stichting
mathematisch
centrum

AFDELING MATHEMATISCHE STATISTIEK SW 83/82 APRIL
(DEPARTMENT OF MATHEMATICAL STATISTICS)

B.F. SCHRIEVER

SCALING OF ORDER DEPENDENT CATEGORICAL VARIABLES WITH CORRESPONDENCE ANALYSIS

Preprint

kruislaan 413 1098 SJ amsterdam

Printed at the Mathematical Centre, 413 Kruislaan, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

1980 Mathematics subject classification: Primary: 62H20, 62H17
Secondary: 6207

Scaling of order dependent categorical variables with correspondence analysis [*)]

by

B.F. Schriever

ABSTRACT

In this paper we introduce forms of dependence between categorical variables, which induce successively stronger 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, the orderings are reflected in the correspondence analysis solution. 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 & PHRASES: *Correspondence analysis; multi-correspondence analysis; contingency tables; order dependence; scaling*

---

*) This report will be submitted for publication elsewhere.

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 et al. (1977) for theoretical treatments and applications of correspondence analysis.

Although correspondence analysis considers the two variables I and J, indicating row and column number of the contingency table, as nominal variables, the following phenomenon is frequently observed in practice. For variables, whose categories have an intuitive meaningfull order, this order is often reflected by the order of the one-dimensional correspondence analysis scores. In section 3.3 of this 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 section 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 section 3.3.

Correspondence analysis can be generalized to the case when more than two variables are involved. This generalization is called multi-correspondence analysis (LEBART et al.(1977); it is also called homogeneity analysis GIFI (1981), KESTER & SCHRIEVER (1982), or first order correspondence analysis, HILL (1974)). In this paper we introduce a multivariate generalization of order dependence of order 1, and show that the order of the categories in each variable which is induced by this multi-order dependence is reflected in the one-dimensional multi-correspondence analysis scores. However, a similarly generalized multi-order dependence of order 2 need not imply horseshoes in a two-dimensional graphical representation of multi-correspondence analysis (see section 4.2).

In this paper we only consider correspondence analysis as applied to frequency tables (i.e. tables of relative frequencies or probabilities). In section 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 $(M)DO_2$. The 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. This explains why the earlier mentioned phenomena are often found with real data.

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 singlely 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

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

the determinant formed from the specified elements of A. This determinant is called a *minor* of A of order k if

$$1 \le i_1 < i_2 < \dots < i_k \le n$$

and

$$1 \le j_1 < j_2 < \dots < j_k \le m.$$

<u>DEFINITION 2.1</u>. The matrix A is called *totally positive of order* r (abbreviated TP_r) if all minors of order $\le r$ are positive. If all minors of order $\le r$ are strictly positive, then A is said to be *strictly totally positive of order* r (STP_r).

<u>LEMMA 2.1</u>. *If the matrix* A_1 *of size* $n \times \ell$ *is* TP_r *and the matrix* A_2 *of size* $\ell \times m$ *is* TP_s, *then the matrix* A_1A_2 *is* $TP_{\min(r,s)}$. *In the case that* A_1 *is* STP_r *and* A_2 *is* TP_r *and of full rank,* A_1A_2 *is actually* $STP_{\min(r,s)}$.

<u>PROOF</u>. The proof follows from the Binet-Cauchy formula (cf. GANTMACHER, (1977), vol I, p.9)

$$A_1A_2\begin{pmatrix} i_1 i_2 \dots i_k \\ j_1 j_2 \dots j_k \end{pmatrix} = \sum_{1 \le h_1 < h_2 < \dots < h_k \le \ell} A_1\begin{pmatrix} i_1 i_2 \dots i_k \\ h_1 h_2 \dots h_k \end{pmatrix} A_2\begin{pmatrix} h_1 h_2 \dots h_k \\ j_1 j_2 \dots j_k \end{pmatrix}. \qquad \square$$

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 S_x^- and S_x^+.

