stichting
mathematisch
centrum

AFDELING MATHEMATISCHE STATISTIEK SW 80/82 FEBRUARI
(DEPARTMENT OF MATHEMATICAL STATISTICS)

B.F. SCHRIEVER

ORDERING PROPERTIES IN 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, 62H25
Secondary: 6207

Ordering properties in correspondence analysis*)

by

B.F. Schriever

ABSTRACT

In this paper strong forms of bivariate dependence, which can be interpreted as order relations, are considered. It is proved that, under quite general conditions, such order relations if present in frequency tables are preserved by correspondence analysis. Some models for frequency tables having these strong forms of dependence are given. The results are obtained by using some theory of total positivity.

KEY WORDS & PHRASES: *Correspondence analysis, frequency table, bivariate dependence, order relations, total positivity, scaling, horseshoes*

---

*) This report will be submitted for publication elsewhere.

1. INTRODUCTION

There exist many ways to describe 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*. Theoretical treatments and applications of this technique can also be found in the recent publications of GIFI (1981), HILL (1974) and LEBART et al. (1977).

Correspondence analysis can be regarded as a method of scaling by assigning one-dimensional correspondence analysis scores to the categories of the variables describing the rows and columns of the contingency table. In situations where these variables are nominal variables, the most important aspect of assigning scores to categories is perhaps the ordering of the categories which is implied by these scores. In this paper we introduce a form of dependence between two variables I and J indicating row and column number of the contingency table. We call this dependence *order dependence* because it induces an ordering over the categories of the variables. We then prove that an order dependence between I and J is reflected in the order of the one-dimensional correspondence analysis scores. This supports the use of correspondence analysis as a one-dimensional scaling technique.

Usually correspondence analysis is used as a multidimensional scaling technique and the results are presented in a plot. In this graphical representation of correspondence analysis each row and each column of the contingency table is represented as a point. When both row points and column points lie on a convex or concave curve, we speak of a *horseshoe* in the graphical representation. We prove that a horseshoe occurs when the two variables I and J have a still stronger form of dependence, called *TP_3-dependence*. In fact, we prove a generalization of these results.

In this paper we only consider correspondence analysis as applied to *frequency tables* (i.e. tables of relative frequencies or probabilities). In Section 3 we give many examples of probability models for frequency tables in which the two variables I and J are TP_3-dependent. The abundance of examples demonstrates that TP_3-dependence is quite common in practical models. Although this does not imply the TP_3 character for random samples from such populations, one may nevertheless hope that contingency tables are also often TP_3 or close to it and hence that the order relations of correspondence analysis remain valid. This explains why the typical horseshoe is often found in practice (cf. earlier references).

2. CORRESPONDENCE ANALYSIS, ORDER DEPENDENCE AND TOTALLY POSITIVE DEPENDENCE

2.1. Correspondence analysis

Let P be a frequency table of order n×m, i.e. P is an n×m matrix with non-negative real elements p_{ij} (i = 1,...,n; j = 1,...,m) such that

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

Denote by

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

and

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

the row and column sums of P, respectively. 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.

The identity matrix will be denoted by I and the column vector having all its components equal to unity will be denoted by e; the order of this matrix and vector will be clear from the context. The transpose of a matrix or a vector will be denoted by the superscript ^{T}.

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. The dependence between I and J can be analysed with correspondence analysis. This technique is based on the following definition.

DEFINITION 2.1. 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*, t = 1,2,...,min(m,n), which satisfy

$$(2.1)\qquad \begin{cases} \lambda_t u^{(t)} = R^{-1}Pv^{(t)} \\ \lambda_t v^{(t)} = C^{-1}P^T u^{(t)}, \\ \qquad\qquad \text{where } \lambda_t \text{ is maximal subject to} \\ u^{(t)T}Ru^{(t)} = 1, \quad v^{(t)T}Cv^{(t)} = 1, \\ u^{(t)T}Ru^{(s)} = 0, \quad v^{(t)T}Cv^{(s)} = 0 \qquad (s = 1,2,\dots,t-1). \end{cases}$$

In order to derive properties of the solutions $u^{(t)}$ and $v^{(t)}$, we first prove that a solution of correspondence analysis can be found by solving eigenvalue problems.

