ON THE CONNECTION BETWEEN THE DISTRIBUTION OF EIGENVALUES IN MULTIPLE CORRESPONDENCE ANALYSIS AND LOG-LINEAR MODELS *

Authors: S. Ben Ammou
– Department of Quantitative Methods, Faculté de Droit et des Sciences Economiques et Politiques de Sousse, Tunisia, (saloua.benammou@fdseps.rnu.tn)

G. Saporta
– Chaire de Statistique Appliquée & CEDRIC, Conservatoire National des Arts et Métiers, Paris, France (saporta@cnam.fr)

Received: December 2002 Revised: September 2003 Accepted: October 2003

Abstract:

- Multiple Correspondence Analysis (MCA) and log-linear modeling are two techniques for multi-way contingency table analysis having different approaches and fields of applications. Log-linear models are interesting when applied to a small number of variables. Multiple Correspondence Analysis is useful in large tables. This efficiency is balanced by the fact that MCA is not able to explicit the relations between more than two variables, as can be done through log-linear modeling. The two approaches are complementary. We present in this paper the distribution of eigenvalues in MCA when the data fit a known log-linear model, then we construct this model by successive applications of MCA. We also propose an empirical procedure, fitting progressively the log-linear model where the fitting criterion is based on eigenvalue diagrams. The procedure is validated on several sets of data used in the literature.

Key-Words:

- *Multiple Correspondence Analysis; eigenvalues; log-linear models; graphical models; normal distribution.*

AMS Subject Classification:

- 49A05, 78B26.

---

*We thank Professor M. Bourdeau for his careful reading.

1. INTRODUCTION

Multiple Correspondence Analysis and log-linear modeling are two very different, but mutually beneficial approaches to analyzing multi-way contingency tables: log-linear models are profitably applied to a small number of variables. Multiple Correspondence Analysis is useful in large tables. This efficiency is balanced by the fact that MCA is not able to explicit relations between more than two variables, as can be done through log-linear modeling. The two approaches are complementary. After a short reminder on MCA and log-linear approaches, we study the distribution of eigenvalues in MCA under modeling hypotheses, especially in the case of independence. At the end we propose an algorithmic approach for fitting log-linear models where the fitting criterion is based on eigenvalues diagram.

2. A SHORT SURVEY OF MULTIPLE CORRESPONDENCE ANALYSIS AND LOG-LINEAR MODELS

We first introduce MCA and log-linear modelling, then we present some works using both methods.

2.1. Multiple Correspondence Analysis

Correspondence Analysis (CA) has quite a long history as a method for the analysis of categorical data. The starting point of this history is usually set in 1935 [28], and since then CA has been reinvented several times. We can distinguish simple CA (CA of contingency tables) and MCA or Multiple Correspondence Analysis (CA of so-called indicator matrices). MCA traces back to Guttman [23], Burt [8] or Hayashi [25]. In France, in the 1960s, Benzecri [6] proposes, other developments of this method. Outside France, MCA has been developed by J. de Leeuw since 1973 [22] under the name of Homogeneity Analysis, and the name of Dual Scaling by Nishisato [38].

Multiple Correspondence Analysis (MCA) is a multidimensional descriptive technique of categorical data. A theoretical version of Multiple Correspondence Analysis of p variables can be defined as the limit, when the number of statistical units increases, of the CA of a complete disjunctive table.

Let X be a complete disjunctive table of p categorical variables $X_1, X_2, ...,$ X_p, with respectively $m_1, m_2, ..., m_p$ modalities observed over a sample of n individuals. CA of this complete disjunctive table is equivalent to the analysis of B [8], where $B = X'X$ is the Burt table associated with X. The two analyses have the same factors, but the eigenvalues in MCA equal to the squared root of the eigenvalues in the CA of the associated Burt table. MCA of X corresponds to the diagonalization of the matrix $\frac{1}{p}(D^{-1}X'X) = \frac{1}{p}(D^{-1}B)$ where $D = \mathrm{Diag}(X'X) = \mathrm{Diag}(B)$.

The structure of the eigenvalue diagram depends on the variable interactions. It is well known that in the case of pairwise independent variables, the q non-trivial eigenvalues are theoretically equal to to $\frac{1}{p}$, where

$$q = \sum_{i=1}^{p} m_i - p \; . \tag{1}$$

2.2. Log-linear modeling

Log-linear modeling is a well-known method for studying structural relationships between categorical variables in a multiple contingency table when all the variables have no particular role. Relatively recent and not as well known in France as MCA, log-linear modeling has many classical references. After first works of Birch [7] in 1963 and Goodman [17], we must mention the basic books of Haberman [24], Bishop, Fienberg & Holland [8], Fienberg [15].

More Recently, Dobson [12], Agresti [1], Christensen [10] have written syntheses on the subject supplemented with personal contributions.

Whittaker [41] devotes a large part of his book to log-linear models before defining associated graphical models.

2.2.1. Log-linear modeling in the binomial case

Let $X = (X_1, X_2, ..., X_p)$ be a k-dimensional random vector, with values in $\{0,1\}^k$. The expression for the k-dimensional probability density of X is:

$$\begin{aligned} f_k(X) \;=\; & p(0,0,...,0)^{(1-x_1)(1-x_2)\cdots(1-x_k)} \cdot p(1,0,...,0)^{x_1(1-x_2)\cdots(1-x_k)} \\ & \cdot\, p(0,1,...,0)^{(1-x_1)x_2\cdots(1-x_k)} \cdots\, p(0,0,...,1)^{(1-x_1)(1-x_2)\cdots x_k} \\ & \cdots\, p(1,1,...,0)^{x_1x_2\cdots(1-x_k)} \cdots\, p(1,1,...,1)^{x_1x_2\cdots x_k} \; . \end{aligned}$$

We can write the density function as a log-linear expansion:

$$\begin{aligned} \log[f_k(X)] \;=\; & u_o + \sum_{i=1}^{k} u_i\, x_i + \sum_{\substack{i,j=1,\\ i\neq j}}^{k} u_{ij}\, x_i\, x_j + \sum_{\substack{i,j,\ l=1,\\ i\neq j\neq l}}^{k} u_{ijl}\, x_i\, x_j\, x_l \\ & + \cdots + u_{123\ldots k}\, x_1\, x_2 \cdots x_k \end{aligned}$$

where $u_o = \log[p(0,0,...,0)]$, $u_i = \log[\frac{p(0,0,...,0,1,0,...0)}{p(0,0,...,0)}]$ and the u-terms $u_{ij}, ..., u_{123\ldots k}$ are a log cross product ratio in the (k,k) probability table. The u-term u_{ij} is set to zero when X_i and X_j are independent variables.

2.2.2. Log-linear modeling in the multinomial case

Let $X = (X_1, X_2, ..., X_k)$ be a k-dimensional random vector, with values in $\{0, 1, ..., m_1-1\} \times \{0, 1, ..., m_2-1\} \times ... \times \{0, 1, ..., m_k-1\}$ instead of in $\{0, 1\}^k$ as in the preceding case.

The generalisation to the k-dimensional cross-classified multinomial distribution is the log-linear expansion:

$$\log[f_k(X)] \;=\; u_o + \sum_{i=1}^{k} u_i(x) + \sum_{\substack{i,j=1,\\ i\neq j}}^{k} u_{ij}(x) + \sum_{\substack{i,j,\ l=1,\\ i\neq j\neq l}}^{k} u_{ijl}(x) + \;\cdots\; + \; u_{123...k}(x)\ .$$

Each u-term is a coordinate projection function with the coordinates indicated by its index; and each u-term is constrained to be zero whenever one of its indicated coordinates is zero.

The importance of log-linear expansions rests with the fact that many interesting hypotheses can be generated by setting some u-terms to zero.

We are interested particularly in this paper with graphical and hierarchical log-linear models.

2.2.2.1. Graphical log-linear models

Let $G = (K, E)$ be the independence graph of the k-dimensional random vector X, with k vertices in $K = \{1, 2, ..., k\}$ and edge set E. G is the set of pairs (i, j) such that whenever (i, j) is not in E the variables X_i and X_j are independent conditionally on the other variables.

Given an independence graph G, the cross classified multinomial distribution for the random vector X is a graphical model for X, if the distribution of X is different from constraints of the form that for all pair of coordinates not in the edge set E of G, the u-terms constraining the selected coordinates are identically zero.

2.2.2.2. Hierarchical log-linear models

A graphical model satisfies constraints of the form that all u-terms 'above' a fixed point have to be zero to get conditional independence. A larger class of models, the hierarchical models, is obtained by allowing more flexibility in setting the u-terms to zero.

A log-linear model is hierarchical if, whenever one particular u-term is constrained to zero then all higher u-terms containing the same set of subscripts are also set to zero.

We note here that every distribution with a log-linear expansion has an interaction (or independence) graph, and a hierarchical log-linear model is graphical if and only if its maximal u-terms correspond to cliques in the independence graph.

When all the u-terms are non-zero, we have the **saturated** model.

In the case when only the u_i are non-zero, the model is called the **mutual independence model**:

$$\log[f_k(X)] \;=\; u_o(x) \;+ \sum_{i=1}^{k} u_i(x) \;.$$

When only u_i and some of u_{ij} are non-zero, the model is called a **conditional independence model**:

$$\log[f_k(X)] \;=\; u_o(x) \;+ \sum_{i=1}^{k} u_i(x) \;+ \sum_{i,j} u_{ij}(x) \;.$$

These conditional independence models refer to simple interactions between some variables.

2.2.3. Parameters estimation and related tests

Theoretical frequencies are generally estimated using the maximum-likelihood method. Weighted regression, or iterative methods can be also used as well since log-linear models are particular cases of the generalized linear model. Usually the classical χ^2 or the G^2 tests (the likelihood ratio) are used to assess log-linear models. The values of the two statistics increase with the number of variables, and decrease with the number of interactions. The closer the statistics are to zero, the better the models.

Model selection becomes difficult when the number of variables grow: e.g. with four variables there are 167 different hierarchical models. To avoid the "combinatory explosion" we can use criterions based on the Kullback information like the Akaike criterion:

$$AIC \;=\; -2\log(\widehat{L}) + 2\,k \qquad (\textit{An Information criterion})\;,$$

or the Schwartz criterion:

$$BIC \;=\; -2\log(\widehat{L}) + k\log(n) \qquad (\textit{Bayesian Information criterion})\;,$$

where $\widehat{L}$ is the maximum of the likelihood function (L), and k the number of parameters maximising L.

2.3. Multiple Correspondence Analysis and log-linear model as complementary tools of analysis

In this section, we present some works that show how CA (or MCA) and log-linear modeling can be related. This leads to a better understanding of CA, and to a combined use of both methods.