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

$$X = (x^{(1)}, x^{(2)}, \ldots, x^{(r)}).$$

<u>LEMMA 2.2</u>. *The* r *largest eigenvalues of a* STP_r *matrix* A *are strictly positive and distinct*

$$\lambda_1 > \lambda_2 > \ldots > \lambda_r > |\lambda_{r+1}|.$$

Furthermore

$$X\begin{pmatrix} i_1 i_2 \ldots i_k \\ 1\ 2\ \ldots\ k \end{pmatrix} > 0 \qquad 1 \leq i_1 < i_2 < \ldots < i_k \leq n; \quad k = 1,\ldots,r$$

and for arbitrary real numbers $c_k, c_{k+1}, \ldots, c_\ell$ $(1 \leq k \leq \ell \leq r,\ \sum_{t=k}^{\ell} c_t^2 > 0)$ *the bounds* S_x^- *and* S_x^+ *of the vector*

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

satisfy

$$k-1 \leq S_x^- \leq S_x^+ \leq \ell - 1.$$

<u>PROOF</u>. 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*

(2.1) $$S_y^- \le r-1 \Rightarrow S_x^- \le S_y^- \quad \text{for all } y \in \mathbb{R}^m,$$

moreover

(2.2) *for all* y *such that* $S_x^- = S_y^- \le r-1$, *the first non-zero component of* x *and* y *have the same sign*

(ii) *If* A *is* STP_r, *then*

(2.3) $$S_y^- \le r-1 \Rightarrow S_x^+ \le S_y^- \quad \text{for all } y \neq 0.$$

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

PROOF. The lemma is a special case of a result in KARLIN (1968), p.233. □

Furthermore we also need the following lemma.

LEMMA 2.4. *Any* TP_r *matrix of rank* $\ge r$ *can be approximated elementwise as closely as desired by means of a* 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 $n \times m$, i.e. P is an $n \times m$ matrix with positive real elements

$$p_{ij} \geq 0 \qquad i = 1,\ldots,n;\ j = 1,\ldots,m$$

such that

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

Denote by

$$r_i = \sum_{j=1}^{m} p_{ij} \qquad i = 1,\ldots,n$$

and

$$c_j = \sum_{i=1}^{n} p_{ij} \qquad j = 1,\ldots,m$$

the row and column sums of P. Let these marginals of P form the diagonal elements of the diagonal matrices R and C, respectively. We assume that R and C are non-singular.

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 $u^{(t)} = (u_1^{(t)},\ldots,u_n^{(t)})^T$, called the *row factors*, and $v^{(t)} = (v_1^{(t)},\ldots,v_m^{(t)})^T$, called the *column factors*, for $t = 1,2,\ldots, \min(m,n)$, which satisfy

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

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

where λ_t is maximal subject to

$$u^{(t)^T}\ R\ u^{(t)} = 1, \qquad v^{(t)^T}\ C\ v^{(t)} = 1$$

$$u^{(t)^T}\ R\ u^{(s)} = 0, \qquad v^{(t)^T}\ C\ v^{(s)} = 0 \qquad s = 1,2,\ldots,t-1.$$

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 the equation

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

that if two rows i and i' have (approximately) equal conditional distributions, the corresponding correspondence analysis scores $u_i^{(t)}$ and $u_{i'}^{(t)}$, $t = 1,2,\ldots,\min(m,n)$, 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 (cf. HILL (1974), KESTER & SCHRIEVER (1982), LEBART et al. (1977), p.54 and SCHRIEVER (1982)) that *Vectors* $u^{(t)}$ *and* $v^{(t)}$ *in Definition* 3.1. *exist for* $t = 1,2,\ldots,\min(m,n)$ *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* $\lambda_1^2 \geq \lambda_2^2 \geq \ldots \geq \lambda_{\min(m,n)}^2$ *are row and column factors of the correspondence analysis solution.*

