

Scaling of Order Dependent Categorical Variables with Correspondence Analysis

B.F. Schriever

Mathematisch Centrum, Amsterdam, The Netherlands

Summary

In this paper we introduce successively stronger forms of ordinal dependence between categorical variables, corresponding to orderings over the categories of the variables. In our main theorem it is proved that if these forms of dependence are present in contingency tables, then the orderings are reflected in the correspondence analysis solution, whatever *a priori* ordering may have been given to the categories. This explains two important order phenomena which frequently occur in practice. Furthermore a multivariate generalization of the main theorem is given. The results in this paper support the use of (multi-) correspondence analysis as a scaling technique for categorical variables.

Key words: Contingency tables; Correspondence analysis; Multicorrespondence analysis; Order dependence; Scaling.

1 Introduction

There exist many ways of describing the association which is present in a contingency table. Hirschfeld (1935) introduced a method which was later (independently) formulated by a number of authors. Benzécri (1973) gives a description of this method under the now well-established name of correspondence analysis. The method can be described from several points of view. We formulate correspondence analysis as a method of scaling: it assigns q -dimensional scores to the categories of the variables describing the rows and columns of the contingency table. (Here q is some integer.) The scaling is performed in such a way that the scores of two row categories are close together if their corresponding rows are more similar, and similarly for columns. Furthermore, it is customary to display the correspondence analysis scores in a q -dimensional graphical representation. In this graphical representation each row and column is represented as a point, which has its q -dimensional correspondence analysis score as coordinates. We refer to the publications of Gifi (1981), Hill (1974), Kester & Schriever (1982) and Lebart, Morineau & Tabard (1977) for theoretical treatments and applications of correspondence analysis.

Correspondence analysis considers the two variables I and J , indicating row and column number of the contingency table, as nominal variables, i.e. under permutation of category labels, the scores undergo the same permutation, however the following phenomenon is frequently observed in practice. For variables whose categories have an intuitively meaningful order, this order is often reflected by the order of the one-dimensional correspondence analysis scores. In § 3.3 of the present paper we prove that this phenomenon is implied by a strong form of dependence between I and J called order dependence of order 1. This form of dependence induces an ordering over the categories of the variables I and J ; see § 3.2. Since the most important aspect of assigning scores to

categories is perhaps the ordering which is induced by these scores, this result supports the use of correspondence analysis as a one-dimensional scaling technique.

Another phenomenon which often occurs in practice is the so-called horseshoe. We speak of a horseshoe in the two-dimensional graphical representation of correspondence analysis when row points and column points lie on convex or concave curves. We prove that a horseshoe occurs when the two variables I and J have a still stronger form of dependence, called order dependence of order 2. In fact we prove a generalization of these results to higher orders in § 3.3.

Correspondence analysis can be generalized to the case when more than two variables are involved. This generalization is called multicorrespondence analysis by Lebart *et al.* (1977); it is also called homogeneity analysis (Gifi, 1981; Kester & Schriever, 1982), or first-order correspondence analysis (Hill, 1974). In the present paper we introduce a multivariate generalization of order dependence of order 1, and show that the order of the categories in each variable which corresponds to this multiorder dependence is reflected in the order of the one-dimensional multicorrespondence analysis scores. However, a similarly generalized multiorder dependence of order 2 need not imply horseshoes in a two-dimensional graphical representation of multicorrespondence analysis; see § 4.2.

In this paper we consider only correspondence analysis as applied to frequency tables (i.e. tables of relative frequencies or probabilities), disregarding problems of sampling variation. In § 5 we give examples of probability models for frequency tables in which the variables are (multi-) order dependent of order 2. In this case the frequency table is said to be $(M)DO_2$. This abundance of examples demonstrates that (multi-) order dependence of order 2 is quite common in practical models. Although this does not imply the $(M)DO_2$ character of random samples from such populations, one may nevertheless expect that contingency tables are also often $(M)DO_2$ or close to it, and hence that the order relations of correspondence analysis remain valid. It is however difficult to derive precise and useful statistical properties of such qualitative aspects of (multi-) correspondence analysis. For instance, the probability of obtaining correct orders or of obtaining horseshoes is easily shown to converge exponentially fast to one as the sample size tends to infinity, but such a result has little practical use.

A better argument for the practical relevance of our results when sampling variation should also be taken into account is the fact which we mentioned before that an intuitively expected ordering and the horseshoe phenomenon are so often found with small-sample real data. Even if, with real data, an obvious deviation from the intuitively expected ordering of categories is found, our experience is that a good explanation can usually be found for this. This makes (multi-) correspondence analysis a very useful exploratory tool for checking that an *a priori* ordering of the categories of a nominal variable is correct.

2 Total positivity

In this paper we make use of matrix theory. Some results of the theory of totally positive matrices are summarized in this section.

We denote matrices by capital letters. The (i, j) element of a matrix A is denoted by a_{ij} ; however the diagonal elements of a diagonal matrix are singly subscripted. Vectors are denoted by lower case letters and are considered as column vectors. The i th component of a vector x is denoted by x_i . The transpose of a matrix or vector is denoted by the superscript T .

The identity matrix is denoted by I and the vector having all its components equal to unity is denoted by e ; the size of this matrix and vector will be clear from the context.

For a matrix A of size $n \times m$ we denote by

$$A \begin{pmatrix} i_1 & i_2 & \dots & i_k \\ j_1 & j_2 & \dots & j_k \end{pmatrix} = \begin{pmatrix} a_{i_1 j_1} & a_{i_1 j_2} & \dots & a_{i_1 j_k} \\ a_{i_2 j_1} & a_{i_2 j_2} & \dots & a_{i_2 j_k} \\ \vdots & \vdots & \dots & \vdots \\ a_{i_k j_1} & a_{i_k j_2} & \dots & a_{i_k j_k} \end{pmatrix}$$

the determinant formed from the specified elements of A . This determinant is called a *minor* of A of order k if $1 \leq i_1 < i_2 < \dots < i_k \leq n$ and $1 \leq j_1 < j_2 < \dots < j_k \leq m$.