LEMMA 2.1. *Vectors* $u^{(t)}$ *and* $v^{(t)}$ *in Definition* 2.1 *exist for* t = 1,2,...,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* $\lambda_1^2 \ge \lambda_2^2 \ge \dots \ge \lambda_{\min(m,n)}^2 \ge 0$ *satisfy* (2.1). *Furthermore,* $u^{(1)} = e$, $v^{(1)} = e$ *and* $\lambda_1 = 1$.

PROOF. It follows from (2.1) that

$$(2.2)\qquad \begin{cases} \lambda_t^2 u^{(t)} = R^{-1}PC^{-1}P^T u^{(t)} \\ \qquad\qquad \text{where } \lambda_t \text{ is maximal subject to} \\ u^{(t)T}Ru^{(t)} = 1, \quad u^{(t)T}Ru^{(s)} = 0 \qquad (s = 1,2,\dots,t-1), \end{cases}$$

i.e. $u^{(t)}$ is an eigenvector of $R^{-1}PC^{-1}P^T$ corresponding to the eigenvalue λ_t^2. Analogously, it follows that $v^{(t)}$ is an eigenvector of $C^{-1}P^TR^{-1}P$ corresponding to the eigenvalue λ_t^2.

Conversely, note that the eigenvalues of $R^{-1}PC^{-1}P^T$ and $R^{-\frac{1}{2}}PC^{-1}P^TR^{-\frac{1}{2}}$ coincide. Since $R^{-\frac{1}{2}}PC^{-1}P^TR^{-\frac{1}{2}}$ is symmetric and positive semi-definite, these eigenvalues are real, non-negative and there exists a system of n orthonormal eigenvectors of this matrix. Premultiplying these eigenvectors by $R^{-\frac{1}{2}}$ will give vectors $u^{(t)}$ $(t = 1,\dots,n)$ which satisfy (2.2). Arrange the eigenvalues of $R^{-1}PC^{-1}P^T$ in decreasing order $\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_n^2$. Suppose that $\lambda_t \neq 0$ and define

$$v^{(t)} := C^{-1}P^Tu^{(t)}/\lambda_t,$$

then

$$R^{-1}Pv^{(t)} = R^{-1}PC^{-1}P^Tu^{(t)}/\lambda_t = \lambda_t u^{(t)}$$

and

$$v^{(t)^T}Cv^{(t)} = u^{(t)^T}Ru^{(t)} = 1, \quad v^{(t)^T}Cv^{(s)} = u^{(t)}Ru^{(s)}\ \lambda_s/\lambda_t = 0,$$

$$(s = 1,\dots,t-1).$$

Moreover, it follows that

$$C^{-1}P^TR^{-1}Pv^{(t)} = \lambda_t^2 v^{(t)}$$

and hence $v^{(t)}$ is an eigenvector of $C^{-1}P^TR^{-1}P$ corresponding to eigenvalue λ_t^2. If $\lambda_t = 0$ for some $t \leq \min(m,n)$, then

$$R^{-1}PC^{-1}P^Tu^{(t)} = 0$$

and hence

$$u^{(t)}RR^{-1}C^{-1}P^Tu^{(t)} = 0$$

and it follows that $C^{-\frac{1}{2}}P^Tu^{(t)} = 0$ and thus $C^{-1}P^Tu^{(t)} = 0$.
Similarly, $C^{-1}P^TR^{-1}Pv^{(t)} = 0$ implies $R^{-1}Pv^{(t)} = 0$. It is seen that the eigenvectors of $R^{-1}PC^{-1}P^T$ and $C^{-1}P^TR^{-1}P$ satisfy (2.1).