Furthermore it can be proved that $\lambda_1 = 1$ and that the first row and column factor can always be taken to be trivial, i.e. $u^{(1)} = e$ and $v^{(1)} = e$. In the sequal 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 (cf. FISHER (1940)). This gives us the following interpretation of the row and column factors. The first non-trivial 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 non-trivial, 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 $(\lambda_2 u_i^{(2)}, \lambda_3 u_i^{(3)}, \ldots, \lambda_{q+1} u_i^{(q+1)})$ and column j has coordinates $(\lambda_2 v_j^{(2)}, \lambda_3 v_j^{(3)}, \ldots, \lambda_{q+1} v_j^{(q+1)})$. Thus two row points (column points) lie close to one another if the corresponding conditional distributions are approximately equal. It can be proved (cf. GIFI (1981), p.134, KESTER & SCHRIEVER (1982) and LEBART et al. (1977), p.49) that when $q \geq$ rank $(P)-1$ the converse holds also; in the case that $q <$ 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 forms of dependence between the variables I and J which can be interpreted as order relations; they induce successively stronger orderings over the categories of these two variables. These forms of dependence are called *order dependence of order* r, r = 1,2,3,... .

In the case that the family of induced distributions of $J \mid I = i$ is stochastically (strictly) increasing, i.e.

$$P\{J \leq j_0 \mid I = i\} \text{ is (strictly) decreasing in } i \text{ for each } j_0$$

the variable J induces a (strict) ordering over the categories of the variable I. LEHMANN (1966) speaks of positive regression dependence of J on I. In this case the $n \times m$ frequency table P satisfies

$$(3.1) \qquad 1 \le i < i' \le n \Rightarrow \sum_{j \le j_0} P_{ij}/r_i \ge \sum_{j \le j_0} P_{i'j}/r_{i'} \qquad j_0 = 1,\dots,m-1,$$

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 $n \times n$

$$S_n = \begin{pmatrix} 1 & 1 & \cdot & \cdot & 1 \\ & 1 & \cdot & \cdot & 1 \\ & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot \\ & & & & 1 \end{pmatrix}, \text{ with inverse } S_n^{-1} = \begin{pmatrix} 1 & -1 & & & \\ & 1 & \cdot & & \\ & & \cdot & \cdot & \\ & & & \cdot & -1 \\ & & & & 1 \end{pmatrix}$$

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) element of Q_R equals

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

It follows that

$$(3.1) \iff Q_R \textit{ is} \text{(S)}TP_1.$$

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 M_r, if the vector of differences $\delta = (\delta_1,\dots,\delta_{m-1})^T$, where

$$\delta_j = \psi_j - \psi_{j+1} \qquad j = 1,\dots,m-1,$$

has a number of changes of sign which satisfies

$$S_\delta^- = r-1.$$

The vector ψ is said to be *strictly monotone of order* r, denoted by SM_r, if

$$S_\delta^- = S_\delta^+ = r-1.$$

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

$$\phi_i = \sum_{j=1}^{m} \psi_j \, P_{ij}/r_i \qquad i = 1,\ldots,n,$$

then a stochastically (strictly) increasing family, of distributions of $J \mid I = i$ is charactarized by (see theorem 3.1 below)

$$\psi \textit{ is } (S)M_1 \Rightarrow \phi \textit{ is } (S)M_1,$$

i.e. the family preserves (strictly) order 1 monotonicity of functions.

The ordering of distributions in a stochastically increasing family becomes stronger when this family also preserves order 2 monotonicity of functions. To illustrate this we 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. In the case that this family also preserves order 2 monotonicity of functions, any two-sided test has a power function which is monotone of order ≤ 2, e.g. first decreases than increases. This corresponds to a stronger ordering of the distributions, but still a natural one. In the case that the family of conditional distributions of $J \mid I = i$ preserves order 2 monotonicity of functions, we have

$$\psi \textit{ is } M_r \Rightarrow \phi \textit{ is } M_s, \textit{ where } s \leq r; \quad r = 1,2.$$

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

<u>THEOREM 3.1</u>. *If the matrix* Q_R *is* TP_r, *then*

$$\psi \textit{ is } M_t \;\Rightarrow\; \phi \textit{ is } M_s, \textit{ where } 1 \le s \le t; \quad t = 1,2,\ldots,r.$$