Definition 2.1. The matrix A is called *totally positive of order r* (abbreviated TP_r) if all minors of order $\leq r$ are positive. If all minors of order $\leq r$ are strictly positive, then A is said to be *strictly totally positive of order r* (STP_r).

LEMMA 2.1. If the matrix A_1 of size $n \times l$ is TP_r and the matrix A_2 of size $l \times m$ is TP_s , then the matrix $A_1 A_2$ is $TP_{\min(r,s)}$. In the case that A_1 is STP_r and A_2 is TP_s and of full rank, $A_1 A_2$ is actually $STP_{\min(r,s)}$.

Proof. The proof follows from the Binet–Cauchy formula (Gantmacher, 1977, **1**, p. 9):

$$A_1 A_2 \begin{pmatrix} i_1 & i_2 & \dots & i_k \\ j_1 & j_2 & \dots & j_k \end{pmatrix} = \sum_{1 \leq h_1 < h_2 < \dots < h_k \leq l} A_1 \begin{pmatrix} i_1 & i_2 & \dots & i_k \\ h_1 & h_2 & \dots & h_k \end{pmatrix} A_2 \begin{pmatrix} h_1 & h_2 & \dots & h_k \\ j_1 & j_2 & \dots & j_k \end{pmatrix}.$$

An important property of (strictly) totally positive matrices is the number of changes of sign of the eigenvectors. In counting the number of changes of sign (of the sequence of components) of a vector, zero components are permitted to take on arbitrary signs. So the number of changes of sign of a vector x will vary between two bounds \mathcal{P}_x^- and \mathcal{P}_x^+ .

In the next lemma the vectors $x^{(1)}, x^{(2)}, \dots, x^{(r)}$ denote the eigenvectors of an $n \times n$ matrix A corresponding to the r ‘largest’ eigenvalues $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_r|$, and X denotes the $n \times r$ matrix $X = (x^{(1)}, x^{(2)}, \dots, x^{(r)})$.

LEMMA 2.2. The r largest eigenvalues of an STP_r matrix A are strictly positive and distinct: $\lambda_1 > \lambda_2 > \dots > \lambda_r > |\lambda_{r+1}|$. Furthermore, for each $k = 1, \dots, r$ there is a σ_k equal to $+1$ or -1 such that

$$\sigma_k X \begin{pmatrix} i_1 & i_2 & \dots & i_k \\ 1 & 2 & \dots & k \end{pmatrix} > 0$$

for all $1 \leq i_1 < i_2 < \dots < i_k \leq n$, and for arbitrary real numbers c_k, c_{k+1}, \dots, c_l , ($1 \leq k \leq l \leq r$, $\sum_t c_t^2 > 0$, where the sum is over $t = k, \dots, l$), the bounds \mathcal{P}_x^- and \mathcal{P}_x^+ of the vector

$$x = \sum_{t=k}^l c_t x^{(t)}$$

satisfy

$$k - 1 \leq \mathcal{P}_x^- \leq \mathcal{P}_x^+ \leq l - 1.$$

Proof. This lemma is a weaker version of a result of Gantmacher & Krein (1950, p. 349).

Another related property is the variation diminishing property. This gives us a better intuitive grasp of total positivity.

LEMMA 2.3. Consider transformations of the form $x = Ay$, where A is a matrix of size $n \times m$.

(i) If A is TP_r , then, for all $y \in \mathbb{R}^m$,

$$\mathcal{S}_y^- \leq r-1 \Rightarrow \mathcal{S}_x^- \leq \mathcal{S}_y^-; \tag{2.1}$$

moreover, for all y such that

$$\mathcal{S}_x^- = \mathcal{S}_y^- \leq r-1, \tag{2.2}$$

the first nonzero component of x and y have the same sign.

(ii) If A is STP_r , then, for all $y \neq 0$,

$$\mathcal{S}_y^- \leq r-1 \Rightarrow \mathcal{S}_x^+ \leq \mathcal{S}_y^-. \tag{2.3}$$

(iii) Conversely, when $m < n$, then (2.2) and (2.3) imply that A is STP_r . If A is of full rank $m < n$, then (2.1) and (2.2) imply that A is TP_r .

Proof. The lemma is a special case of a result of Karlin (1968, p. 233).

Furthermore we also need the following lemma.

LEMMA 2.4. Any TP_r matrix of rank $\geq r$ can be approximated elementwise as closely as desired by means of an STP_r matrix of the same rank.

Proof. See Gantmacher & Krein (1950, p. 357).

3 Ordering properties in correspondence analysis

3.1 Correspondence analysis

Let P be a frequency table of size $n \times m$, that is P is an $n \times m$ matrix with positive real elements $p_{ij} \geq 0$ ($i = 1, \dots, n$; $j = 1, \dots, m$), such that

$$\sum_{i=1}^n \sum_{j=1}^m p_{ij} = 1.$$

Denote by

$$r_i = \sum_{j=1}^m p_{ij} \quad (i = 1, \dots, n), \quad c_j = \sum_{i=1}^n p_{ij} \quad (j = 1, \dots, m)$$

the row and column sums of P . Let these marginals of P form the diagonal elements of the diagonal matrixes R and C , respectively. We assume that R and C are nonsingular.

Let I and J denote the two variables indicating row and column number of the frequency table P . Note that the variables giving rise to the frequency table may be ordinal or nominal. Correspondence analysis is a technique for analysing the dependence between the two variables I and J . There are several ways to look at this technique; in the next definition we formulate it as a method of scaling.

