements : Ce travail a été financé par le projet d'investigation UPV 038.321-HA041/99 de l'Université du Pays Basque (UPV/EHU) et le projet PB98-0149 de la Direction Génerale de l'Enseignement supérieur et de la Recherche Scientifique du Ministère Espagnol de l'Éducation et de la Culture.

Références

BÉCUE-BERTAUT, Mónica MARÍA & PAGÈS J. (2000), Analyse factorielle multiple intra-tableaux. Application à l'analyse simultanée de plusieurs questions ouvertes, *in* «JADT 2000 : 5e Journées Internationales d'Analyse Statistique des Données Textuelles».

BENZÉCRI J.P. & collaborateurs (1973), *L'analyse des données. 2 : L'analyse des correspondances*, Dunod.

CARROLL J.D. (1968), «Generalization of canonical correlation analysis to three or more sets of variables», *Proc. Am. Psychol Ass.*, pp. 227-228.

CARROLL J.D. & CHANG J.J. (1970), «Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition», *Psychometrika*, **35**(3), 283-319.

CAZES P. (1980a), «L'analyse de certains tableaux rectangulaires décomposés en blocs : généralisation des propriétés rencontrées dans l'étude des correspondances multiples. I. Définitions et applications à l'analyse canonique des variables qualitativesfg, *Les Cahiers de l'analyse des données*, **V**(2), 145-161.

CAZES P. (1980b), «L'analyse de certains tableaux rectangulaires décomposés en blocs : généralisation des propriétés rencontrées dans l'étude des correspondances multiples. II. Questionnaires : variantes de codages et nouveaux calculs de contributions», *Les Cahiers de l'analyse des données*, **V**(4), 387-403.

CAZES P. (1981a), «L'analyse de certains tableaux rectangulaires décomposés en blocs : généralisation des propriétés rencontrées dans l'étude des correspondances multiples. III. Codage simultané de variables qualitatives et quantitatives», *Les Cahiers de l'analyse des données*, **VI**(1), 9-18.

CAZES P. (1981b), «L'analyse de certains tableaux rectangulaires décomposés en blocs : généralisation des propriétés rencontrées dans l'étude des correspondances multiples. IV. Cas modèles», *Les Cahiers de l'analyse des données*, **VI**(2), 135-143.

D'AMBRA L. & LAURO N. (1989), «Non symetrical analysis of three-way contingency tables», *Multiway Data Analysis*, pp. 301-315.

ESCOFIER B. (1983), «Généralisation de l'analyse des correspondances à la comparaison de tableaux de fréquence», *INRIA*, **207**, 1-33.

ESCOFIER B. & PAGÈS J. (1984), «L'analyse factorielle multiple», *Cahiers du Bureau Universitaire de Recherche Opérationnelle*, **42**, 1-68.

ESCOFIER B. & PAGÈS J. (1988), *Analyses factorielles simples et multiples. Objectifs, méthodes et interprétation*, Dunod.

ESCOFIER B. & PAGÈS J. (1998), *Analyses factorielles simples et multiples. Objectifs, méthodes et interprétation*, 3nd, Dunod.

GOWER J.C. (1975), «Generalized procrustes analysis», *Psychometrika*, **40**(1), 33-51.

GOWER J.C. (1984), Procrustes analysis, *in* C.J. Wiley, ed., «Handbook of Applicable Mathematics, 6», Lloyd E.H.

HOTELLING H. (1936), «Relation between two sets of variables», *Biometrika*, **28**, 129-149.

ISRAËLS A. (1987), *Eigenvalue Techniques for Qualitative Data*, DSWO Press.

LAURO N. & D'AMBRA L. (1984), «L'analyse non symétrique des correspondances», *Data Analysis and Informatics, III*, pp. 433-446.

LEBART L., MORINEAU A. & FENELON J.P. (1985), *Tratamiento estadístico de datos*, Marcombo.

LECLERC A. (1973), Étude de certains types de tableaux par l'analyse des correspondances, Thèse de 3^{e} cycle, Université de Paris VI.

L'HERMIER DES PLANTES H. (1976), *STATIS : Structuration de Tableaux à trois indices de la statistique*, Thèse (3c), USTL, Montpellier.

TUCKER L.R. (1958), « An inter-battery method of factor analysis », *Psychometrika*.