<u>PROOF</u>. Let $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_{n-1})^T$ be the vector of differences

$$\varepsilon_i = \phi_i - \phi_{i+1} \qquad i = 1,\ldots,n-1,$$

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

$$\delta_j = \psi_j - \psi_{j+1} \qquad j = 1,\ldots,m-1.$$

Then we have that

$$\varepsilon = Q_R \delta.$$

The result now follows from lemma 2.3 (i). □

Note that under slight non-degeneracy conditions the converse holds also. In the case that Q_R is STP_r we have

$$\psi \textit{ is } SM_t,\ \phi \textit{ is } M_t \;\Rightarrow\; \phi \textit{ is } SM_t; \qquad t = 1,2,\ldots,r.$$

We have described forms of regression dependence of J on I such that the variable J induces successively stronger orderings over the categories of the variable I. Similarly, the variable I induces successively stronger orderings over the categories of the variable J if the $(m-1)\times(n-1)$ matrix

$$Q_C \text{ is } (S)TP_r \qquad r = 1,2,\ldots,$$

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 both conditions are satisfied we say that the variables I and J are (*strictly*) *order dependent of order* r. The frequency table P then satisfies 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. P *is* $(S)TP_{r+1}$ *and* rank (P) $\geq r+1 \Rightarrow$ 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

$$Q_R \begin{pmatrix} i_1 & i_2 & \dots & i_k \\ j_1 & j_2 & \dots & j_k \end{pmatrix} > 0 \quad 1 \leq i_1 < i_2 < \dots < i_k \leq n-1; \; k = 1,\dots,r$$

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 w.r.t. 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 is induced by order dependence, is reflected in the ordering of the components of the correspondence analysis row and column factors.

THEOREM 3.2. *Let the* $n \times m$ *frequency table* P *be* SDO_r, *then correspondence analysis applied to* P *yields*

(i) *eigenvalues* $1 = \lambda_1 \geq \lambda_2 > \lambda_3 > \ldots > \lambda_{r+1} > \lambda_{r+2}$
and

(ii) *row and column factors* $u^{(t)}$ *and* $v^{(t)}$ *which are* SM_{t-1} *and start oscillating in the same direction,* $t = 2,3,\ldots,r+1$.
Moreover, for arbitrary real numbers $c_k, c_{k+1}, \ldots, c_\ell$ $(2 \leq k \leq \ell \leq r+1,$ $\sum_{t=k}^{\ell} c_t^2 > 0)$ *the vectors*

$$u = \sum_{t=k}^{\ell} c_t u^{(t)} \quad \textit{and} \quad v = \sum_{t=k}^{\ell} c_t v^{(t)} \quad \textit{are } M_s,$$

where

$$k-1 \leq s \leq \ell-1.$$

(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 $i = 1,\ldots,n-1$.

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

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

Note that $u^{(t)}$ is an eigenvector of $R^{-1}PC^{-1}P^T$ corresponding to λ_t^2 iff

$$\bar{x}^{(t)} = S_n^{-1}u^{(t)}$$

is an eigenvector of $\bar{Q}$ corresponding to λ_t^2. Since $R^{-1}PC^{-1}P^T$ has row sums equal to unity, the elements $\bar{q}_{in}$ $i = 1,\ldots,n-1$ vanish. Hence the vector $(0,\ldots,0,1)^T$ is an eigenvector of $\bar{Q}$ corresponding to $\lambda_1^2 = 1$. It follows that $\bar{x}^{(t)} = (\bar{x}_1^{(t)},\ldots,\bar{x}_{n-1}^{(t)},\bar{x}_n^{(t)})^T$ is an eigenvector of $\bar{Q}$ corresponding to λ_t^2 iff $x^{(t)} = (\bar{x}_1^{(t)},\ldots,\bar{x}_{n-1}^{(t)})^T$ is an eigenvector of Q corresponding to λ_t^2, $t = 2,3,\ldots,\min(m,n)$. From

$$\bar{x}^{(t)} = S_n^{-1}u^{(t)}$$

it follows that

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

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

$$\lambda_2^2 > \lambda_3^2 > \ldots > \lambda_{r+1}^2 > \lambda_{r+2}^2$$

and that for arbitrary real numbers $c_k, c_{k+1},\ldots,c_\ell$ $(2 \le k \le \ell \le r+1)$, the vector