CA is often introduced without any reference to other methods of statistical treatment of categorical data with proven usefulness and flexibility.

A major difference between CA and most other techniques for categorical data analysis lies in the use of probability models. In log-linear analysis (LLA), for example, a distribution is assumed under which the data are collected, then a log-linear model for the data is hypothesized and estimations are made under the assumption that this probability model is true, and finally these estimates are compared with the observed frequencies to evaluate the log-linear model. In this way it is possible to make inferences about the population on the basis of the sample data.

In CA, it is claimed that no underlying distribution has to be assumed and no model has to be hypothesized, but a decomposition of the data is obtained to study the 'structure' in the data.

A vast literature has been devoted to the assessment of CA (or MCA) and LLA. We briefly report here some of that literature.

Several works compare CA or MCA and LLA. Daudin and Trecourt [11] demonstrate empirically that LLA is better adapted to study global relationships between the variables, in the sense that marginal liaisons are eliminated in the computation of profiles.

Goodman [17],[18],[19],[20],[21] defines two models belonging to the same family: the saturated row column correspondence analysis model as a generalization of MCA, and the row column association model as a generalization of LLA. He demonstrates, with illustrations by examples, that using these models is better than using the classical methods.

Baccini, Mathieu and Mondot [3] use an example to compare the two methods. Jmel [30], De Falguerolles, Jmel and Whittaker [13],[14] use graphical models compared to MCA.

Other works use CA or MCA and LLA as a combined approach to contingency table analysis: Van der Heijden and de Leeuw [26],[27],[28], Novak and Hoffman [39] and others, use CA as a tool for the exploration of the residuals from log-linear models, and give an example of the procedure.

Worsley [42] shows that in certain cases CA leads directly to the appropriate log-linear model.

Lauro and Decarli [31] used AC as a procedure for the identification of best log-linear models.

3. EIGENVALUES IN CORRESPONDENCE ANALYSIS

It is well known that MCA is an extension of CA, although we first present eigenvalues in CA, and some simple rules for the selection of the number of eigenvalues.

3.1. Asymptotic distribution of eigenvalues in Correspondence Analysis

Let N be a contingency table with m_1 rows and m_2 columns, and let us assume that N is the realization of a multinomial distribution $M(n, p_{ij})$ which is realistic. In this framework the observed eigenvalues μ_i are estimators of the eigenvalues λ_i of nP, where P is the table of unknown joint probabilities.

Lebart [32] and O'Neill [34],[35],[36] proved the following result:
if $\mu_i = 0$ then λ_i has the same distribution as the corresponding eigenvalues of a $(m_1-1)(m_2-1)$ degrees of freedom from the Wishart matrix: $W_{(m_1-1)(m_2-1)}(r, l)$ where $r = \min(m_1-1,\, m_2-1)$.

If $\mu_j = 0$ then $\sqrt{\lambda_j}$ is asymptotically normally distributed, but with parameters depending on the unknown p_{ij}. Since it is difficult to test this hypothesis, some authors have proposed a bootstrap approach, which unfortunately is not valid: since the empirical eigenvalues, on which the replication is based, are generally not null, we cannot observe the distribution based on the Wishart matrix.

3.2. Malinvaud's test

Based upon the reconstitution formula, which is a weighted singular value decomposition of N:

$$n_{ij} \;=\; \frac{(n_{i\cdot}\, n_{\cdot j})}{n} \left(1 + \frac{\displaystyle\sum_k (a_{ik}\, b_{ki})}{\sqrt{\lambda_k}} \right) ,$$

where a_{ik}, b_{ik} are the factorial components associated to the row and column profiles.

We may use a chi-square test comparing the observed n_{ij} from a sample of size n to the expected frequencies under the null-hypothesis H_k of only k non zeros. The μ_i weighted least squares estimates of these expectations are precisely the $\widetilde{n_{ij}}$ of the reconstitution formula with its first k terms. We then compute the classical chi-square goodness of fit statistic:

$$Q_k = \sum_i \sum_j \frac{(\widetilde{n}_{ij} - n_{ij})^2}{\widetilde{n}_{ij}} .$$

If $k = 0$ (independence), Q_0 is compared to a chi-square with $(m_1 - 1)\,(m_2 - 1)$ degrees of freedom. Under H_k, Q_k is asymptotically distributed like a chi-square with $(m_1 - k - 1)\,(m_2 - k - 1)$ degrees of freedom. However Q_k suffers from the following drawback: if n_{ij} is small, $\widetilde{n}_{ij}$ can be negative and the test statistic cannot be used. This is not the case for the modification proposed by E. Malinvaud [37] who proposed to use $\frac{(n_{i.}\,n_{.j})}{n}$ instead of $\widetilde{n}_{ij}$ for the denominator. Furthermore, this leads to a simple relation with the sum of the discarded eigenvalues:

$$Q'_k = \sum_i \sum_j \frac{(\widetilde{n}_{ij} - n_{ij})^2}{\frac{(n_{i.} n_{.j})}{n}} = n\,(\lambda_{k+1} + \lambda_{k+2} + ... + \lambda_r) .$$

Q'_k is also asymptotically distributed like a chi-square with $(p - k - 1)\,(q - k - 1)$ degrees of freedom.

4. BEHAVIOUR OF EIGENVALUES IN MCA UNDER MODELING HYPOTHESES

Let $X = (X_1|X_2|...|X_p)$ be a disjunctive table of p categorical variables X_i (with respectively m_i modalities) observed on a sample of n individuals, and q the number of non trivial eigenvalues (as defined in § 2.1).

Multiple Correspondence Analysis is the CA of disjunctive table X.

The rank of X: $\operatorname{rank}(X) = \min(q{+}1; n)$, so equals $q{+}1$ if $n > q{+}1$.

We suppose, without loss of generality, that n is large enough, which is the usual case. CA factors are the eigenvectors of the matrix $\frac{1}{p}\,D^{-1}B$ (where B and D are defined in § 2.1). So $D^{-1}B$ is a diagonal unit matrix.

Its trace is: $\operatorname{Tr}(D^{-1}B) = \sum_{i=1}^{p} m_i$ and $\frac{1}{p}\operatorname{Tr}(D^{-1}B) = \frac{1}{p}\sum_{i=1}^{p} m_i$.

Since $\sum_{i=1}^{q} \mu_i = \frac{1}{p}\sum_{i=1}^{p} m_i - 1$, we can conclude that

(2)
$$\frac{1}{q}\sum_{i=1}^{q} \mu_i = \frac{1}{p}$$

and

$$\sum_{i=1}^{q}(\mu_i)^2 = \frac{1}{p^2}\sum_{i=1}^{p}(m_i-1) + \frac{1}{p^2}\sum_{i\neq j}\sum \varphi_{ij}^2 \tag{3}$$

where φ_{ij}^2 is the observed Pearson's φ^2 crossing of X_i with X_j, and

$$\varphi^2 = \frac{1}{n}\sum_i\sum_j \frac{\left(n_{ij} - \frac{n_{i\cdot}\,n_{\cdot j}}{n}\right)^2}{\frac{n_{i\cdot}\,n_{\cdot j}}{n}} = \frac{\chi^2}{n} \ ,$$

($n_{i\cdot}$ and $n_{\cdot j}$ are margin effectives).

Although MCA is an extension of CA, the results of § 3 are not valid and one cannot use Malinvaud's test: elements of X being 0 or 1 and not frequencies, Q_k and Q'_k do not follow a chi-square distribution.

However it is possible to get information about the dispersion of the q eigenvalues in particular cases [5].

4.1. Distribution of eigenvalues in MCA under independence hypothesis

Under the hypothesis of pairwise independence of the variables X_i, all $\varphi_{ij}^2 = 0$ and equation (3), becomes

$$\sum_{i=1}^{q}(\mu_i)^2 = \frac{1}{p^2}\sum_{i=1}^{p}(m_i-1) \ .$$

Using (2) we get

$$\sum_{i=1}^{q}(\mu_i)^2 = \frac{1}{p^2}\, q \ ,$$

and finally:

$$\sum_{i=1}^{q}(\mu_i)^2 = \frac{1}{p^2} = \left[\frac{1}{q}\sum_i(\mu_i)\right]^2 .$$

Since the mean of the squared μ_i equals their squared means only if all the terms are equal, we can conclude that all the eigenvalues have the same value, so that:

$$\mu_i = \frac{1}{p} \ , \qquad \forall\, i \ .$$

We conclude that the theoretical MCA (i.e. for the population), under the hypothesis of pairwise independence of the variables X_i leads to one q-multiple non-trivial non-zero eigenvalue $\lambda = \frac{1}{p}$. And the eigenvalue diagram has the particular shape shown in *Figure 1*:

λ_I	Eigenvalues diagram
λ_1	****************************
λ_2	****************************
λ_3	****************************
λ_4	****************************
λ_5	****************************
$\vdots$	****************************
λ_q	****************************

Figure 1: Theoretical eigenvalues diagram in the independence case.

This result is not true when we have a finite sample, since sampling fluctuations make the observed $\varphi^2_{ij} \neq 0$. Although the trace of $\frac{1}{p}(D^{-1}B)$ and $\overline{\mu}$ the mean of the observed non-trivial eigenvalues, are constants, we observe q different non-trivial eigenvalues $\mu_i \neq \frac{1}{p}$, and the eigenvalue diagram takes the shape shown in *Figure 2*:

λ_I	Eigenvalues diagram
λ_1	****************************
λ_2	***************************
λ_3	**************************
λ_4	*************************
λ_5	************************
$\vdots$	***********************
λ_q	**********************

Figure 2: Observed eigenvalues diagram in the independence case.

4.1.1. Dispersion of eigenvalues

Let S^2_μ be the measure of μ_i around $\frac{1}{p}$ given by:

$$S^2_\mu \;=\; \frac{1}{q}\sum_{i=1}^{q}\left(\mu_i - \frac{1}{p}\right)^2 = \frac{1}{q}\sum_{i=1}^{q}(\mu_i)^2 - \frac{1}{p^2}\;,$$

which implies

$$\sum_{i=1}^{q}(\mu_i)^2 = q\left(S^2_\mu + \frac{1}{p^2}\right)\;.$$

Using equations (1)&(3), we have:

$$\sum_{i=1}^{q}(\mu_i)^2 \;=\; \frac{q}{p^2} + \frac{1}{p^2}\sum_{i\neq j}\sum \varphi^2_{ij} \;=\; \frac{q}{p^2} + \frac{1}{n\,p^2}\sum_{i\neq j}\sum \chi^2_{ij}\;.$$

Under the hypothesis of pairwise independence of the variables, the χ^2_{ij} are realizations of $\chi^2_{(m_i-1)(m_j-1)}$ variables, so their expected values are $(m_i-1)\,(m_j-1)$.