TUCKER L.R. (1972), « Relations between multidimensional scaling and three-mode factor analysis », *Psychometrika*, **37**(1), 3-27.

ZÁRRAGA A. & GOITISOLO B. (2000), « Estudio comparativo de análisis alternativos de tablas disyuntivas incompletas », *Documentos de Trabajo. Biltoki*, **8**, 1-29.

Annexe
Résultats de l'analyse simultanée

LÉGENDE :

- H : homme
- F : femme
- d^2 : distance, au carré, des points à l'origine
- p_i, p_j : poids des profils-lignes et profils-colonnes respectivement $i \in \mathcal{I},\ j \in \mathcal{J}$
- $F_s, G_s : s = 1, 2, 3$, coordonnées des projections des profils-lignes, profils-colonnes respectivement sur les trois premiers axes
- $A_s : s = 1, 2, 3$, contributions des points à la formation des axes
- $R_s : s = 1, 2, 3$, qualité de représentation des points sur les axes

INERTIE TOTALE : 2.610824 [3]

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE
1	1.8661	71.48	71.48
2	0.3216	12.32	83.79
3	0.2659	10.18	93.98
4	0.0612	2.35	96.32
5	0.0488	1.87	98.19
6	0.0212	0.81	99.00
7	0.0137	0.52	99.53
8	0.0047	0.18	99.70
9	0.0036	0.14	99.84
...			

[3] $2.610824 = 0.3952\frac{1}{0.3047} + 0.4244\frac{1}{0.3231}$

On remarquera que la première valeur propre est supérieure à l'unité (mais inférieure à $G = 2$) du fait de la pondération retenue. Comme on l'a noté à la page 51 il suffit de diviser par $\alpha = \alpha_1 + \alpha_2 = 3.2822 + 3.0954 = 6.3776$ pour obtenir les valeurs propres $0.2926, 0.0504, 0.0417, \ldots$ d'une inertie totale de 0.4094, valeurs comparables en ordre de grandeur à celles obtenues dans les analyses séparées des tableaux.

RÉSULTATS POUR LES COMPROMIS										
	d^2	F_1	F_2	F_3	A_1	A_2	A_3	R_1	R_2	R_3
clot	0.59	0.72	-0.02	-0.25	25.87	0.13	21.03	0.88	0.00	0.10
clac	0.21	0.45	-0.05	0.00	2.94	0.23	0.00	0.94	0.01	0.00
toba	0.63	0.32	0.38	0.60	1.91	15.87	46.96	0.16	0.24	0.58
writ	1.21	-1.08	-0.04	-0.02	24.81	0.18	0.07	0.97	0.00	0.00
book	0.16	0.24	-0.12	0.24	0.57	0.82	3.71	0.37	0.09	0.34
reco	0.58	0.03	-0.63	0.13	0.00	13.23	0.65	0.00	0.68	0.03
hous	0.32	0.47	0.14	0.17	1.66	0.91	1.51	0.68	0.06	0.09
swee	1.24	-1.10	0.07	0.08	14.89	0.32	0.57	0.98	0.00	0.01
toys	2.25	-1.27	0.65	-0.43	16.56	25.03	13.60	0.71	0.19	0.08
jewe	0.97	-0.71	-0.65	0.11	8.41	39.90	1.49	0.52	0.43	0.01
perf	0.26	0.09	-0.20	0.23	0.07	2.03	3.37	0.03	0.15	0.20
hobb	0.27	0.18	0.10	0.29	0.35	0.70	6.48	0.12	0.04	0.31
othe	0.16	0.33	-0.08	-0.07	1.94	0.64	0.57	0.69	0.04	0.03