Definition 3.1. A solution of correspondence analysis applied to the frequency table P consists of real vectors $u^{(t)} = (u_1^{(t)}, \dots, u_n^{(t)})^T$, called the *row factors*, and $v^{(t)} = (v_1^{(t)}, \dots, v_m^{(t)})^T$, called the *column factors*, for $t = 1, 2, \dots, \min(m, n)$, which satisfy

$$\lambda_t u^{(t)} = R^{-1} P v^{(t)}, \quad \lambda_t v^{(t)} = C^{-1} P^T u^{(t)},$$

where λ_t is maximal subject to

$$\begin{aligned} u^{(t)T} R u^{(t)} &= 1, & v^{(t)T} C v^{(t)} &= 1, \\ u^{(t)T} R u^{(s)} &= 0, & v^{(t)T} C v^{(s)} &= 0 \quad (s = 1, 2, \dots, t-1). \end{aligned}$$

Let P also denote the (empirical) joint distribution of I and J induced by the frequency table P . The components of the i th row of the matrix $R^{-1}P$ can be interpreted as the conditional probabilities $P\{J=j \mid I=i\}$. It follows from $\lambda_t u^{(t)} = R^{-1}P v^{(t)}$ that if two rows i and i' have (approximately) equal conditional distributions, the corresponding correspondence analysis scores $u_i^{(t)}$ and $u_{i'}^{(t)}$, $t = 1, 2, \dots, \min(m, n)$, are also (approximately) equal. Of course a similar property holds for columns j and j' which have (approximately) equal conditional distributions.

The solution of correspondence analysis can be found by solving eigenvalue problems. It can be proved (Hill, 1974; Kester & Schriever, 1982; Lebart *et al.*, 1977, p. 54; Schriever, 1982) that the vectors $u^{(t)}$ and $v^{(t)}$ in Definition 3.1 exist for $t = 1, 2, \dots, \min(m, n)$ and are eigenvectors of the matrices $R^{-1}PC^{-1}P^T$ and $C^{-1}P^TR^{-1}P$, respectively, corresponding to the eigenvalue λ_t^2 . Conversely, the eigenvectors, suitably normalized, of $R^{-1}PC^{-1}P^T$ and $C^{-1}P^TR^{-1}P$ corresponding to the eigenvalues $\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_{\min(m, n)}^2$ are row and column factors of the correspondence analysis solution. Clearly $u^{(t)}$ and $v^{(t)}$ are uniquely determined, up to a change of sign, when the corresponding eigenvalue λ_t^2 is simple. Furthermore it can be proved that $\lambda_1 = 1$ and that the first row and column factor can always be taken to be trivial, that is $u^{(1)} = e$ and $v^{(1)} = e$. In the sequel we assume, in particular in the case $1 = \lambda_1 = \lambda_2$, that $u^{(1)} = e$ and $v^{(1)} = e$.

Hill (1974) showed that correspondence analysis is algebraically equivalent to Fisher's contingency table analysis (Fisher, 1940). This gives us the following interpretation of the row and column factors. The first nontrivial row and column factor, $u^{(2)}$ and $v^{(2)}$, can be interpreted as 'optimal' scores of the categories of the variables I and J : they define derived variables with maximal correlation. The vectors $u^{(t)}$ and $v^{(t)}$ define scores with similar properties subject to orthogonality to previous sets of scores.

In practice not all $\min(m, n)$ factors are computed, but only the first q nontrivial, where q is an integer $< \min(m, n)$. Moreover, they are usually presented in one or more plots. These plots show aspects of the q -dimensional graphical representation of correspondence analysis where each row and each column of P is represented as a point; row i has coordinates $(\lambda_2 u_i^{(2)}, \lambda_3 u_i^{(3)}, \dots, \lambda_{q+1} u_i^{(q+1)})$, and column j has coordinates $(\lambda_2 v_j^{(2)}, \lambda_3 v_j^{(3)}, \dots, \lambda_{q+1} v_j^{(q+1)})$. Thus two row points (column points) lie close to one another if the corresponding conditional distributions are approximately equal. It can be proved (Gifi, 1981, p. 134; Kester & Schriever, 1982; Lebart *et al.*, 1977, p. 49) that when $q \geq \text{rank}(P) - 1$ the converse holds also; in the case that $q < \text{rank}(P) - 1$ it holds approximately.

Further results and properties of correspondence analysis can be found in the references given in this subsection.

3.2 Order dependence

In this subsection we introduce successively stronger forms of ordinal dependence between the variables I and J . These forms of dependence are called *order dependence of order r* ($r = 1, 2, 3, \dots$).

Suppose that the rows and columns of the frequency table P can be indexed such that the family of induced distributions of $J \mid I = i$ is stochastically (strictly) increasing, that is $P\{J \leq j_0 \mid I = i\}$ is (strictly) decreasing in i for each j_0 . Then the conditional distributions of the variable $J \mid I = i$ are stochastically ordered in a sequence which is identical to the order

given by the row index i . Lehmann (1966) speaks of positive regression dependence of J on I . In this case the $n \times m$ frequency table P satisfies

$$1 \leq i < i' \leq n \Rightarrow \sum_{j \leq i_0} p_{ij}/r_i \geq \sum_{j \leq i_0} p_{i'j}/r_{i'} \quad (j_0 = 1, \dots, m-1) \tag{3.1}$$

with strict inequality in case of strict regression dependence. In order to write (3.1) in matrix notation we introduce the upper triangular matrix S_n of size $n \times n$ with unit elements on and above the diagonal and with all other elements zero. The inverse of S_n is the matrix with unit elements on the (main) diagonal, with the elements adjacent and above the diagonal (i.e. on the first super diagonal) equal to -1 , and with all other elements zero.

Furthermore define

$$\bar{Q}_R = S_n^{-1} R^{-1} P S_m,$$

and let the $(n-1) \times (m-1)$ matrix Q_R be obtained by deleting the last row and column of \bar{Q}_R . Thus the (i, j) th element of Q_R equals

$$P\{J \leq j \mid I = i\} - P\{J \leq j \mid I = i+1\} \quad (i = 1, \dots, n-1; j = 1, \dots, m-1).$$

It follows that (3.1) is equivalent to Q_R is (S)TP₁.