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

satisfies

$$k-1 \le S_x^- \le S_x^+ \le \ell-1.$$

Furthermore, in the case that $r \ge 2$ we have

$$X\begin{pmatrix} i & i' \\ 2 & 3 \end{pmatrix} > 0 \qquad 1 \le i < i' \le n-1;$$

i.e.

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

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

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

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 3.3. *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* M_{t-1} *and start oscillating in the same direction;* $t = 2,3,\dots,r+1$. *Moreover for arbitrary real numbers* $c_k, c_{k+1},\dots,c_\ell$ $(2 \le k \le \ell \le r+1, \sum_{t=k}^{\ell} c_t^2 > 0)$ *the vectors*

$$u = \sum_{t=k}^{\ell} c_t u^{(t)} \text{ and } v = \sum_{t=k}^{\ell} c_t v^{(t)} \text{ are } M_s$$

where

$$k-1 \le s \le \ell-1.$$

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 3.2, lemma 2.4 and continuity considerations. □

REMARK. GANTMACHER & KREIN 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 3.2 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 3.2 which are essentially weaker.

As was mentioned in the introduction, the two most important consequences of the theorems 3.2 and 3.3 are

(i) If the two variables I and J are (strictly) order dependent of order ≥ 1, the ordering over the categories of I and J implied by the dependence is reflected in the order of the components of the first non-trivial row and column factor, $u^{(2)}$ and $v^{(2)}$. Furthermore, when the rows and columns of the frequency table P are permuted, the components of the factors $u^{(t)}$ and $v^{(t)}$ undergo the same permutation. Therefore, it follows from the theorems 3.2 and 3.3 that if P is SDO_r, there exists only one permutation of rows and one permutation of columns such that P is DO_r. This permutation is determined by the order of the components of $u^{(2)}$ and $v^{(2)}$. This supports the use of the components of the first non-trivial row and column factors, $u^{(2)}$ and $v^{(2)}$, as scores for the categories of the variables I and J respectively.

(ii) 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 MULTI-CORRESPONDENCE ANALYSIS

4.1. Multi-correspondence 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* $(u^{(t)T},v^{(t)T})^T = (u_1^{(t)},\ldots,u_n^{(t)},v_1^{(t)},\ldots,v_m^{(t)})^T$, *where* $u^{(t)}$ *and* $v^{(t)}$ *are solutions of correspondence analysis applied to* P *corresponding to* λ_t, *are eigenvecotrs 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 $1 + \lambda_t$; $t = 1,2,\ldots,\min(m,n)$.

PROOF. Trivial. □

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

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

$$P_{j\ell} = P_{\ell j}^T \qquad j,\ell = 1,\ldots,k.$$

Denote by C_j the diagonal matrix P_{jj}; $j = 1,\ldots,k$, and denote by C the diagonal matrix of size $m \times m$, where $m = \sum_{j=1}^k m_j$, with diagonal elements the diagonal elements of $C_1,C_2,\ldots,C_k$. Assume that C is non-singular.

DEFINITION 4.1. A solution of multi-correspondence 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

$$(4.1)\quad \begin{cases} \lambda_t v^{(t)} = \begin{pmatrix} C_1^{-1} & & & 0 \\ & C_2^{-1} & & \\ & & \ddots & \\ 0 & & & C_k^{-1} \end{pmatrix} \begin{pmatrix} C_1 & P_{12} & \cdots & P_{1k} \\ P_{12}^T & C_2 & \cdots & P_{2k} \\ \vdots & \vdots & & \vdots \\ P_{1k}^T & P_{2k}^T & \cdots & C_k \end{pmatrix} v^{(t)} \\ \text{where } \lambda_t \text{ is maximal subject to} \\ v^{(t)T}Cv^{(t)} = k;\ v^{(t)T}Cv^{(s)} = 0\quad s = 1,2,\dots,t-1. \end{cases}$$