RÉSULTATS POUR LES PROFILS-LIGNES PARTIELS								
	p_i	d^2	F_1	F_2	F_3	R_1	R_2	R_3
Hclot	12.33	2.23	1.00	-0.42	0.04	0.45	0.08	0.00
Hclac	5.61	0.56	0.51	-0.30	0.11	0.47	0.16	0.02
Htoba	8.43	0.87	0.38	0.09	0.49	0.17	0.01	0.27
Hwrit	13.95	1.57	-0.87	0.31	-0.27	0.48	0.06	0.05
Hbook	5.87	0.34	0.34	-0.32	0.17	0.35	0.31	0.09
Hreco	5.29	1.02	0.18	-0.64	0.15	0.03	0.40	0.02
Hhous	3.48	0.73	0.57	-0.14	0.22	0.44	0.03	0.07
Hswee	6.94	1.87	-0.96	0.52	-0.34	0.49	0.14	0.06
Htoys	7.71	4.19	-1.29	1.06	-0.76	0.39	0.27	0.14
Hjewe	5.31	0.63	-0.34	-0.20	-0.08	0.19	0.07	0.01
Hperf	2.33	1.20	0.67	-0.06	0.41	0.37	0.00	0.14
Hhobb	10.75	0.46	0.18	0.02	0.39	0.07	0.00	0.33
Hothe	11.99	0.29	0.30	-0.27	-0.04	0.31	0.25	0.01
Fclot	37.39	0.73	0.56	0.21	-0.41	0.43	0.06	0.23
Fclac	8.17	0.33	0.39	0.15	-0.10	0.47	0.07	0.03
Ftoba	8.79	1.63	0.26	0.68	0.71	0.04	0.28	0.31
Fwrit	6.51	4.01	-1.40	-0.55	0.34	0.49	0.08	0.03
Fbook	3.23	0.28	0.11	0.15	0.32	0.04	0.08	0.36
Freco	0.97	0.90	-0.33	-0.60	0.08	0.12	0.40	0.01
Fhous	3.52	0.57	0.37	0.43	0.12	0.25	0.32	0.02
Fswee	4.67	3.31	-1.27	-0.48	0.59	0.49	0.07	0.11
Ftoys	2.59	4.25	-1.24	-0.06	0.13	0.36	0.00	0.00
Fjewe	10.54	2.53	-0.98	-0.96	0.25	0.38	0.36	0.02
Fperf	6.51	0.24	-0.25	-0.28	0.13	0.26	0.32	0.07
Fhobb	1.65	0.41	0.17	0.33	0.02	0.07	0.26	0.00
Fothe	5.46	0.33	0.37	0.20	-0.11	0.42	0.12	0.04

RÉSULTATS POUR LES PROFILS-COLONNES											
	p_j	d^2	G_1	G_2	G_3	A_1	A_2	A_3	R_1	R_2	R_3
H <12	13.67	2.66	-1.33	0.75	-0.49	13.04	24.18	12.40	0.67	0.21	0.09
H12-14	27.00	0.86	-0.90	0.05	-0.02	11.78	0.17	0.04	0.94	0.00	0.00
H15-17	17.07	0.38	-0.08	-0.49	0.17	0.07	12.89	1.95	0.02	0.64	0.08
H18-20	8.08	1.01	0.90	-0.33	-0.05	3.54	2.79	0.08	0.81	0.11	0.00
H21-29	16.55	1.55	1.17	-0.21	-0.18	12.20	2.35	2.12	0.88	0.03	0.02
H30-39	6.65	1.38	1.14	-0.06	-0.04	4.65	0.06	0.04	0.94	0.00	0.00
H40-49	3.94	1.45	1.03	0.15	0.35	2.26	0.29	1.85	0.73	0.02	0.09
H50-64	4.48	1.77	0.90	0.30	0.75	1.94	1.28	9.45	0.45	0.05	0.32
H 65+	2.55	2.62	0.64	0.60	1.27	0.56	2.86	15.57	0.16	0.14	0.62
F <12	7.67	4.24	-2.00	0.29	-0.03	16.47	2.07	0.02	0.95	0.02	0.00
F12-14	16.84	2.37	-1.41	-0.53	0.25	17.93	14.77	4.10	0.84	0.12	0.03
F15-17	12.03	0.53	-0.23	-0.64	0.02	0.34	15.17	0.02	0.10	0.77	0.00
F18-20	7.82	0.34	0.54	-0.07	-0.11	1.21	0.11	0.37	0.84	0.01	0.04
F21-29	17.22	0.71	0.73	0.05	-0.39	4.96	0.14	9.98	0.76	0.00	0.22
F30-39	14.37	0.79	0.75	0.21	-0.40	4.29	2.00	8.68	0.71	0.06	0.20
F40-49	8.14	0.66	0.74	0.23	-0.16	2.38	1.31	0.78	0.82	0.08	0.04
F50-64	10.09	0.76	0.59	0.45	0.39	1.91	6.27	5.85	0.46	0.26	0.20
F 65+	5.82	2.30	0.40	0.79	1.10	0.49	11.27	26.71	0.07	0.27	0.53