A well-known upperbound for an eigenvalue μ of a matrix $A = (a_{ij})$ is

$$|\mu| \leq \max_i \sum_j |a_{ij}|$$

(cf. WILKINSON (1965), p.58). Since the row sums of $R^{-1}PC^{-1}P^T$ are all equal to unity, it follows that $\lambda_1^2 \leq 1$. The vectors $u^{(1)} = e$, $v^{(1)} = e$ satisfy (2.1) with $\lambda_1 = 1$. □

HILL (1974) shows that 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 conditional on the orthogonality of the derived row and column factors for previous values of t.

2.2. Total positivity

In this section we summarize some theory of total positivity. For a matrix $A = (a_{ij})$ of order n×m we denote by

$$A\begin{pmatrix} i_1 i_2 \ldots i_p \\ j_1 j_2 \ldots j_p \end{pmatrix} := \begin{vmatrix} a_{i_1 j_1} & a_{i_1 j_2} & \cdots & a_{i_1 j_p} \\ a_{i_2 j_1} & a_{i_2 j_2} & \cdots & a_{i_2 j_p} \\ \vdots & \vdots & & \vdots \\ a_{i_p j_1} & a_{i_p j_2} & \cdots & a_{i_p j_p} \end{vmatrix}$$

the determinant from the specified elements of A. This determinant is called a *minor* of A of order p if $1 \le i_1 < i_2 < \ldots < i_p \le n$ and $1 \le j_1 < j_2 < \ldots < j_p \le m$.

DEFINITION 2.2. The matrix A is called *totally positive of order* r (abbreviated TP_r) if all minors of order ≤ r are non-negative. If all minors of order ≤ r are positive, then A is said to be *strictly totally positive of order* r (STP_r).

DEFINITION 2.3. A square matrix A is called *oscillatory of order* r (OS_r) if A is TP_r and there exists a positive integer q such that A^q is STP_r.

LEMMA 2.2. *If the matrix* A_1 *of order* $n \times \ell$ *is* TP_r (STP_r) *and the matrix* A_2 *of order* $\ell \times m$ *is* TP_s (STP_s) *then the matrix* $A_1 A_2$ *is* $TP_{\min(r,s)}$ ($STP_{\min(r,s)}$).

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

$$A_1A_2\begin{pmatrix} i_1 i_2 \dots i_p \\ j_1 j_2 \dots j_p \end{pmatrix} = \sum_{1\le k_1<k_2<\dots<k_p\le \ell} A_1\begin{pmatrix} i_1 i_2 \dots i_p \\ k_1 k_2 \dots k_p \end{pmatrix} A_2\begin{pmatrix} k_1 k_2 \dots k_p \\ j_1 j_2 \dots j_p \end{pmatrix}.$$

□

It is easily seen that for any diagonal matrix D with positive diagonal elements, DA is (S)TP_r iff A is (S)TP_r.

LEMMA 2.3. *Let the matrix* A *of order* n×n *be* TP_r, *then* A *is* OS_r *if*

(i) $|i-j| \le 1 \Rightarrow a_{ij} > 0$,

(ii) $A\begin{pmatrix} i & i+1 & \dots & i+p-1 \\ i & i+1 & \dots & i+p-1 \end{pmatrix} > 0 \quad (i = 1,\dots,n-p+1;\ p = 1,\dots,r).$

PROOF. See Appendix. □

REMARK. If the n×n matrix A is TP_n, the conditions (i) and (ii) in Lemma 2.3 are necessary also. Moreover, in this case (ii) can be replaced by

(ii') A *is non-singular*

(cf. GANTMACHER & KREIN (1950), p.139-140, or KARLIN (1968), p.88-93).

An important property of oscillatory matrices is the number of changes of sign of the eigenvectors. In counting the number of changes of sign (of the coordinates) of a vector $u = (u_1,\dots,u_n)^T$, zero coordinates are permitted take on arbitrary signs. So the number of changes of sign of a vector will vary between two bounds S_u^- and S_u^+. In the next lemma the vectors $u^{(1)},\dots,u^{(r)}$ denote the eigenvectors of an n×n matrix A corresponding to the r "largest" eigenvalues $|\lambda_1| \ge |\lambda_2| \ge \dots \ge |\lambda_r|$.