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 $1/\sqrt{k}$ from the variable factors $v^{(t)}$ and the eigenvalues differ by a factor $1/k$, $t = 1,2,\dots,m$. Therefore, all the eigenvalues λ_t of multi-correspondence analysis are positive and the first variable factor can be taken trivially as

$$v^{(1)} = e$$

with

$$\lambda_1 = k.$$

Furthermore it follows that

$$\lambda_t e^T C_j v^{(j,t)} = e^T C_j (C_j^{-1} P_{j1},\dots,C_j^{-1} P_{jk}) v^{(t)} = e^T C v^{(t)} = 0$$

$$\text{for } j = 1,\dots,k;\ t = 2,3,\dots,m.$$

Thus the non-trivial factors of each variable are also centered.

The interpretation of multi-correspondence analysis variable factors differs somewhat from the interpretation of the row and column factors in correspondence analysis. The first non-trivial variable factors $v^{(j,2)}$, $j = 1,\ldots,k$, can be interpreted as "optimal" scores for the categories of the variables $J_1,\ldots,J_k$: they define derived variables such that the first principal component of their correlation matrix has maximal variance. (cf. HILL (1974))

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

For further results and properties of multi-correspondence analysis we refer to GIFI (1981), KESTER & SCHRIEVER (1982) and LEBART et al. (1977).

4.2. Multi-correspondence analysis and multi-order dependence

In subsection 3.3 it was proved that the first set of r non-trivial row and column factors of correspondence analysis reflects the ordering of categories implied by (strictly) order dependence of order r. In the present subsection we show that a similar property holds for multi-correspondence analysis w.r.t. a (strictly) multi-order dependence of order 1. Moreover, it need not hold w.r.t. multi-order 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 $P_{j\ell}$, $j \neq \ell$; $j,\ell = 1,\ldots,k$, are $(S)DO_r$.

In this case the variables $J_1,J_2,\ldots,J_k$ are said to be *(strictly) multi-order dependent of order* r.

THEOREM 4.2. *Let the* k*-dimensional frequency table* P *be* $SMDO_1$, *then multi-correspondence analysis applied to* P *yields*

(i) *eigenvalues* $k = \lambda_1 \geq \lambda_2 > \lambda_3$
and

(ii) *first non-trivial variable factors* $v^{(j,2)}$, $j = 1,\ldots,k$, *which are all* *strictly increasing or all strictly decreasing in their components. (i.e. they are all* SM_1 *and start oscillating in the same direction)*

PROOF. Let S denote the m × m matrix

$$S = \begin{pmatrix} S_{m_1} & & & 0 \\ & S_{m_2} & & \\ & & \ddots & \\ 0 & & & S_{m_k} \end{pmatrix}, \quad \text{with inverse } S^{-1} = \begin{pmatrix} S^{-1}_{m_1} & & & 0 \\ & S^{-1}_{m_2} & & \\ & & \ddots & \\ 0 & & & S^{-1}_{m_k} \end{pmatrix}$$

Furthermore, let $\bar{Q}$ denote the m × m matrix

$$\bar{Q} = S^{-1}BS$$

and let the matrix Q of size (m-k)×(m-k) be obtained by deleting the rows and columns corresponding to the k indices $m_1, m_1+m_2, \ldots, m_1+m_2+\ldots+m_k$. Similarly to the first part of the proof of theorem 3.2 it follows that (ii) holds iff the eigenvector of Q corresponding to the largest eigenvalue λ_2 is strictly negative or strictly positive.