We can then easily compute $E(\sum_{i=1}^{q}(\mu_i)^2)$, and get:

$$E\left(\sum_{i=1}^{q}(\mu_i)^2\right) = \frac{q}{p^2} + \frac{1}{p^2}\,\frac{1}{n}\,\sum_{i\neq j}\sum(m_i-1)\,(m_j-1)\ .$$

Finally:

$$E(S^2_\mu) = \frac{1}{q}\,E\left(\sum_{i=1}^{q}(\mu_i)^2\right) - \frac{1}{p^2}$$

and we obtain:

$$E(S^2_\mu) = \frac{1}{p^2}\,\frac{1}{n}\,\frac{1}{q}\,\sum_{i\neq j}\sum(m_i-1)\,(m_j-1)\ .$$

Now, since $E(S^2_\mu)=\sigma^2$, we may assume that $\frac{1}{p}\pm 2\,\sigma$ contains roughly 95% of the eigenvalues. Moreover, since the kurtosis of the set of eigenvalues is lower than for a normal distribution, this proportion is actually probably larger then 95%.

4.1.2. Estimation of the Burt table

Let X be the disjunctive table associated to p categorical variables X_i, with m_i modalities respectively, observed on a sample of n individuals, where $X_i = (X_{i1}, X_{i2}, ..., X_{im_i})$, X is a matrix made (of p-block) of p blocks X_i

$$X = (X_1 \mid X_2 \mid ... \mid X_i \mid ... \mid X_p)\ .$$

Let $(X^j_{i1}, X^j_{i2}, ..., X^j_{ip})$ be the observed value of X_i on the j^{th} individual.

We can write

$$X = \begin{bmatrix} X^1_{11} & \cdots & X^1_{1m_1} & X^1_{21} & \cdots & X^1_{2m_2} & \cdots & X^1_{p1} & \cdots & X^1_{pm_p} \\ X^2_{11} & \cdots & X^2_{1m_1} & X^2_{21} & \cdots & X^2_{2m_2} & \cdots & X^2_{p1} & \cdots & X^2_{pm_p} \\ & \vdots & & & \vdots & & \vdots & & \vdots & \\ X^n_{11} & \cdots & X^n_{1m_1} & X^n_{21} & \cdots & X^n_{2m_2} & \cdots & X^n_{p1} & \cdots & X^n_{pm_p} \end{bmatrix}.$$

The Burt table of X is then

$$B = \begin{bmatrix} X'_1X_1 & X'_1X_2 & \cdots & X'_1X_p \\ X'_2X_1 & X'_2X_2 & \cdots & X'_2X_p \\ \vdots & \vdots & \ddots & \vdots \\ X'_pX_1 & X'_pX_2 & \cdots & X'_pX_p \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1p} \\ B_{21} & B_{22} & \cdots & B_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ B_{p1} & B_{p2} & \cdots & B_{pp} \end{bmatrix},$$

where

$$
\begin{aligned}
B_i &= B_{ii} \\
&= X_i' X_i = \begin{bmatrix}
\sum_{j=1}^{n} (X_{1i}^j)^2 & \sum_{j=1}^{n} (X_{1i}^j)\,(X_{2i}^j) & \cdots & \sum_{j=1}^{n} (X_{1i}^j)\,(X_{m_i i}^j) \\
\sum_{j=1}^{n} (X_{2i}^j)\,(X_{1i}^j) & \sum_{j=1}^{n} (X_{2i}^j)^2 & \cdots & \sum_{j=1}^{n} (X_{2i}^j)\,(X_{m_i i}^j) \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{j=1}^{n} (X_{m_i i}^j)\,(X_{1i}^j) & \sum_{j=1}^{n} (X_{m_i i}^j)\,(X_{2i}^j) & \cdots & \sum_{j=1}^{n} (X_{m_i i}^j)^2
\end{bmatrix}
\end{aligned}
$$

and

$$X_{ki}^j = \begin{cases} 0 \\ 1 \end{cases}$$

with $\sum_{k=1}^{m_i} X_{ki}^j = 1$. Since there is only one k in $\{1, ..., m_i\}$ such as $X_{ji}^k = 1$, all other being zero, we obtain:

$$\sum_{k=1}^{n} (X_{ki}^j)^2 = \sum_{k=1}^{n} X_{ki}^j \qquad \text{in } \{1, ..., n\}, \quad \forall\, k \in \{1, ..., m_i\}$$

and

$$\sum_{k=1}^{n} (X_{ki}^j)\,(X_{k'i}{}^j) = 0 \qquad \forall\, k, \quad k \in \{1, ..., m_i\} \; .$$

And so can conclude that $\forall\, i = 1, ..., p$ the diagonal sub-matrices of the Burt table are themselves diagonal matrices:

$$X_i' X_i = B_i = \begin{bmatrix}
\sum_{j=1}^{n} (X_{1i}^j)^2 & & & & 0 \\
& \ddots & & & \\
& & \sum_{j=1}^{n} (X_{ki}^j)^2 & & \\
& & & \ddots & \\
0 & & & & \sum_{j=1}^{n} (X_{m_i i}^j)^2
\end{bmatrix} .$$

Furthermore, we know that

$$\sum_{k=1}^{m_i} \left(\sum_{j=1}^{n} X_{ki}^j \right) = \sum_{k=1}^{m_i} (n_{ki}) = n \; ,$$

where

$$n_{ki} = \sum_{j=1}^{n} X_{ki}^j = n_i^k$$

is the number of individuals that have the k^{th} modality of the i^{th} variable (for $1 \le i \le p$ and $1 \le k \le m_i$).

So the diagonal sub-matrices of the Burt table are:

$$B_i = B_{ii} = \begin{bmatrix} n_i^1 & & & & 0 \\ & \ddots & & & \\ & & n_i^k & & \\ & & & \ddots & \\ 0 & & & & n_i^{m_i} \end{bmatrix} \qquad \text{where } \sum_{k=1}^{m_i} \frac{n_{ki}}{n} = 1 \quad \forall\, i = 1, ..., p\ .$$

Consider now two independent variables X_α and X_β amongst the p variables having respectively m_α and m_β modalities.

Let B_α be the (m_α, m_α) square matrix $B_\alpha = X'_\alpha X_\alpha$, and $B_{\alpha\beta}$ the (m_α, m_β) rectangular matrix $B_{\alpha\beta} = X'_\alpha X_\beta$.

We have

$$(B_\alpha)_{ii} = \sum_{k=1}^{n} X_{i\alpha}^k = X_{\cdot i}^{\alpha} \qquad \text{and} \qquad (B_\alpha)_{ij} = 0 \quad \text{if } i \neq j\ ,$$

and where $(B_{\alpha\beta})_{ij} = X_{i\alpha}^k\, X_{i\beta}^k \leq n$.

Under the hypothesis that X_α and X_β are independent

$$(B_{\alpha\beta})_{ij} = \frac{(B_\alpha)_{ij}\,(B_\beta)_{ij}}{n} = \frac{X_{\cdot i}^{\alpha}\, X_{\cdot i}^{\beta}}{n}\ .$$

Since $X_{\cdot i}^{\alpha} = n_i^{\alpha}$ and $X_{\cdot i}^{\beta} = n_i^{\beta}$, we can write

$$\left[(B_{\alpha\beta})_{ij} = \sum_{k=1}^{n} X_{ki}^{\alpha}\, X_{kj}^{\beta} = \frac{X_{\cdot i}^{\alpha}\, X_{\cdot i}^{\beta}}{n} = \frac{n_i^{\alpha}\, n_j^{\beta}}{n} \right]$$

and, more generally, we can conclude that

$$X'_i X_j = B_{ij} = \begin{bmatrix} \frac{n_1^i n_1^j}{n} & \frac{n_1^i n_2^j}{n} & \cdots & \frac{n_1^i n_{m_j}^j}{n} \\ \frac{n_2^i n_1^j}{n} & \frac{n_2^i n_2^j}{n} & \cdots & \frac{n_2^i n_{m_j}^j}{n} \\ \vdots & \vdots & & \vdots \\ \frac{n_{m_i}^i n_1^j}{n} & \frac{n_{m_i}^i n_2^j}{n} & \cdots & \frac{n_{m_i}^i n_{m_j}^j}{n} \end{bmatrix}$$

if the p variables are mutually independent.

Now consider a sample of p multinomial random variables X_i. Let $p_i^k = p_{ik}$ be the probability that an individual be in the k^{th} category of the i^{th} variable, and p_{ij}^k be the probably that the j^{th} individual be in the k^{th} category of the i^{th} variable.

The observed Burt table is:

$$B = X'X = \begin{bmatrix} X_1'X_1 & X_1'X_2 & \cdots & X_1'X_p \\ X_2'X_1 & X_2'X_2 & \cdots & X_2'X_p \\ \vdots & \vdots & \vdots & \vdots \\ X_p'X_1 & X_p'X_2 & \cdots & X_p'X_p \end{bmatrix},$$

with

$$X_i'X_i = N_i = \begin{bmatrix} \sum_{j=1}^{n}(X_{ij}^1)^2 & & & & 0 \\ & \ddots & & & \\ & & \sum_{j=1}^{n}(X_{ki}^j)^2 & & \\ & & & \ddots & \\ 0 & & & & \sum_{j=1}^{n}(X_{m_i i}^j)^2 \end{bmatrix} = \text{diag}\{n_i^1, ..., n_i^{m_i}\}\ .$$

But $n_i^k = \sum_{j=1}^{n}(X_{ki}^i)^2 = np_i^k$ and $\sum_{k=1}^{m_i} p_i^k = 1$, so that $\sum_{k=1}^{m_i} n_i^k = n\sum_{k=1}^{m_i} p_i^k = n,\ \forall\, i = 1, ..., p$

$$\text{and}\quad X_i'X_j = \begin{bmatrix} n\,p_i^1 & & & & 0 \\ & \ddots & & & \\ & & n\,p_i^k & & \\ & & & \ddots & \\ 0 & & & & n\,p_i^{m_i} \end{bmatrix}.$$

Since X_i and X_j are independent variables, $X_i'X_j = N_{ij}$ and $(N_{ij})_{kk'} = (X_i'X_j)_{kk'} = n\,p_i^k p_j^{k'}$, which implies

$$X_i'X_j = N_{ij} = \begin{bmatrix} n\,p_1^i p_1^j & n\,p_1^i p_2^j & \cdots & n_1^i n_{m_j}^j \\ n\,p_2^i p_1^j & n\,p_2^i p_2^j & \cdots & n\,p_2^i p_{m_j}^j \\ \vdots & \vdots & & \vdots \\ n\,p_{m_i}^i p_1^j & n\,p_{m_i}^i p_2^j & \cdots & n\,p_{m_i}^i p_{m_j}^j \end{bmatrix}.$$