A fundamental property of a stochastically increasing family is that it preserves monotonicity of functions. To be more specific, a vector $\psi = (\psi_1, \dots, \psi_m)^T$ is said to be *monotone of order r* , denoted by \mathcal{M}_r , if the vector of differences $\delta = (\delta_1, \dots, \delta_{m-1})^T$, where

$$\delta_j = \psi_j - \psi_{j+1} \quad (j = 1, \dots, m-1)$$

has a number of changes of sign which satisfies $\mathcal{S}_\delta^- = r-1$. The vector ψ is said to be *strictly monotone of order r* , denoted by \mathcal{SM}_r , if

$$\mathcal{S}_\delta^- = \mathcal{S}_\delta^+ = r-1.$$

Note that monotonicity is a kind of oscillatory property. Now, if we define for any vector ψ the vector $\phi = (\phi_1, \dots, \phi_n)^T$ of expectations by

$$\phi_i = \sum_{j=1}^m \psi_j p_{ij}/r_i \quad (i = 1, \dots, n)$$

then a stochastically (strictly) increasing family of distributions of $J \mid I = i$ is characterized by

$$\psi \text{ is } (\mathcal{S})\mathcal{M}_1 \Rightarrow \phi \text{ is } (\mathcal{S})\mathcal{M}_1$$

(see Theorem 3.1 below); i.e. the family preserves (strictly) order 1 monotonicity of functions.

In the case that the stochastically increasing family also preserves order 2 monotonicity of functions, the distribution functions of the family are ordered according to a stronger criterium. The difference between consecutive distribution functions is not only always positive, as in (3.1), but must also have further regularity properties. We say, in short, that the ordering of distributions in the family is stronger.

To illustrate that this stronger ordering of distributions is natural, consider the problem of testing hypotheses about the parameter (row number) of the one-parameter family of distributions based on one observation of the variable (column number). If the family is stochastically increasing, then any one-sided test has a monotone (of order 1) power function, as one would hope. In the case that this family also preserves order 2 monotonicity of functions, then any two-sided test has the desirable property that its power function is monotone of order ≤ 2 , i.e. first decreases then increases.

In the case that the family of conditional distributions of $J | I = i$ preserves order 2 monotonicity of functions, we have

$$\psi \text{ is } \mathcal{M}_r \Rightarrow \phi \text{ is } \mathcal{M}_s,$$

where $s \leq r$ and $r = 1, 2$.

Generally, the ordering of the distributions of $J | I = i$ becomes successively stronger if the family preserves order r monotonicity of functions for $r = 1, 2, \dots$. The next theorem shows that this can be formulated by means of the matrix Q_R .

THEOREM 3.1. *If the matrix Q_R is TP_r , then*

$$\psi \text{ is } \mathcal{M}_t \Rightarrow \phi \text{ is } \mathcal{M}_s,$$

where $1 \leq s \leq t$ and $t = 1, 2, \dots, r$.

Proof. Let $\varepsilon = (\varepsilon_1, \dots, \varepsilon_{n-1})^T$ be the vector of differences

$$\varepsilon_i = \phi_i - \phi_{i+1} \quad (i = 1, \dots, n-1),$$

and let $\delta = (\delta_1, \dots, \delta_{m-1})^T$ be the vector of differences

$$\delta_j = \psi_j - \psi_{j+1} \quad (j = 1, \dots, m-1).$$

Then we have that $\varepsilon = Q_R \delta$. The result now follows from Lemma 2.3 (i).

Note that under slight nondegeneracy conditions the converse holds also. In the case that Q_R is STP_r , we have

$$\psi \text{ is } \mathcal{SM}_t, \phi \text{ is } \mathcal{M}_t \Rightarrow \phi \text{ is } \mathcal{SM}_t \quad (t = 1, 2, \dots, r).$$

We have described forms of regression dependence of J on I such that the conditional distributions of $J | I = i$ are successively more strongly ordered with respect to the categories of I . Similar ordering properties hold with respect to the categories of J when the $(m-1) \times (n-1)$ matrix

$$Q_C \text{ is } (s)\text{TP}_r \quad (r = 1, 2, \dots),$$

where the matrix Q_C is obtained by deleting the last row and column of

$$\bar{Q}_C = S_m^{-1} C^{-1} P^T S_n.$$

If the categories of I and J can be indexed such that both conditions hold, we have a strong ordinal relation between the two categorical variables. In this case we say that the variables I and J are (strictly) order dependent of order r . The frequency table P then satisfies, possibly after a permutation of rows and columns, the following definition.

Definition 3.2. The frequency table P is called (strictly) doubly ordered of order r (abbreviated (s)DO $_r$) if the matrices Q_R and Q_C are both (s)TP $_r$ and have rank $\geq r$.

The next theorem gives sufficient conditions for a frequency table P to be (s)DO $_r$.

THEOREM 3.2. *If P is (s)TP $_{r+1}$ and $\text{rank}(P) \geq r+1$ then P is (s)DO $_r$.*

Proof. First consider the case that P is STP_{r+1} . Since S_m is TP_m and has rank m , it follows from Lemma 2.1 that $R^{-1}PS_m$ is STP_{r+1} . Hence in particular

$$R^{-1}PS_m \begin{pmatrix} i & i+1 & \dots & i+k-1 & i+k \\ j_1 & j_2 & \dots & j_k & m \end{pmatrix} = Q_R \begin{pmatrix} i & i+1 & \dots & i+k-1 \\ j_1 & j_2 & \dots & j_k \end{pmatrix} > 0$$

for $1 \leq j_1 < j_2 < \dots < j_k \leq m-1$; $i = 1, \dots, n-k-1$ and $k = 1, \dots, r$.

Application of the result of Karlin (1968, p. 60) for fixed indices j_1, \dots, j_k yields

$$Q_R \begin{pmatrix} i_1 & i_2 & \dots & i_k \\ j_1 & j_2 & \dots & j_k \end{pmatrix} > 0$$

for $1 \leq i_1 < i_2 < \dots < i_k \leq n-1$ and $k = 1, \dots, r$. It follows that Q_R is STP_r . The restriction that all minors are strictly positive can be dropped by appealing to Lemma 2.4 and continuity. The same arguments hold with respect to the matrix Q_C .