Since the marginal frequency tables $P_{j\ell}$, $j \neq \ell$; $j,\ell = 1,\ldots,k$ 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* $v^{(j,2)}$, $j = 1,\ldots,k$, *of multi-correspondence analysis applied to* P *which are all increasing or all decreasing in their components (i.e. they are all* 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 $v^{(j,2)}$ reflect the correct ordering of categories, in the case that the variables $J_1,\ldots,J_k$ are - possibly after one permutation of the categories of each variable - (strictly) multi-order dependent of order ≥ 1. This supports the use of multi-correspondence analysis as a one-dimensional scaling technique.

We now briefly explain why these results can not be extended to multi-order dependence of order 2, i.e. the variable factors $v^{(j,3)}$, $j = 1,\ldots,k$, of multi-correspondence analysis applied to a k-dimensional frequency table P *need not* be SM_2 when P is $SMDO_2$. For instance, suppose that $v^{(j,3)}$ is SM_2 for $j = 2,3,\ldots,k$. 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

$$C_1^{-1}P_{1j}v^{(j,3)} \qquad j = 2,3,\ldots,k$$

are SM_2, their sum $v^{(1,3)}$ need not be SM_2. Examples of this can be given. But if the vectors

$$C_1^{-1}P_{1j}v^{(j,3)} \qquad j = 2,3,\ldots,k$$

all attain their maximum (minimum) at the same place, then $v^{(1,3)}$ is actually SM_2.

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 not given but can be found in 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 *discretisation of bivariate density functions*. Let f be a bivariate density function w.r.t. a product measure $\sigma_1 \times \sigma_2$ on $\mathbb{R}^2$. The frequency table P is said to be a discretisation of f if there exist two partitions $\{E_i\}_{i=1}^n$ and $\{F_j\}_{j=1}^m$ of $\mathbb{R}$ such that

$$p_{ij} = \int_{E_i} \int_{F_j} f(x,y)\ d\sigma_2(y) d\sigma_1(x) \text{ and } \sigma_1(E_i) > 0,\ \sigma_2(F_j) > 0$$
$$i = 1,\dots,n;\ j = 1,\dots,m.$$

It turns out that discretisations 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 trinominal, the negative trinominal 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 *log-linear model*

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

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 non-trivial 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 TP_2 and DO_1 property; however he does not prove that this ordering is reflected in the first non-trivial row and column factor of correspondence analysis.

ACKNOWLEDGEMENT. *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 suggestion.*

REFERENCES

[1] BENZÉCRI, J.P. (1973), *L'Analyse de données II: l'Analyse des correspondances*, Dunod, Paris.

[2] FISHER, R.A. (1940), *The precision of discriminant functions*, Ann. Eugen. Lond., 10, p.422-429.

[3] GANTMACHER, F.R. (1977), *Matrix theory,* vol. I, vol II, Chelsea.

[4] GANTMACHER, F.R. & M.G. KREIN (1950), *Oscillation matrices and kernels and small vibrations of mechanical systems*, AEC-translation-4481.

[5] GIFI, A. (1981), *Non-linear multivariate analysis*, Dept. of Data theory, University of Leiden.

[6] GOODMAN, L.A. (1981), *Association models and canonical correlation for contingency tables*, JASA, 76, p.320-334.

[7] HILL, M.O. (1974), *Correspondence analysis: a neglected multivariate method*, Appl. Statist., 23, p.340-354.

[8] HIRSCHFELD, H.O. (1935), *A connection between correlation and contingency*, Proc. Camb. Phil. Soc., 31, p.520-524.

[9] KARLIN, S. (1968), *Total Positivity,* vol. I, Stanford University Press.

[10] KESTER, N.K. & B.F. SCHRIEVER (1982), *Analysis of association of categorical variables by numerical scores and graphical representation,* report in preparation (preprint), Mathematical Centrum, Amsterdam.

[11] LEBART, L., A. MORINEAU & N. TABARD (1977), *Techniques de la description statistique*, Dunod, Paris.

[12] LEHMANN, E.L. (1966), *Some concepts of dependence*, AMS 37, p.1137-1153.

[13] SCHRIEVER, B.F. (1982), *Ordering properties in correspondence analysis*, Mathematisch Centrum report SW80, Amsterdam.