The maximum-likelihood estimator of p_i^k is $\hat{p}_i^k = \frac{n_i^k}{n}$, so

$$\hat{N}_i = \begin{bmatrix} n_i^1 & & & & 0 \\ & \ddots & & & \\ & & n_i^k & & \\ & & & \ddots & \\ 0 & & & & n_i^{m_i} \end{bmatrix} = B_{ii}$$

and

$$\hat{N}_{ij} = \begin{bmatrix} \frac{n_1^i n_1^j}{n} & \frac{n_1^i n_2^j}{n} & \cdots & \frac{n_1^i n_{m_j}^j}{n} \\ \frac{n_2^i n_1^j}{n} & \frac{n_2^i n_2^j}{n} & \cdots & \frac{n_2^i n_{m_j}^j}{n} \\ \vdots & \vdots & & \vdots \\ \frac{n_{m_i}^i n_1^j}{n} & \frac{n_{m_i}^i n_2^j}{n} & \cdots & \frac{n_{m_i}^i n_{m_j}^j}{n} \end{bmatrix} = B_{ij} \ .$$

We can conclude that the the maximum-likelihood estimator $\hat{B}$ of the theoretical Burt table is $\tilde{B}$ the observed one. Using the invariance functional propriety we can affirm that the maximum-likelihood estimators of the eigenvalues of $D^{-1}B$ are the eigenvalues of $D^{-1}\tilde{B}$, so that each μ_i is the maximum-likelihood estimator of $\lambda_i = \lambda$.

Maximum-likelihood estimators are asymptotically normal, and so, asymptotically, each μ_i is normally distributed. But due to the fact that eigenvalues are ordered, the eigenvalues are not identically and independently distributed. However:

$$E(\mu_1) > \frac{1}{p}, \quad E(\mu_q) < \frac{1}{p} \qquad \text{but} \;\; E(\mu_1) \underset{n\to\infty}{\longrightarrow} \frac{1}{p} \;\; \text{and} \;\; E(\mu_q) \underset{n\to\infty}{\longrightarrow} \frac{1}{p} \ .$$

Furthermore the eigenvalue variances are not the same. And from simulations of large samples of n observations ($n = 100$, ..., $n = 10\,000$), we notice that the convergence of the eigenvalue distribution to a normal one is slow, especially for the extremes (μ_1 and μ_q), even for very large samples [4].

4.2. Distribution of eigenvalues in MCA under non-independence hypotheses

4.2.1. Distribution of the theoretical eigenvalues

Let μ be an eigenvalue of $D^{-1}X'X$. Since μ can be also obtained by diagonalization of $\frac{1}{p}XD^{-1}X'$, μ is a solution of $\frac{1}{p}XD^{-1}X'z = z$, where z is an eigenvector associated to μ.

So

$$\frac{1}{p}\left(\sum_{i=1}^{p} X_i\big(X_i'X_i\big)^{-1}X_i'\right)z \,=\, \mu\, z \quad\Longleftrightarrow\quad \frac{1}{p}\sum_{i=1}^{p} P_i\, z \,=\, \mu\, z\ ,$$

where $P_i = \sum_{i=1}^{p} X_i(X_i'X_i)^{-1}X_i'$ is the orthogonal projector on the space spanned by linear combinations of the indicators of variables categories X_i.

Let A_i the centered projector associated to P_i:

$$A_i = P_i - \frac{1_{m_i m_i}}{n} \qquad \text{where } 1_{m_i m_i} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} .$$

And so we get

$$\frac{1}{p}\sum_{i=1}^{p} A_i\, z \,=\, \mu\, z\ . \tag{4}$$

4.2.1.1. The Case of two-way interactions

Let us assume that among the p studied variables, there is a two-way interaction between X_j and X_k, and that the $(p-2)$ reminding variables are mutually independent. Multiplying equation (4) by A_j we get:

$$\frac{1}{p}\Big(\underbrace{A_jA_1}_{0} + \underbrace{A_jA_2}_{0} + \cdots + \underbrace{A_jA_j}_{Aj} + \cdots + A_jA_k + \cdots + \underbrace{A_jA_p}_{0}\Big) z \,=\, \mu\, A_j\, z\ ,$$

since all variables are pairwise independent except X_j, X_k, and the A_i are orthogonal projectors. Thus:

$$A_jA_k\, z \,=\, (p\,\mu - 1)\, A_j\, z\ . \tag{5}$$

Similarly, multiplying (4) by A_k, we get:

$$A_kA_j\, z \,=\, (p\,\mu - 1)\, A_k\, z\ . \tag{6}$$

Now let us multiply (5) by A_k to get:

$$A_kA_jA_k\, z \,=\, (p\,\mu - 1)\, A_kA_j\, z\ .$$

Using (6) we obtain

$$A_kA_j\,\underbrace{A_k\, z}_{z_1} \,=\, (p\,\mu - 1)^2\,\underbrace{A_k\, z}_{z_1}\ .$$

With the notation $\lambda = (p\,\mu - 1)^2$, we finally write:

$$A_k A_j\, z_1 \;=\; \lambda\, z_1 \;. \tag{7}$$

Equation (7) implies that λ is an eigenvalue of the product of the centered projector $A_k A_j$ associated to the eigenvector z_1.

In general: $\forall\, j, k = 1, ..., p$, if there is an interaction between X_j and X_k, the orthogonal projector $A_j A_k$ admits a non zero eigenvalue $\lambda = (p\,\mu - 1)^2$.
If $\lambda \neq 0 \Leftrightarrow \mu \neq \frac{1}{p}$, the trace of Burt table being constant, there is, at least, another eigenvalue not equal to $\frac{1}{p}$.

Let n_0 be the number of eigenvalue non equal to $\frac{1}{p}$, so that $\sum_{i=1}^{n_0} \lambda_i = \frac{n_0}{p}$.

Theoretically, (except for the particular case, where $\lambda = 1$, for which we have $\mu = \frac{2}{p}$ and $\mu' = 0$), the number of non-trivial-eigenvalues greater than $\frac{1}{p}$ is equal to the number of non-trivial eigenvalues smaller than $\frac{1}{p}$.

The eigenvalue diagram shape is shown on *Figure 3*:

λ_I	Eigenvalues diagram
λ_1	***************************
λ_2	**************************
λ_3	***********************
λ_4	***********************
λ_5	***********************
$\vdots$	*****************
λ_q	****************

Figure 3: Theoretical eigenvalues diagram in two-way interaction case.

The number n_0 depends on the number of categories of X_j and X_k, on the number of variables and on the number of dependent variables.

Let n_1 be the multiplicity of $\frac{1}{p}$, we will show that $n_1 = q - 2\min((m_j - 1);\,(m_k - 1))$, when $p > 2$, and when there is only one two-way interaction between the variables.

This result can be shown as follows:

Let us consider equation (4), and suppose, without loss of generality, that X_1 and X_2 are dependant. So, upon multiplication by A_3: $\frac{1}{p}\sum_{i=1}^{p} A_i z = \mu z$ becomes $\frac{1}{p}(A_3 A_1 + A_3 A_2 + A_3 A_3 + \cdots + A_3 A_P)\, z = \mu\, A_3\, z$, and we get $\mu = \frac{1}{p}$.

Now multiply equation (4) by A_2 and A_1 in turn to get:

$$\begin{cases} \Big(A_1A_1 + A_1A_2 + A_1A_3 + \cdots + A_1A_P\Big)\, z \;=\; p\,\mu A_1\, z \\ \Big(A_2A_1 + A_2A_2 + A_2A_3 + \cdots + A_2A_P\Big)\, z \;=\; p\,\mu A_2\, z \end{cases} \Longleftrightarrow$$

$$\Longleftrightarrow \quad \begin{cases} (A_1 + A_1A_2)\, z \;=\; p\,\mu\, A_1\, z \\ (A_2A_1 + A_2)\, z \;=\; p\,\mu\, A_2\, z \end{cases}$$

$$\Longleftrightarrow \quad \begin{cases} A_1A_2\, b \;=\; \lambda\, z \\ A_2A_1\, b \;=\; \lambda\, z \end{cases}$$

where $\lambda = (p\,\mu - 1)^2$, $a = A_1\, z$ and $b = A_2\, z$.

We recognize here the CA equations, so that the CA of Burt tables, when only two variables are dependent is equivalent to the CA of the contingency tables crossing the two dependent variables. It is well known that the number of eigenvalue in CA equals $q - 2\min((m_j - 1); (m_k - 1))$, and for all non trivial λ_i, there corresponds the values μ_i and μ_i' such that:

$$\mu_i = \frac{1 + \sqrt{\lambda_i}}{p} \qquad \text{and} \qquad \mu_i' = \frac{1 - \sqrt{\lambda_i}}{p}\ .$$

Finally, the CA of the Burt table may have $2\min((m_j–1);(m_k–1))$ eigenvalues non trivial and not equal to $\frac{1}{p}$, associated to the CA of the contingency table. So the number of supplementary eigenvalues equals $q - 2\min((m_j - 1); (m_k - 1))$.

There is, in addition, one n_1 multiple eigenvalue, where n_1 is at least $q - 2\min((m_j - 1); (m_k - 1))$.

4.2.1.2. The case of higher order interactions

Since the Burt table is constructed with pairwise cross products of variables, its observation cannot give us information about multiway interactions.

However the observation of the bi-dimensional theoretical Burt sub-tables, for all pairwise variable combinations, can provide us with all the two-way interactions.

The theoretical Burt table can reveal the existence of higher order interactions in the following case:

If $B_{ij} \neq B_{ii}\, 1_{m_jm_j} B_{jj}$ and $B_{ik} \neq B_{ii}\, 1_{m_km_k} B_{kk}$: there may be a triple interaction between X_i, X_j and X_k.

In general, a Burt table doesn't give either the order of the interactions, or supplementary information on the eigenvalue behavior.

4.2.2. Distribution of observed eigenvalues

Exceptionally, with a small number of interactions, we observe the particular shape of the eigenvalue diagram exhibited in *Figure 4*, where we can distinguish eigenvalues near $\frac{1}{p}$ (theoretically equal to $\frac{1}{p}$), and so we are able to recognize the existence of the independent variables in the analysis.

λ_I	Eigenvalues diagram
λ_1	****************************
λ_2	***************************
λ_3	***********************
λ_4	**********************
λ_5	*********************
$\vdots$	**************
$\vdots$	*************
λ_q	************

Figure 4: Observed eigenvalues diagram in a two-way interaction case.