3.3 Correspondence analysis and order dependence

In this subsection we show that the ordering over the categories of the variables I and J , which corresponds to order dependence, is reflected in the ordering of the components of the correspondence analysis row and column factors.

THEOREM 3.3. *Let the $n \times m$ frequency table P be SDO_r , then correspondence analysis applied to P yields:*

- (i) *eigenvalues $1 = \lambda_1 \geq \lambda_2 > \lambda_3 > \dots > \lambda_{r+1} > \lambda_{r+2}$,*
- (ii) *row and column factors $u^{(t)}$ and $v^{(t)}$ which are \mathcal{SM}_{t-1} and start oscillating in the same direction, for $t = 2, 3, \dots, r+1$. Moreover, for arbitrary real numbers c_k, c_{k+1}, \dots, c_l ($2 \leq k \leq l \leq r+1$, $\sum_t c_t^2 > 0$, where the sum is over $t = k, \dots, l$), the vectors*

$$u = \sum_{t=k}^l c_t u^{(t)}, \quad v = \sum_{t=k}^l c_t v^{(t)}$$

are \mathcal{M}_s , where $k-1 \leq s \leq l-1$.

- (iii) *In the case that $r \geq 2$, the row points (column points) in the two-dimensional graphical representation of correspondence analysis lie on a strictly convex or a strictly concave curve.*

Proof. Let

$$\bar{Q} = S_n^{-1} R^{-1} P C^{-1} P^T S_n$$

and let Q be obtained by deleting the last row and column of \bar{Q} . We have that $Q = Q_R Q_C$, since the (i, m) elements of \bar{Q}_R vanish for $i = 1, \dots, n-1$.

We first prove that for $t = 2, 3, \dots, \min(m, n)$ the vector $x^{(t)} = (x_1^{(t)}, \dots, x_{n-1}^{(t)})^T$ is an eigenvector of Q corresponding to the eigenvalue λ_t^2 if and only if the eigenvector $u^{(t)}$ of $R^{-1} P C^{-1} P^T$ satisfies

$$x_i^{(t)} = u_i^{(t)} - u_{i+1}^{(t)} \quad (i = 1, \dots, n-1).$$

Note that $u^{(t)}$ is an eigenvector of $R^{-1} P C^{-1} P^T$ corresponding to λ_t^2 if and only if $\bar{x}^{(t)} = S_n^{-1} u^{(t)}$ is an eigenvector of \bar{Q} corresponding to λ_t^2 . Since $R^{-1} P C^{-1} P^T$ has row sums equal to unity, the elements \bar{q}_{in} for $i = 1, \dots, n-1$ vanish. Hence the vector $(0, \dots, 0, 1)^T$ is an eigenvector of \bar{Q} corresponding to $\lambda_1^2 = 1$. For $t = 2, 3, \dots, \min(m, n)$ we have that $\bar{x}^{(t)} = (\bar{x}_1^{(t)}, \dots, \bar{x}_{n-1}^{(t)}, \bar{x}_n^{(t)})^T$ is an eigenvector of \bar{Q} corresponding to λ_t^2 if and only if $x^{(t)} = (\bar{x}_1^{(t)}, \dots, \bar{x}_{n-1}^{(t)})^T$ is an eigenvector of Q corresponding to λ_t^2 and $(\lambda_t^2 - 1)\bar{x}_n^{(t)} = \sum_i \bar{x}_i^{(t)}$, where the sum is over $i = 1, \dots, n-1$. From $\bar{x}^{(t)} = S_n^{-1} u^{(t)}$ it follows that

$$x_i^{(t)} = \bar{x}_i^{(t)} = u_i^{(t)} - u_{i+1}^{(t)} \quad (i = 1, \dots, n-1).$$

Note that λ_2^2 is the largest eigenvalue of Q .

Since Q_R and Q_C are STP_r , it follows from Lemma 2.1 that Q is STP_r , also. Application of

Lemma 2.2 yields $\lambda_2^2 > \lambda_3^2 > \dots > \lambda_{r+1}^2 > \lambda_{r+2}^2$, and that for arbitrary real numbers c_k, c_{k+1}, \dots, c_l ($2 \leq k \leq l \leq r+1$), the vector

$$x = \sum_{t=k}^l c_t x^{(t)}$$

satisfies $k-2 \leq \mathcal{P}_x^- \leq \mathcal{P}_x^+ \leq l-2$. Furthermore, in the case that $r \geq 2$ we have

$$\sigma_2 X \begin{pmatrix} i & i' \\ 2 & 3 \end{pmatrix} > 0 \quad (1 \leq i < i' \leq n-1);$$

that is

$$1 \leq i < i' \leq n-1 \Rightarrow \sigma_2 x_i^{(3)} / x_i^{(2)} < \sigma_2 x_{i'}^{(3)} / x_{i'}^{(2)},$$

and hence the row points in the two-dimensional graphical representation lie on a strictly convex or strictly concave curve.

Since the same arguments hold for the matrix $Q_C Q_R$ with eigenvectors $y^{(t)} = (y_1^{(t)}, \dots, y_{m-1}^{(t)})^T$, where

$$y_j^{(t)} = v_j^{(t)} - v_{j+1}^{(t)} \quad (j = 1, \dots, m-1),$$

similar results hold for the column factors $v^{(t)}$ ($t = 2, 3, \dots, r+1$). It follows from (2.2) that $u^{(t)}$ and $v^{(t)}$ start oscillating in the same direction.

Note that in the case that $1 = \lambda_1 = \lambda_2 > \lambda_3$, the vectors $u^{(2)}$ and $v^{(2)}$ are uniquely determined, since we agreed that $u^{(1)} = e, v^{(1)} = e$ always.