When the number of interaction grows, we cannot distinguish eigenvalues theoretically equal to $\frac{1}{p}$ from the eigenvalues non equal to $\frac{1}{p}$.

To detect the existence or interactions, we can first check if the observed variables are mutually independent. In that case, the eigenvalues distribution diagram should have a particular shape (see § 4.1.), with more than 95% of the eigenvalues within the confidence interval $\frac{1}{p} \pm 2\,\sigma$ (see § 4.1.1).

If there is one or more eigenvalues out of the confidence interval, we can then assume the existence of one or more two-way interaction between variables.

5. AN EMPIRICAL PROCEDURE FOR FITTING LOG-LINEAR MODELS BASED ON THE MCA EIGENVALUE DIAGRAM

We propose an empirical procedure for progressively fitting a log-linear model where the fitting test at each step is based on the MCA eigenvalues diagram.

Let X_i, X_j and X_k, three categorical variables, with respectively m_i, m_j and m_k modalities, and let a cross variable with $(m_i \times m_j)$ modalities. We suppose that X_{ij} and X_k, have the same behavior if $m_k = m_i \times m_j$.

Under the hypothesis that two dependant variables X_i and X_j have the same behaviour as the variable X_k with the same characteristics of the cross variable X_{ij}, we propose here an empirical procedure for fitting progressively, with p steps, the log-linear model where the fitting criterion at each step is based on the MCA eigenvalue diagram. Distribution of observed eigenvalues

5.1. Description of the procedure steps

The first step of the procedure consist to test the pairwise independence hypothesis of the variables. To detect existence of interactions, we must first check if all variables are mutually independent. For that matter, we calculate the eigenvalues of MCA on all the p variables, and construct the related confidence interval: the eigenvalue distribution diagram should have a particular shape (cf. § 4.1.). If all the eigenvalues belong to the confidence interval $\frac{1}{p} \pm 2\,\sigma$ (cf. § 4.1.1), we can conclude that the p variables are mutually independent. The log-linear model associated to the variables is a simple additive one:

$$\log[f_p(X)] \;=\; u_0(x) + \sum_{i=1}^{p} u_i(x) \ ,$$

and the procedure is stopped.

If one or more eigenvalue are not in the confidence interval, we conclude that there is at least one double interaction between variables, and we go to the second step of the procedure.

In the second step, we look at all two-way interaction u-terms. We isolate one variable amongst the p variables that we note X_p, without loss of generality, and so we obtain a set of $(p-1)$ variables X_i, and apply the first step to test pairwise independence of the $(p-1)$ variables.

If the $(p-1)$ variables are independent, we can conclude that the doubles interactions are amongst X_p and at least one of the X_i, so we construct correspondent cross variables X_{ip} by using the first step to test independence between variables (X_i, X_p) where $i = 1, ..., p-1$. The log-linear model associated to the variables is:

$$\log[f_p(X)] \;=\; u_0(x) + \sum_{i=1}^{p} u_i(x) + \sum_{i=1}^{p-1} u_{ip}(x)\, \delta_{ip} \ ,$$

and the procedure stopped, (with $\delta_{ip} = 1$ if the interaction between X_p and X_i exists, otherwise it is set to zero.)

If the $(p-1)$ variables are not independent, we can conclude that there is double interaction between X_i and X_j where $i, j = 1, ..., p$–1, and perhaps between X_i and X_p.

We can construct correspondent cross variables X_{ip} and X_{ij} by using the first step to test independence of variables (X_i, X_p) and variables (X_i, X_j) where $i, j = 1, ..., p-1$. The log-linear model associated to the variables is:

$$\log[f_p(X)] = u_0(x) + \sum_{i=1}^{p} u_i(x) + \sum_{i=1}^{p-1} u_{ip}(x)\, \delta_{ip} + \text{terms due to the interaction between three or more variables}$$

and we go to the third step of the procedure

In the third step, we look at three-way interaction u-terms, by testing the dependence of variables X_i and cross variables X_{jk}, where $i, j, k = 1, ..., p$ and i, j, k are different, and construct cross variables X_{ijk}. The independence test is based on the eigenvalue pattern of the related MCA as described in the first step.

Continuing this way, in the k^{th} step, we look at k-way interaction u-terms, ... and in the least step we look at the p-way interaction u-term.

This algorithm is summarized in *Figure 5.*

5.2. An example for a graphical model

For this example we use a data set given by Haberman [24] that was used in Falguerolles *et al.* [14] to fit a graphical model. The data reports attitudes toward non therapeutic abortions among white subjects crossed with three categorical variables describing the subjects.

The data set is a contingency table observed for 3181 individuals, crossing four three modality variables X_1, X_2, X_3 and X_4, defined in *Table 1.*

The first step of the procedure consists of testing the pairwise independence hypothesis of the variables. We first transform the contingency table in a complete disjunctive table, then calculate the parameters (defined in § 2.1 and § 4.1.1) needed for the test (*Table* 2).

MCA on the four variables gives the eigenvalues diagram of *Figure 6.*

The shape of eigenvalues diagram refers clearly to the existence of dependent variables.

Eigenvalues λ_1, λ_7 and λ_8 are not in the interval I_c, so there is at least two dependent variables: there is one or more two-way interactions between variables.

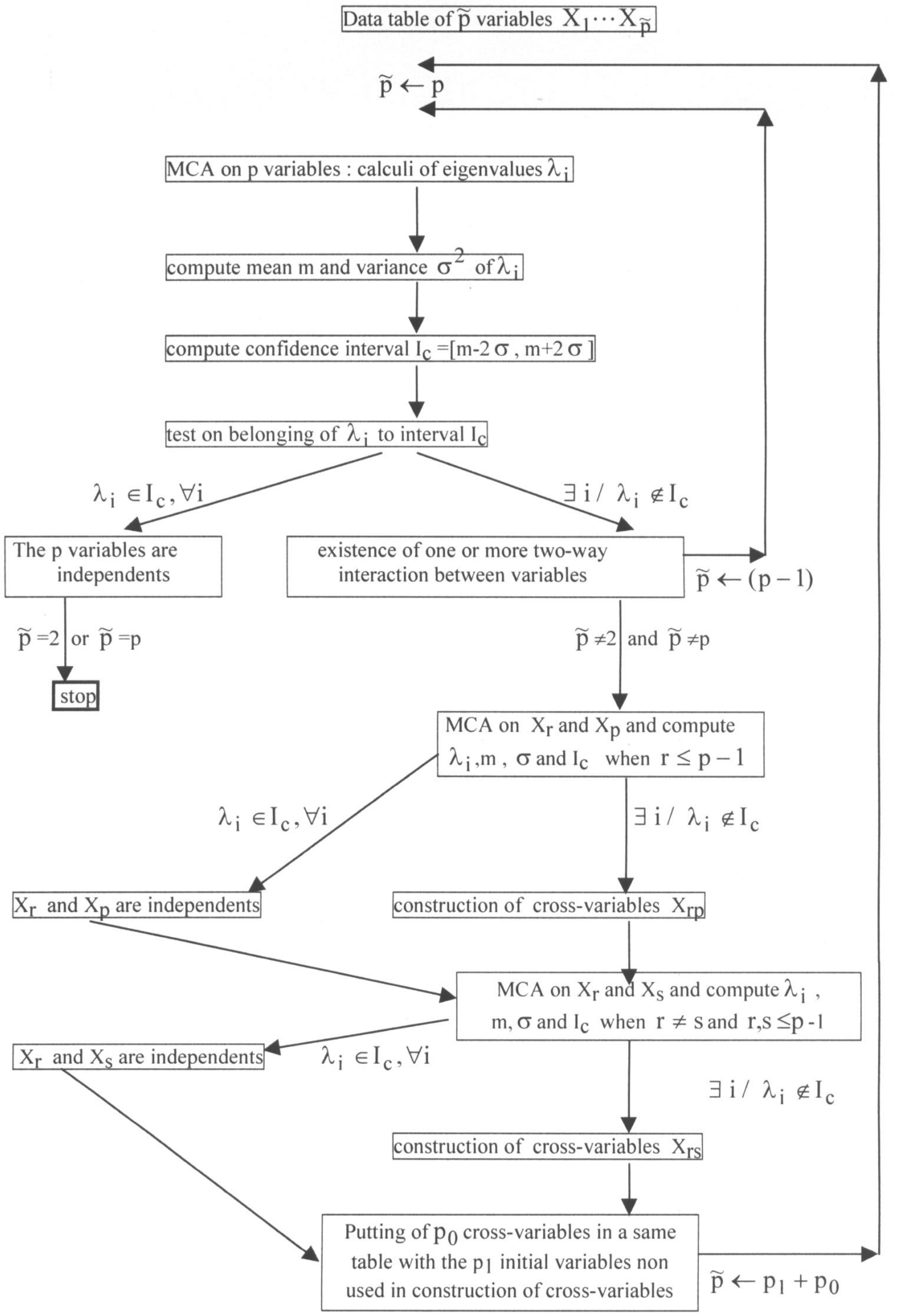

Figure 5: Block diagram for the Empirical procedure.

Table 1: Attitudes toward non therapeutic abortions among white.

Year X_1	Religion: X_2	Education in years: X_3	Attitude: X_4 positive	mixed	negative
1972	northern Protestant	≤ 8	09	16	41
		9–12	85	52	105
		≥ 13	77	30	38
	southern Protestant	≤ 8	08	08	46
		9–12	35	29	54
		≥ 13	37	15	22
	Catholic	≤ 8	11	14	38
		9–12	47	35	115
		≥ 13	25	12	42
1973	northern Protestant	≤ 8	17	17	42
		9–12	102	38	84
		≥ 13	88	15	31
	southern Protestant	≤ 8	14	11	34
		9–12	61	30	59
		≥ 13	49	11	19
	Catholic	≤ 8	06	16	26
		9–12	60	29	108
		≥ 13	31	18	50
1974	northern Protestant	≤ 8	23	13	32
		9–12	106	50	88
		≥ 13	79	21	31
	southern Protestant	≤ 8	05	15	37
		9–12	38	39	54
		≥ 13	52	12	32
	Catholic	≤ 8	08	10	24
		9–12	65	39	89
		≥ 13	37	18	43

Table 2: Parameters needed for the test
(first step of the example for a graphical model).

n	p	m_1	m_2	m_3	m_4	q	m	σ	I_c
3181	4	3	3	3	3	8	0.25	0.0109	[0.2283, 0.2717]

$\lambda_1 = 0.3221$	**************************
$\lambda_2 = 0.2704$	*********************
$\lambda_3 = 0.2599$	********************
$\lambda_4 = 0.2531$	*******************
$\lambda_5 = 0.2451$	******************
$\lambda_6 = 0.2393$	*****************
$\lambda_7 = 0.2277$	****************
$\lambda_8 = 0.1823$	***********

Figure 6: Eigenvalues diagram
(first step of the example for a graphical model).