THEOREM 3.4. *Let the $n \times m$ frequency table P be DO_r , then there exist row and column factors, $u^{(t)}$ and $v^{(t)}$, of correspondence analysis applied to P such that $u^{(t)}$ and $v^{(t)}$ are \mathcal{M}_{t-1} and start oscillating in the same direction, for $t = 2, 3, \dots, r+1$. Moreover for arbitrary real numbers c_k, c_{k+1}, \dots, c_l ($2 \leq k \leq l \leq r+1, \sum_t c_t^2 > 0$, where the sum is over $t = k, \dots, l$), the vectors*

$$u = \sum_{t=k}^l c_t u^{(t)}, \quad v = \sum_{t=k}^l c_t v^{(t)}$$

are \mathcal{M}_s , where $k-1 \leq s \leq l-1$. Furthermore, in the case that $r \geq 2$, there exists a two-dimensional graphical representation of correspondence analysis such that the row points (column points) lie on a convex or concave curve.

Proof. The proof follows from Theorem 3.3, Lemma 2.4 and continuity considerations.

Remark 1. Note that when λ_r^2 is a simple eigenvalue, then $u^{(t)}$ and $v^{(t)}$ are uniquely determined (up to a change of sign) and thus have the stated monotonicity property.

Remark 2. Gantmacher & Krein (1950) prove Lemma 2.2 in the case that the matrix A is TP_r and has some power which is STP_r . The conditions of Theorem 3.3 imply that $Q_R Q_C$ and $Q_C Q_R$ are both STP_r . Since we only need that some powers of these matrices are STP_r , these conditions are somewhat too strong. However, it seems hard to find simple sufficient conditions for Theorem 3.3 which are essentially weaker.

As was mentioned in the Introduction, the two most important consequences of Theorems 3.3 and 3.4 are as follows. First, when the rows and columns of the frequency table are permuted, the components of the row and column factors undergo the same permutation. It follows from Theorems 3.3 and 3.4 that if a permutation of rows and columns exists such that P is SDO_r , there exists no other permutation such that P is DO_r .

This unique permutation is then determined by the order of the components of $u^{(2)}$ and $v^{(2)}$. Thus when the two variables I and J are (strictly) order dependent of order ≥ 1 the ordering over the categories of I and J corresponding to the dependence is reflected in the order of the components of the first nontrivial row and column factor, $u^{(2)}$ and $v^{(2)}$. This supports the use of the components of $u^{(2)}$ and $v^{(2)}$ as scores for the categories of the variables I and J respectively.

Secondly, if the two variables are (strictly) order dependent of order ≥ 2 , a horseshoe occurs in the two-dimensional graphical representation of correspondence analysis.

4 Ordering properties in multicorrespondence analysis

4.1 Multicorrespondence analysis

Correspondence analysis can be generalized to the case that more than two variables are involved. In order to see how this can be done, we give the following equivalent formulation of correspondence analysis.

LEMMA 4.1. *The vectors $(u^{(t)T}, v^{(t)T})^T = (u_1^{(t)}, \dots, u_n^{(t)}, v_1^{(t)}, \dots, v_m^{(t)})^T$, where $u^{(t)}$ and $v^{(t)}$ are solutions of correspondence analysis applied to P corresponding to λ_t , are eigenvectors of the matrix*

$$B = \begin{pmatrix} R^{-1} & 0 \\ 0 & C^{-1} \end{pmatrix} \begin{pmatrix} R & P \\ P^T & C \end{pmatrix}$$

corresponding to an eigenvalue $1 + \lambda_t$ for $t = 1, 2, \dots, \min(m, n)$.

Proof. Trivial.

Note that $(u^{(t)T}, -v^{(t)T})^T$ is an eigenvector of B corresponding to an eigenvalue $1 - \lambda_t$, for $t = 1, 2, \dots, \min(m, n)$. Furthermore if $\min(m, n) = m$, the vectors $(u^{(t)T}, 0^T)^T$ are eigenvectors of B corresponding to an eigenvalue 1 ($t = m + 1, \dots, n$).

Now consider the case that we have a k -dimensional frequency table P of size $m_1 \times m_2 \times \dots \times m_k$. Let the variables J_1, J_2, \dots, J_k denote the variables indicating the category numbers on the dimensions $1, 2, \dots, k$ of P respectively. Furthermore, let P_{jl} denote the $m_j \times m_l$ marginal bivariate frequency table of the variables J_j and J_l , for $j, l = 1, \dots, k$. Note that

$$P_{jl} = P_{lj}^T \quad (j, l = 1, \dots, k).$$

Denote by C_j the diagonal matrix P_{jj} ($j = 1, \dots, k$), and denote by C the diagonal matrix of size $m \times m$, where $m = \sum_j m_j$ with the sum over $j = 1, \dots, k$, with diagonal elements the diagonal elements of C_1, C_2, \dots, C_k . Assume that C is nonsingular.

Definition 4.1. A solution of multicorrespondence analysis applied to the k -dimensional frequency table P consists of real vectors

$$v^{(1,t)} = (v_1^{(1,t)}, \dots, v_{m_1}^{(1,t)})^T, \dots, v^{(k,t)} = (v_1^{(k,t)}, \dots, v_{m_k}^{(k,t)})^T$$

called the *variable factors*, for $t = 1, 2, \dots, m$, such that the vectors $v^{(t)} = (v^{(1,t)T}, \dots, v^{(k,t)T})^T$ satisfy

$$\lambda_t v^{(t)} = C^{-1} \begin{pmatrix} C_1 & P_{12} & \dots & P_{1k} \\ P_{21} & C_2 & \dots & P_{2k} \\ \vdots & \vdots & \dots & \vdots \\ P_{k1} & P_{k2} & \dots & C_k \end{pmatrix} v^{(t)}, \tag{4.1}$$

where λ_t is maximal subject to

$$v^{(t)\top} C v^{(t)} = k, \quad v^{(t)\top} C v^{(s)} = 0 \quad (s = 1, 2, \dots, t-1).$$

Let the $m \times m$ matrix B denote the product of the two matrices written on the right-hand side of (4.1). Note that the eigenvalue problem (4.1) is equivalent to correspondence analysis applied to the symmetric matrix CB , only the row and column factors differ by a factor $k^{-\frac{1}{2}}$ from the variable factors $v^{(t)}$ and the eigenvalues differ by a factor k^{-1} ($t = 1, 2, \dots, m$). Therefore, all the eigenvalues λ_t of multicorrespondence analysis are positive and the first variable factor can be taken trivially as $v^{(1)} = e$ with $\lambda_1 = k$. Furthermore it follows that

$$\lambda_t e^\top C_j v^{(j,t)} = e^\top C_j (C_j^{-1} P_{j1}, \dots, C_j^{-1} P_{jk}) v^{(t)} = e^\top C v^{(t)} = 0 \quad (j = 1, \dots, k; t = 2, 3, \dots, m).$$

Thus the nontrivial factors of each variable are also centred.