The second step consists of the detection of two-way interactions. In a first time, we use our first step with only three variables X_1, X_2 and X_3.

With the values of n and m_i (for $i = 1, ..., 3$) still the same, the other parameters become (*Table 3*):

Table 3: Parameters for the test
(second step of the example for a graphical model).

q	m	σ	I_c
6	0.33333	0.0118	[0.3097, 0.3569]

We get the following eigenvalue diagram (*Figure 7*):

$\lambda_1 = 0.3606$	**************************
$\lambda_2 = 0.3448$	*************************
$\lambda_3 = 0.3385$	************************
$\lambda_4 = 0.3305$	**********************
$\lambda_5 = 0.3025$	*********************

Figure 7: Eigenvalues diagram
(second step of the example for a graphical model).

λ_1 and λ_5 are not in interval I_c, so there is one or more two-way interaction between X_1, X_2 and X_3, as also as interactions between X_4 and others.

In a second step we look at the interactions between X_4 and X_i ($i = 1, 2, 3$).

For $i = 1$ to $i = 3$ we look at the eigenvalues of the MCA of X_4 with X_i, and calculate their variances and intervals I_c.

Crossing X_1 with X_4 we get (*Table 4*):

Table 4: MCA on X_1 and X_4 (parameters and eigenvalues).

q	m	σ	I_c	λ_1	λ_2	λ_3	λ_4
4	0.5	0.0125	[0.4750, 0.5250]	0.5389	0.5156	0.4644	0.4611

Crossing X_2 with X_4 we get (*Table 5*):

Table 5: MCA on X_2 and X_4 (parameters and eigenvalues).

q	m	σ	I_c	λ_1	λ_2	λ_3	λ_4
4	0.5	0.0125	[0.4750, 0.5250]	0.5741	0.5076	0.4924	0.4259

Crossing X_3 with X_4 we get (*Table 6*):

Table 6: MCA on X_3 and X_4 (parameters and eigenvalues).

q	m	σ	I_c	λ_1	λ_2	λ_3	λ_4
4	0.5	0.0125	[0.4750, 0.5250]	0.6112	0.5041	0.4959	0.3979

In the three cases, λ_1 and λ_4 are not in the intervals I_c, so there is a two-way interaction between X_1 and X_4, X_2 and X_4 and between X_3 and X_4, so we can construct cross variables X_{4i} having 9 modalities ($i = 1, 2, 3$).

Now, we use the first step with only two variables X_1 and X_2, after we look for interactions between X_3 and X_i ($i = 1, 2$).

Crossing X_1 with X_2 we get (*Table 7*):

Table 7: MCA on X_1 and X_2 (parameters and eigenvalues).

q	m	σ	I_c	λ_1	λ_2	λ_3	λ_4
4	0.5	0.0125	[0.4750, 0.5250]	0.5153	0.5045	0.4955	0.4848

All the eigenvalues are in the confidence interval, so X_1 and X_2 are independent conditionally on the other, and there is no cross variable X_{12}. The corresponding u-term u_{12} equals to zero.

Let us now look, when $i = 1$ and $i = 2$, at the eigenvalues of the MCA of X_3 with X_i, with their variances and intervals I_c:

Crossing X_1 with X_3 we get (*Table 8*):

Table 8: MCA on X_1 and X_3 (parameters and eigenvalues).

q	m	σ	I_c	λ_1	λ_2	λ_3	λ_4
4	0.5	0.0125	[0.4750, 0.5250]	0.5134	0.5023	0.4978	0.4866

All the eigenvalues are in the confidence interval I_c, so X_1 and X_3 are independent conditionally on the other, and there is no cross variable X_{13}: the corresponding u-term u_{13} equals to zero.

Crossing now X_2 with X_3 we get (*Table 9*):

Table 9: MCA on X_2 and X_3 (parameters and eigenvalues).

q	m	σ	I_c	λ_1	λ_2	λ_3	λ_4
4	0.5	0.0125	[0.4750, 0.5250]	0.5401	0.5128	0.4872	0.4599

Here, λ_1 and λ_4 are not in the interval I_c, so there is a two-way interaction between X_2 and X_3, u_{23} is not set to zero, and we can add the cross variable X_{32} (as well as X_{23}) with 9 modalities to the model.

The third step consists of the detection of triple interactions between variables, that is to two-way interactions between the variables X_i and the cross variables X_{jk}.

We first put the cross variables (X_{41}, X_{42}, X_{43}, X_{32}) with the initial variables that were deemed non dependent in the second step of the procedure, i.e. X_1 and X_2, and then we use the first step of the procedure with the set of obtained variables.

So we get the following results (*Table 10* and *Figure 8*):

Table 10: MCA on X_1, X_2, X_{41}, X_{42}, X_{43} and X_{32} (parameters third step of the example for a graphical model).

q	m	σ	I_c
36	0.1667	0.0168	[0.1331, 0.2003]

$\lambda_1 = 0.5201$	**************************
$\lambda_2 = 0.5006$	*************************
$\lambda_3 = 0.3447$	******************
$\lambda_4 = 0.3347$	******************
$\lambda_5 = 0.3303$	******************
$\lambda_6 = 0.3193$	*****************
$\lambda_7 = 0.1810$	************
$\lambda_8 = 0.1796$	***********
$\lambda_9 = 0.1732$	***********
$\lambda_{10} = 0.1710$	***********
$\lambda_{11} = 0.1664$	***********
$\lambda_{12} = 0.1627$	***********
$\lambda_{13} = 0.1626$	***********
$\lambda_{14} = 0.1578$	***********
$\lambda_{15} = 0.1538$	**********
$\lambda_{16} = 0.1423$	*********

Figure 8: MCA on X_1, X_2, X_{41}, X_{42}, X_{43} and X_{32} (eigenvalues diagram, third step of the example for a graphical model).

The first six eigenvalues are not in I_c: there is one or more two-way interaction between the initial variables X_i, and the crossed ones X_{ik}, so there exists a triple interaction between simple variables.

We drop X_{32} and use the first step with the five other variables to get the following results (*Table 11* and *Figure 9*):

Table 11: MCA on X_1, X_2, X_{41}, X_{42} and X_{43} (parameters for the test).

q	m	σ	I_c
28	0.2	0.0162	[0.1671, 0.2324]

$\lambda_1 = 0.6105$	**************************
$\lambda_2 = 0.6006$	**************************
$\lambda_3 = 0.4143$	****************
$\lambda_4 = 0.4028$	****************
$\lambda_5 = 0.3982$	****************
$\lambda_6 = 0.3831$	***************
$\lambda_7 = 0.2262$	**********
$\lambda_8 = 0.2220$	**********
$\lambda_9 = 0.2162$	**********
$\lambda_{10} = 0.2083$	*********
$\lambda_{11} = 0.2054$	*********
$\lambda_{12} = 0.2017$	*********
$\lambda_{13} = 0.1952$	*********
$\lambda_{14} = 0.1986$	*********
$\lambda_{15} = 0.1952$	*********
$\lambda_{16} = 0.1928$	*********
$\lambda_{17} = 0.1878$	********
$\lambda_{18} = 0.1837$	********
$\lambda_{19} = 0.1815$	********
$\lambda_{20} = 0.1711$	********

Figure 9: MCA on X_1, X_2, X_{41}, X_{42} and X_{43} (eigenvalues diagram, third step of the example for a graphical model).

The first six eigenvalues are not in I_c, so there is at least one two-way interaction between the variables. We know that simple variables X_1, X_2 and the crossed variables X_{41}, X_{42}, X_{43} are dependent so we have to test dependence between X_1 and X_{32} only. Crossing X_1 and X_{32} we get the following results (*Table 12)*:

Table 12: MCA on X_1 and X_{32} (parameters and eigenvalues).

q	m	σ	I_c
10	0.5	0.0159	[0.4682, 0.5318]

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}
0.5287	0.5194	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.4806	0.4713

All the eigenvalues are in the confidence interval I_c, so X_1 and X_{32} are independent conditionally on the other, and there is no cross variable X_{132}. The corresponding u-term u_{123} equals zero.

Now we can drop the cross variable X_{43}. The remaining variables X_1, X_2, X_{41}, X_{42} are dependent by construction. We have only to test for dependence between X_1 and X_{43}.

Crossing X_1 with X_{43}, we get the same parameter as the crossing of X_1 and X_{32}, and the following eigenvalues (*Table 13*):

Table 13: MCA on X_1 and X_{43} (eigenvalues).

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}
0.5445	0.5232	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.4768	0.4555

We remark that λ_1 and λ_{10} are not in the interval I_c, so X_1 and X_{43} seem to be dependent. But we have to fit a graphical model, that is a particular case of hierarchical models (as defined in § 2.2.2.2, a log-linear models is hierarchical if, whenever one particular u-term is constrained to zero then all higher u-terms containing the same set of subscripts are also set to zero).

Here the u-term u_{13} is set to zero, so the u-term u_{134} is also set to zero.

Crossing X_2 with X_{43}, we get the same parameter as the crossing of X_1 and X_{32}, and the following eigenvalues (*Table 14*):

Table 14: MCA on X_2 and X_{43} (eigenvalues).

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}
0.5871	0.5466	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.4534	0.4143

Eigenvalues λ_1, λ_2, λ_9 and λ_{10} are not in the interval I_c, the u-terms u_{23} and u_{24} are not set to zero, and since X_2 and X_{43} are not dependent the u-term u_{234} is not set to zero.

Crossing X_1 with X_{42} (or equivalently X_2 with X_{41}) we get the same parameter as the crossing of X_1 and X_{32}, and the following eigenvalues:

Table 15: MCA on X_1 and X_{42} (eigenvalues).

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}
0.5434	0.5289	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.4711	0.4566

Eigenvalues λ_1 and λ_{10} are not in the interval I_c, the u-term u_{14} is equal to zero, X_1 and X_{42} are dependent, and the u-term u_{124} is set to zero.

Finally, variables X_1 and X_{41} are dependent by construction.