The interpretation of multicorrespondence analysis variable factors differs somewhat from the interpretation of the row and column factors in correspondence analysis. The first nontrivial variable factors $v^{(j,2)}$ ($j = 1, \dots, k$) can be interpreted as 'optimal' scores for the categories of the variables J_1, \dots, J_k : they define derived variables such that the first principal component of their correlation matrix has maximal variance (Hill, 1974). Note that the technique considers only marginal bivariate associations.

Similarly to correspondence analysis, the variable factors $v^{(j,t)}$ ($j = 1, \dots, k$; $t = 2, 3, \dots, q+1$) of multicorrespondence analysis are displayed in a q -dimensional graphical representation.

For further results and properties of multicorrespondence analysis we refer to Gifi (1981), Kester & Schriever (1982) and Lebart *et al.* (1977).

4.2 Multicorrespondence analysis and multiorder dependence

In § 3.3 it was proved that (strict) order dependence of order r in the frequency table implies that for $t = 1, \dots, r$ the t th nontrivial row and column factor are (strictly) monotone of order t . In the present subsection we show that a similar property holds for multicorrespondence analysis with respect to a (strictly) multiorder dependence of order 1. However, it need not hold with respect to multiorder dependence of higher order.

Definition 4.2. The k -dimensional frequency table P is called (strictly) *multivariate doubly ordered of order r* , abbreviated (S)MDO $_r$, if all the marginal bivariate frequency tables P_{jl} ($j \neq l$; $j, l = 1, \dots, k$) are (S)DO $_r$.

The variables J_1, J_2, \dots, J_k are called (strictly) *multiorder dependent of order r* if the categories of the variables can be indexed such that P is (S)MDO $_r$. This form of dependence considers only marginal bivariate associations.

THEOREM 4.2. *Let the k -dimensional frequency table P be SMDO $_1$, then multicorrespondence analysis applied to P yields:*

- (i) eigenvalues $k = \lambda_1 \geq \lambda_2 > \lambda_3$,
- (ii) first nontrivial variable factors $v^{(j,2)}$ ($j = 1, \dots, k$) which are all strictly increasing or all strictly decreasing in their components (i.e. they are all \mathcal{SM}_1 and start oscillating in the same direction).

Proof. Let S denote the $m \times m$ block matrix with diagonal blocks $S_{m_1}, S_{m_2}, \dots, S_{m_k}$ and all off-diagonal blocks zero. Furthermore, let \bar{Q} denote the $m \times m$ matrix $\bar{Q} = S^{-1}BS$, and let the matrix Q of size $(m-k) \times (m-k)$ be obtained by deleting the rows and columns

corresponding to the k indices $m_1, m_1+m_2, \dots, m_1+m_2+\dots+m_k$. Similarly to the first part of the proof of Theorem 3.3 it follows that (ii) holds if and only if the eigenvector of Q corresponding to the largest eigenvalue λ_2 is strictly negative or strictly positive.

Since the marginal frequency tables P_{jl} ($j \neq l$; $j, l = 1, \dots, k$) are SDO_1 it follows that the elements of Q are positive; the elements of Q are even strictly positive except on diagonal blocks. It follows that Q^2 is STP_1 .

Application of Lemma 2.2 with $r=1$ (i.e. the theorem of Perron–Frobenius) to Q^2 yields the result (ii) and $\lambda_2 > \lambda_3$.

THEOREM 4.3. *Let the k -dimensional frequency table P be MDO_1 ; then there exist variable factors $v^{(i,2)}$ ($j = 1, \dots, k$) of multicorrespondence analysis applied to P which are all increasing or all decreasing in their components (i.e. they are all \mathcal{M}_1 and start oscillating in the same direction).*

Proof. The proof follows from application of Lemma 2.4, Theorem 4.2 and continuity considerations.

These theorems show that the order of the components of the variable factors $v^{(i,2)}$ reflect the correct ordering of categories, in the case that the variables J_1, \dots, J_k are (strictly) multi-order dependent of order ≥ 1 . This supports the use of multicorrespondence analysis as a one-dimensional scaling technique.

We now briefly explain why these results can not be extended to multiorder dependence of order 2, i.e. the variable factors $v^{(i,3)}$ ($j = 1, \dots, k$) of multicorrespondence analysis applied to a k -dimensional frequency table P need not be \mathcal{SM}_2 when P is SMDO_2 . For instance, suppose that $v^{(i,3)}$ is \mathcal{SM}_2 for $j = 2, 3, \dots, k$. It follows from the eigenvalue problem (4.1) that

$$(\lambda_3 - 1)v^{(1,3)} = \sum_{j=2}^k C_1^{-1} P_{1j} v^{(j,3)}.$$

Although the vectors $C_1^{-1} P_{1j} v^{(j,3)}$ ($j = 2, 3, \dots, k$) are \mathcal{SM}_2 , their sum $v^{(1,3)}$ need not be \mathcal{SM}_2 . Examples of this can be given. But if the vectors $C_1^{-1} P_{1j} v^{(j,3)}$ ($j = 2, 3, \dots, k$) all attain their maximum (minimum) at the same place, then $v^{(1,3)}$ is actually \mathcal{SM}_2 .

Furthermore, since $v^{(i,t)}$ and $v^{(i,s)}$, $t \neq s$, need not be orthogonal with respect to C_j it is possible that $v^{(i,3)}$ is also \mathcal{SM}_1 .