The procedure stops here because we can't have more than triple interactions in a hierarchical model when all the two-way interactions are not present. We obtain the following model (see *Figure 10* for the associated graph):

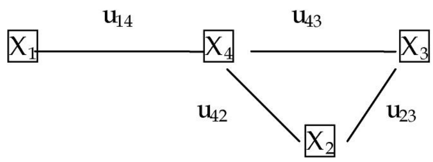

Figure 10: Lattice diagram (example for a graphical model).

$$\log[f_4(X)] = u_0 + u_1 x_1 + u_2\, x_2 + u_3\, x_3 + u_4\, x_4 + u_{32}\, x_2 x_3 + u_{41} x_4 x_1 + u_{42}\, x_4 x_2 + u_{43}\, x_4 x_3 + u_{432}\, x_4 x_3 x_2 \;.$$

5.3. An example for a saturated model

Here we use a data set given by Israëls [29] that was also used by Van der Heijden et al. [28] about 'shop-lifting' habits.

Table 16 is a contingency table crossing three variables: sex (2 modalities), age (9 modalities) and type of goods (13 modalities: Clothing (C), Clothing accessories (Ca), Provision-Tobacco (PT), Writing materials (Wm), Books (B), Records (R), Household goods (Hg), Sweets (S), Toys (T), Jewellery (J), Perfume (P), Hobbies tools(Ht), and Others(O)) observed over 33 101 individuals.

In the first step, we test the pairwise independence of variables X_1, X_2 and X_3. We first transform the contingency table in a complete disjunctive table, then compute the parameters (defined in § 2.2 & § 4.1.1) needed for the test to get (*Table 17*).

A MCA on the three variables gives the eigenvalue diagram of *Figure 11*.

The eigenvalue diagram shows clearly that the variables are not independent: only 8 eigenvalues ($\lambda_7, ..., \lambda_{15}$) are in the confidence interval.

Using the second step of the procedure, we get the two-way interactions.

Table 16: Multicontingency table for the shop-lifting data.

Sex:	Age:	Goods: X_3												
X_1	X_2	C	Ca	PT	Wm	B	R	Hg	S	T	J	P	Ht	O
Male	≤ 11	81	66	150	667	67	24	47	430	743	132	32	197	209
	12–14	138	204	340	1409	259	272	117	637	684	408	57	547	550
	15–17	304	193	229	527	258	368	98	246	116	298	61	402	454
	18–20	384	149	151	84	146	141	61	40	13	71	52	138	252
	21–29	942	297	313	92	251	167	193	30	16	130	111	280	624
	30–39	359	109	136	36	96	67	75	11	16	31	54	200	195
	40–49	178	53	121	36	48	29	50	5	6	14	41	152	88
	50–64	137	68	171	37	56	27	55	17	3	11	50	211	90
	≥ 65	45	28	145	17	41	7	29	28	8	10	28	111	34
Female	≤ 11	71	19	59	224	19	7	22	137	113	162	70	15	24
	12–14	241	98	111	463	60	32	29	240	98	138	178	29	58
	15–17	477	114	58	91	50	27	41	80	14	548	141	9	72
	18–20	436	108	76	18	32	12	32	12	10	303	70	14	67
	21–29	1180	207	132	30	61	21	65	16	12	74	104	30	157
	30–39	1009	165	121	27	43	9	74	14	31	100	81	36	107
	40–49	517	102	93	23	31	7	51	10	8	48	46	24	66
	50–64	488	127	214	27	57	13	79	23	17	22	69	35	64
	≥ 65	173	64	215	13	44	0	39	42	6	12	41	11	55

Table 17: Parameters needed for the test
(first step of the example for a satured model).

n	p	m_1	m_2	m_3	q	m	σ	I_c
33101	3	2	9	13	21	0.3333	0.0061	[0.3211, 0.3455]

$\lambda_1 = 0.5759$	***
$\lambda_2 = 0.4256$	***********************************
$\lambda_3 = 0.3966$	********************************
$\lambda_4 = 0.3899$	*******************************
$\lambda_5 = 0.3542$	****************************
$\lambda_6 = 0.3494$	****************************
$\lambda_7 = 0.3407$	***************************
$\lambda_8 = 0.3384$	**************************
$\lambda_9 = 0.3344$	**********************
$\lambda_{10} = 0.3333$	**********************
$\lambda_{11} = 0.3333$	**********************
$\lambda_{12} = 0.3333$	**********************
$\lambda_{13} = 0.3322$	*********************
$\lambda_{14} = 0.3271$	*********************
$\lambda_{15} = 0.3260$	*********************
$\lambda_{16} = 0.3177$	********************
$\lambda_{17} = 0.3103$	*******************
$\lambda_{18} = 0.2802$	******************
$\lambda_{19} = 0.2632$	****************
$\lambda_{20} = 0.1925$	************
$\lambda_{21} = 0.1423$	*******

Figure 11: MCA on X_1, X_2 and X_3
(eigenvalues diagram, third step of the example for a saturated model).

MCA of X_1 and X_3 gives the following results (*Table 18* and *Figure 12*):

Table 18: MCA on X_1 and X_3
(parameters).

n	p	q	m	σ	I_c
33101	2	13	0.5	0.00002	[0.5000, 0.5000]

$\lambda_1 = 0.7032$	**
$\lambda_2 = 0.5000$	*************************
$\lambda_3 = 0.5000$	*************************
$\lambda_4 = 0.5000$	*************************
$\lambda_5 = 0.5000$	*************************
$\lambda_6 = 0.5000$	*************************
$\lambda_7 = 0.5000$	*************************
$\lambda_8 = 0.5000$	*************************
$\lambda_9 = 0.5000$	**********************
$\lambda_{10} = 0.5000$	**********************
$\lambda_{11} = 0.5000$	**********************
$\lambda_{12} = 0.5000$	**********************
$\lambda_{13} = 0.2968$	**********

Figure 12: MCA on X_1 and X_3
(eigenvalues diagram, second step of the example for a saturated model).

The first and the last eigenvalues are not in the confidence interval so the u-term u_{13} is not set to zero.

We notice here the peculiar form of the eigenvalues diagram, due to the fact that multiple eigenvalue $\lambda = \frac{1}{2}$ that have a multiplicity $11 = m_3 - m_1$ is an artificial one (cf. § 4.2.1.1).

MCA of X_2 and X_3 gives the following results (*Table 19* and *Figure 13*):

Table 19: MCA on X_2 and X_3
(parameters).

n	p	q	m	σ	I_c
33101	2	20	0.5	0.0001	[0.4998, 0.5002]

The 8 first and the 8 last eigenvalues are not in the confidence interval so the u-term u_{23} is not set to zero.

$\lambda_1 = 0.7852$	**
$\lambda_2 = 0.6074$	*******************************
$\lambda_3 = 0.5903$	*****************************
$\lambda_4 = 0.5346$	**************************
$\lambda_5 = 0.5245$	*************************
$\lambda_6 = 0.5112$	************************
$\lambda_7 = 0.5109$	************************
$\lambda_8 = 0.5019$	***********************
$\lambda_9 = 0.5000$	********************
$\lambda_{10} = 0.5000$	********************
$\lambda_{11} = 0.5000$	********************
$\lambda_{12} = 0.5000$	********************
$\lambda_{13} = 0.4981$	*******************
$\lambda_{14} = 0.4891$	******************
$\lambda_{15} = 0.4888$	******************
$\lambda_{16} = 0.4755$	*****************
$\lambda_{17} = 0.4654$	****************
$\lambda_{18} = 0.4097$	************
$\lambda_{19} = 0.3926$	***********
$\lambda_{20} = 0.2148$	******

Figure 13: MCA on X_2 and X_3
(eigenvalues diagram, second step of the example for a saturated model).

MCA of X_1 and X_2 gives the following eigenvalue results (*Table 20, Figure 14*):

Table 20: MCA on X_1 and X_2
(parameters).

n	p	q	m	σ	I_c
33101	2	9	0.5	0.0037	[0.4926, 0.5074]

$\lambda_1 = 0.6241$	**
$\lambda_2 = 0.5000$	*************************
$\lambda_3 = 0.5000$	*************************
$\lambda_4 = 0.5000$	*************************
$\lambda_5 = 0.5000$	*************************
$\lambda_6 = 0.5000$	*************************
$\lambda_7 = 0.5000$	*************************
$\lambda_8 = 0.5000$	*************************
$\lambda_9 = 0.3759$	**********

Figure 14: MCA on X_1 and X_2
(eigenvalues diagram, second step of the example for a saturated model).

The first and the last eigenvalues are not in the confidence interval so the u-term u_{12} is not set to zero. At the end of the second step, we obtain all three two-way interactions. To know if the model is a saturated one we can built one of the crossed variables and test its dependence with the remaining simple variable.

MCA of X_{32} with X_1 gives the following eigenvalues:

$$\lambda_1 = 0.7285 \;, \qquad \lambda_2 = \lambda_3 = \cdots = \lambda_{116} = 0.5 \;,$$
$$\lambda_{117} = 0.2715 \quad \text{and} \quad I_c = [0.4615,\, 0.5384] \;.$$

The first and the last eigenvalues are not in the confidence interval so the u-term u_{123} is not set to zero.

At the end we get the following saturated model:

$$\log[f_3(X)] = u_0 + u_1 x_1 + u_2\, x_2 + u_3\, x_3 + u_{12}\, x_1 x_2 + u_{23}\, x_2 x_3 + u_{13}\, x_1 x_3 + u_{123}\, x_1 x_2 x_3 \;.$$

5.4. An example for a mutual independence model

Here we use a data set given by Andersen [2] as a contingency table crossing four variables observed over 299 individuals corresponding to a retrospective study of ovary cancer, defined in Table 21:

Table 21: Retrospective study of ovary cancer.

X_1 stage	X_2 operation	X_3 survival	X_4 X-ray No	X_4 X-ray Yes
Early	radical	no	10	17
	limited	yes	41	64
		no	1	3
		yes	13	9
Advanced	radical	no	38	64
	limited	yes	6	11
		no	3	13
		yes	1	5

In the first step of procedure, we test for the pairwise independence of variables X_1, X_2, X_3 and X_4. We first transform the contingency table in a complete disjunctive table, then compute the parameters (see § 4.1.1) needed for the test.

The MCA on the four variables gives the following results (*Table 22* and *Figure 15*):

Table 22: Parameters needed for the test
(first step of the example for a mutual independence model).

n	p	m_1	m_2	m_3	m_4	q	m	σ	I_c
299	4	2	2	2	2	4	0.25	0.0250	[0.2000, 0.3000]

$\lambda_1 = 0.4145$	***********************************
$\lambda_2 = 0.2512$	********************
$\lambda_3 = 0.2449$	*******************
$\lambda_4 = 0.0894$	*********

Figure 15: MCA on X_1, X_2, X_3 and X_4
(eigenvalues diagram, first step of the example for a mutual independence model).

The eigenvalue diagram shows clearly that variables are not independent, only λ_2 and λ_3 are in the confidence interval.