Even in the case that stronger forms of multivariate dependence are present in the k -dimensional table P , for example the form of dependence such that for every pair of variables J_j and J_l , for $j \neq l$, the conditional joint distribution of J_j and J_l given the remaining variables is SDO_r , or the still stronger form in which P is STP_{r+1} in every pair of variables where the remaining variables are kept fixed, then we still have that the variable factors $v^{(i,t)}$ need not be \mathcal{SM}_{t-1} , for $t = 3, \dots, r$. An example of this is a discretization of a particular four-dimensional normal distribution given by Gifi (1981, p. 370).

These considerations indicate that we need some unnatural condition in addition to multivariate ordinal dependence in order to get higher order monotonicity results for multicorrespondence analysis. It is not the definition of multiorder dependence which is at fault, but the actual technique multicorrespondence analysis. It might be more appropriate to seek for another technique which analyses multivariate ordinal dependence using more information than just bivariate marginals.

5 (Multi-) order dependence in practice

In this section we give two important examples of probability models for (s)DO frequency tables. These examples can easily be extended to the multivariate case. The proofs of the

results mentioned in this section are given by Schriever (1982). In these examples it is easier to verify that the frequency tables are (s)TP_{r+1} for some r , which implies, by Theorem 3.2, that they are (s)DO_r.

A class of probability models for frequency tables is obtained by making a *discretization of bivariate density functions*. Let f be a bivariate density function with respect to a product measure $\sigma_1 \times \sigma_2$ on \mathbb{R}^2 . The frequency table P is said to be a discretization of f if there exists two partitions $\{E_i\}$, for $i = 1, \dots, n$, and $\{F_j\}$, for $j = 1, \dots, m$, of \mathbb{R} such that

$$p_{ij} = \int_{E_i} \int_{F_j} f(x, y) d\sigma_2(y) d\sigma_1(x), \quad \sigma_1(E_i) > 0, \quad \sigma_2(F_j) > 0$$

$$(i = 1, \dots, n; j = 1, \dots, m).$$

It turns out that discretizations of bivariate densities are actually STP_{min(m,n)}, or have some power which is STP_{min(m,n)}, when the density f is (s)TP and the elements of the partitions $\{E_i\}$ and $\{F_j\}$ are ordered correctly. Examples are: the bivariate normal, the trinomial, the negative trinomial and various types of the bivariate F , the bivariate gamma, the bivariate beta, the bivariate logistic, the bivariate Pareto, the bivariate Poisson and the bivariate hypergeometric distribution.

Another more specific example for a frequency table P is the *linear by linear interaction model*

$$\log p_{ij} = \mu + \alpha_i + \beta_j + \gamma_i \delta_j \quad (i = 1, \dots, n; j = 1, \dots, m)$$

where

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_i = \sum_j \delta_j = 0.$$

The frequency table P is STP_{min(m,n)} if the rows and columns are indexed such that γ_i and δ_j are both strictly increasing or both strictly decreasing in their indices. Goodman (1981) compares maximum likelihood estimates of γ_i and δ_j in this model with the first nontrivial row and column factor of correspondence analysis. Furthermore he discusses the ordering of rows and columns which is present in this model by means of the TP₂ and DO₁ property, however he does not prove that this ordering is reflected in the first nontrivial row and column factor of correspondence analysis.

Acknowledgements

I am very grateful to Dr. R.D. Gill and Prof. dr. J. Oosterhoff for their suggestions and support of this work. I also thank Prof. dr. J. De Leeuw for his valuable suggestions.

References

- Benzécri, J.P. (1973). *L'Analyse des données II: l'analyse des correspondances*. Paris: Dunod.
- Fisher, R.A. (1940). The precision of discriminant functions. *Ann. Eugen.* **10**, 422–429.
- Gantmacher, F.R. (1977). *Matrix Theory*, **1, 2**. New York: Chelsea.
- Gantmacher, F.R. & Krein, M.G. (1950). *Oscillation Matrices and Kernels and Small Vibrations of Mechanical Systems*; transl. from Russian, issued (1961), AEC-tr-4481, by US Atomic Energy Commission, Washington.
- Gifi, A. (1981). Non-linear multivariate analysis. Dept. of Data Theory, University of Leiden.
- Goodman, L.A. (1981). Association models and canonical correlation in the analysis of cross-classifications having ordered categories. *J. Am. Statist. Assoc.* **76**, 320–334.
- Hill, M.O. (1974). Correspondence analysis: a neglected multivariate method. *Appl. Statist.* **23**, 340–354.
- Hirschfeld, H.O. (1935). A connection between correlation and contingency. *Proc. Camb. Phil. Soc.* **31**, 520–524.
- Karlin, S. (1968). *Total Positivity*, **1**. Stanford University Press.
- Kester, N.K. & Schriever, B.F. (1982). *Analysis of Association of Categorical Variables by Numerical Scores and Graphical Representation*, report SW85. Amsterdam: Mathematisch Centrum.
- Lebart, L., Morineau, A. & Tabard, N. (1977). *Techniques de la description statistique*. Paris: Dunod.
- Lehmann, E.L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37**, 1137–1153.
- Schriever, B.F. (1982). *Ordering Properties in Correspondence Analysis*, report SW80. Amsterdam: Mathematisch Centrum.

Résumé

Dans cet article nous introduisons des façons successivement plus fortes de dépendance ordinale entre deux variables catégoriques qui induisent des ordres successivement plus forts des modalités des variables. Dans notre théorème principal s'avère que dans le cas où ces façons de dépendance figurent aux tableaux de contingence, les ordres sont reflétés dans la solution de l'analyse des correspondances. Cela explique deux phénomènes d'ordre importants qui se présentent souvent en pratique. En plus on donne une généralisation multivariante du théorème principal. Les résultats de cet article supportent l'application de l'analyse des correspondances (multiples) comme une technique qui construit des échelles pur des variables catégoriques.

[Paper received May 1982, revised January 1983]