Let's drop X_4 and use the second step of the procedure. MCA on the three remaining variables gives the following results (*Table 23* and *Figure 16*):

Table 23: MCA on X_1, X_2 and X_3
(parameters).

n	p	q	m	σ	I_c
299	3	3	0.3333	0.0273	[0.2787, 0.3879]

$\lambda_1 = 0.3639$	**********************
$\lambda_2 = 0.3342$	********************
$\lambda_3 = 0.3019$	*******************

Figure 16: MCA on X_1, X_2 and X_3
(eigenvalues diagram).

The eigenvalue diagram shows clearly that variables are independent, since all the eigenvalues are in the confidence interval, so there is surely one or more interaction X_4 and X_i, $i = 1, ..., 3$.

The MCA on X_4 and X_i gives the following results (*Table 24* and *Figure 17*):

Table 24: MCA on X_4, X_i
(parameters).

n	p	q	m	σ	I_c
299	2	2	0.5	0.0283	[0.4434, 0.5566]

X_4 and X_1	X_4 and X_2	X_4 and X_3
$\lambda_1 = 0.5365$ **********	$\lambda_1 = 0.8198$ **********	$\lambda_1 = 0.5058$ **********
$\lambda_2 = 0.4635$ *********	$\lambda_2 = 0.1802$ ***	$\lambda_2 = 0.4942$ *********

Figure 17: Eigenvalues diagram for MCA on X_4 and X_1,
MCA on X_4 and X_2 and MCA on X_4 and X_3.

It's clear that there exists only an interaction between X_4 and X_2, X_1 and X_3 are non dependent of X_4, then $u_{14} = u_{13} = 0$ and $u_{24} \neq 0$ and we build the crossed variable X_{24}.

The MCA of X_1, X_3 and X_{24} gives the following results (*Table 25* and *Figure 18*):

Table 25: MCA on X_1, X_3 and X_{24}
(parameters).

n	p	q	m	σ	I_c
299	3	5	0.3333	0.0273	[0.2787, 0.3879]

$\lambda_1 = 0.3647$	**********************
$\lambda_2 = 0.3624$	**********************
$\lambda_3 = 0.3333$	*********************
$\lambda_4 = 0.3047$	********************
λ_5=0.3016	********************

Figure 18: Eigenvalues diagram for MCA on X_1, X_3 and X_{24}.

The eigenvalue diagram shows that the variables are independent, all the eigenvalues being within the confidence interval, and there is no triple interaction between variables.

We finally obtain the same model as Andersen:

$$\log[f_4(X)] = u_0 + u_1 x_1 + u_2\, x_2 + u_3\, x_3 + u_4\, x_4 + x_{24} x_4 x_2 \ .$$

6. CONCLUSION

Log-linear modeling and MCA are two complementary techniques for the analysis of categorical data. In this framework, we propose a method for fitting progressively log-linear models, using the eigenvalue shape of MCA.

We show that, in MCA, under the independence hypothesis for the variables, each observed eigenvalue is asymptotically normally distributed. These distributions have the same mean, different variances and converge to normal distributions. In this case, the eigenvalue diagram takes a peculiar shape. This shape is different if there is one or more interactions between variables, and we can recognize the log-linear model fitted for the data in some special cases.

Then, based on these results, we propose a simple procedure for progressively fitting log-linear models, where the fitting criterion is based on MCA eigenvalue diagrams: the chosen model is constructed by successive utilizations of MCA (non constrained by the variables number). Finally, we validate this procedure on three sets of data drawn from the literature.

REFERENCES

[1] Agresti, A. (1990). *Categorical Data Analysis*, Wiley–Interscience.

[2] Andersen, E.B. (1991). *The Statistical Analysis of Categorical Data* (Second edition), Springer-Verlag.

[3] Baccini, A.; Mathieu, J.R. and Mondot, A.M. (1987). Comparaison sur un exemple, d'analyse des correspondances multiples et de modélisations, *Revue de Statistique Appliquée*, **XXXV**(3), 21–34.

[4] Ben Ammou, S. (1996). *Comportement des valeurs propres en analyse des correspondances multiples sous certaines hypothèses de modéles*, Doctoral Thesis of University Paris IX, Dauphine.

[5] Ben Ammou, S. and Saporta, G. (1998). Sur la normalité asymptotique des valeurs propres en ACM sous l'hypothèses d'indépendance des variables, *Revue de Statistique Appliquée*, **XLVI**(3), 21–35.

[6] Benzecri, J.P. (1973). *Analyse des Données [The Analysis of Data]* (2 vol), Paris: Dunod.

[7] Birch, M.W. (1963). Maximum likelihood in three-way contingency tables, *J. Royal Satist. Soc. (B)*, **25**, 220–233.

[8] Bishop, Y.M.M.; Fienberg, S.E. and Holland, P.W. with the collaboration of Light, R.J. and Mosteller, F. (1975). *Discrete Multivariate Analysis: Theory and Practice*, The MIT Press.

[9] Burt, C. (1950). The factorial analysis of qualitative data, *British J. of Statist. Psychol.*, **3**(3), 166–185.

[10] CHRISTENSEN, R. (1990). *Log-Linear Models*, Springer-Verlag, New York.

[11] DAUDIN, J.J. and TRECOURT, P. (1980). Analyse factorielle des correspondances et modéle log-linéaire: comparaison des deux méthodes sur un exemple, *Revue de Statistique Appliquée*, **XXVIII**(1).

[12] DOBSON, A. (1983). *An Introduction to Statistical Modelling*, Chapman and Hall, New York.

[13] DE FALGUEROLLES, A. and JMEL, S. (1993). Un modèle graphique pour la sélection de variables qualitatives, *Revue de Statistique Appliquée*, **XLI**(2), 23–41.

[14] DE FALGUEROLLES, A.; JMEL, S. and WHITTAKER, J. (1995). Correspondence analysis and association models constrained by a conditional independence graph, *Psychometrika*, **60**(2), 161–180.

[15] FIENBERG, S.E. (1980). *The Analysis of Cross-Classified Categorical Data*, MIT Press, Cambridge, Mass.

[16] GOODMAN, L.A. (1970). The multivariate analysis of qualitative data: interaction among multiple classifications, *J. of Amer. Statist. Assoc.*, **65**, 226–256.

[17] GOODMAN, L.A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories, *Journal of the American Statistical Association*, **74**, 537–552.

[18] GOODMAN, L.A. (1981). Association models and the bivariate normal for contingency tables with ordered categories, *Biometrika*, **68**, 347–355.

[19] GOODMAN, L.A. (1981). Association models and canonical correlation in the analysis of cross-classifications having ordered categories, *Journal of the American Statistical Association*, **76**, 320–334.

[20] GOODMAN, L.A. (1986). Some useful extensions of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables (with comments), *International Statistical Review*, **54**(3), 243–309.

[21] GOODMAN, L.A. (1991). Measures, models and graphical display in the analysis of cross-classified data, *Journal of the American Statistical Association*, **86**, 1085–1110.

[22] GIFI, A. (1990). *Non Linear Multivariate Analysis*, J. Wiley.

[23] GUTTMAN, L. (1941). *The quantification of a class of attributes: a theory and method of a scale construction.* In "The Prediction of Personal Adjustment" (P. Horst, Ed.), SSRC, New York, 251–264.

[24] HABERMAN, S.J. (1974). *The Analysis of Frequency Data*, University of Chicago, University Press, Chicago.

[25] HAYASHI, C. (1956). Theory and examples of quantification (II), *Proc. of Institute of Statist. Math.*, **4**(2), 19–30.

[26] VAN DER HEIJDEN, P.G.M. and DE LEEUW, J. (1985). Correspondence analysis used complementary to log-linear analysis, *Psychometrika*, **50**(4), 429–447.

[27] VAN DER HEIJDEN, P.G.M. and WORSLEY, K.J. (1986). Comment on "Correspondence analysis used complementary to log-linear Analysis", *Leiden Psychological Reports*, Psychometrics and Research Methodology, Department of Psychology, Leiden University – The Netherlands.

[28] van der Heijden, P.G.M.; de Falguerolles, A. and de Leeuw, J. (1989). A combined approach to contingency table analysis using correspondence analysis and log-linear analysis (with discussion), *Applied Statitics*, **38**, 249–292.

[29] Israëls, A. (1987). *Eigenvalue Techniques for Qualitative Data*, Leiden, The Netherlands, DSWO-Press.

[30] Jmel, S. (1991). Modeles graphiques, analyse en composantes principales et analyse des correspondances multiples: comparaisons sur des exemples. Poster présenté lors des *23èmes Journées de Statistique á Strasbourg*.

[31] Lauro, N.C. and Decarli, A. (1982). Correspondence analysis and log-linear models in multi-way contingency tables. Some remarks on experimental data, *Rivista Internationale di Statistica Metron*, **XL**(1–2).

[32] Lebart, L. (1976). The significance of eigenvalues issued from correspondence analysis, *COMPSTAT, Physica Verlag, Vienne*, 38–45.

[33] de Leeuw, J. (1984). *Canonical Analysis of Categorical Data* (Doctoral dissertation, University of Leiden, 1973), Leiden, DSWO-Press.

[34] O'Neill, M.E. (1978). Asymptotic distributions of the canonical correlations from contingency tables, *Austral. J. Statist.*, **20**(1), 75–82.

[35] O'Neill, M.E. (1978). Distributional expansion for canonical correlations from contingency tables, *J.R. Statist. Soc. B.*, **40**(3), 303–312.

[36] O'Neill, M.E. (1980). A note on the canonical correlations from contingency tables, *Austral. J. Statist.*, **20**(1), 58–66.

[37] Malinvaud, E. (1987). Data analysis in applied socio-economic statistics with special consideration of correspondence analysis, *Marketing Science Conference Proceedings, HEC-ISA, Joy en Josas.*

[38] Nishisato, S. (1980). *Analysis of Categorical Data. Dual Scaling and Its Application*, Univ. of Toronto Press.

[39] Novak, T.P. and Hoffman, D.L. (1990). Residual scaling: an alternative to correspondence analysis for the graphical representation of residuals from log-linear models, *Multivariate Behavioral Research*, **25**, 351–370.

[40] Siciliano, R. (1990). Asymptotic distribution of eigenvalues and statistical tests in non symmetric correspondence analysis, *Statistica Applicata*, **2**(3), 259–276.

[41] Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*, Wiley & Sons Ltd, England.

[42] Worlsley, K.J. (1987). Un exemple d'identification d'un modèle log-linéaire grâce à une analyse des correspondances (avec discussion), *Revue de Statistique Appliquée*, **XXXV**(3), 13–20.