LEMMA 2.4. *The* r *largest eigenvalues of an* OS_r *matrix* A *are positive and distinct*

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

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

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

PROOF. See GANTMACHER & KREIN (1950), p.349. □

2.3. Ordering properties in correspondence analysis

In this section we show that when order relations are present in the frequency table, this ordering is reflected in the ordering of the components of the row and column factors. First we investigate a simple order relation, which we call *doubly ordering*, then we generalize to more complex order relations called TP_r-*ordering*.

DEFINITION 2.4. The n×m frequency table P is called *row ordered* (abbreviated RO) if

$$1 \le i_1 < i_2 \le n \Rightarrow \sum_{j \le j^*} p_{i_1 j}/r_{i_1} \ge \sum_{j \le j^*} p_{i_2 j}/r_{i_2} \qquad (j^* = 1,2,\dots,m-1).$$

If strict inequality holds everywhere, then P is said to be *strictly row ordered* (SRO).
The frequency table P is called *(strictly) column ordered* (abbreviated (S)CO) if P^T is (S)RO.
The frequency table P is called *doubly ordered* (DO) if P is both RO and CO.
P is called SDO if P is both SRO and SCO.

Let *Pr* denote the (empirical) distribution of I and J induced by the frequency table P. Notice that P is RO iff the family of induced distributions of $J|I=i$ is stochastically increasing, i.e.

$$Pr\{J \le j^* \mid I=i\} \text{ is non-increasing in } i \text{ for each } j^*,$$

implying an ordering of the rows of P. This form of dependence between I and J is called *positive regression dependence* of J on I and was considered by LEHMANN (1966). Analogously, P is CO implies an ordering of the columns of P. We shall say that I and J are order dependent when P is DO.

We introduce some more notation. Let S_n be the upper triangular matrix of order n×n:

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

Denote by $\bar{Q}_R := S_n^{-1}R^{-1}PS_m$, $\bar{Q}_C := S_m^{-1}C^{-1}P^TS_n$ and $\bar{Q} := S_n^{-1}R^{-1}PC^{-1}P^TS_n$. The matrices Q_R, Q_C and Q are obtained by deleting the last row and column of the matrices $\bar{Q}_R$, $\bar{Q}_C$ and $\bar{Q}$, respectively. We have the equality $Q = Q_RQ_C$, since the elements $(\bar{Q}_R)_{im} = 0$ $(i = 1,\ldots,n-1)$. With this notation we have

$$P \textit{ is } (S)RO \iff Q_R \textit{ is } (S)TP_1$$

and

$$P \textit{ is } (S)CO \iff Q_C \textit{ is } (S)TP_1.$$

LEMMA 2.5. *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* $R^{-1}PC^{-1}P^T$ *corresponding to* λ_t^2 *satisfies* $u_i^{(t)}-u_{i+1}^{(t)} = x_i^{(t)}$ $(i = 1,\ldots,n-1)$; $t = 2,3,\ldots,n$.

PROOF. $u^{(t)}$ 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, $\bar{Q}_{in} = 0$ $(i = 1,\ldots,n-1)$; hence the vector $(0,0,\ldots,0,1)^T$ is 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. From $\bar{x}^{(t)} = S_n^{-1}u^{(t)}$ it follows that $x_i^{(t)} = \bar{x}_i^{(t)} = u_i^{(t)} - u_{i+1}^{(t)}$ $(i = 1,\ldots,n-1)$. Note that λ_2^2 is the largest eigenvalue of Q. □

THEOREM 2.6. *If* P *is* DO, SCO *and no rows of* P *are proportional, then*

(i) $1 = \lambda_1 > \lambda_2 > \lambda_3$, *unless* P *is a* 2×2 *diagonal matrix in which case* $1 = \lambda_1 = \lambda_2$,

(ii) *the components of the first non-trivial row and column factor* $u^{(2)}$ *and* $v^{(2)}$, *are both strictly increasing or strictly decreasing.*

In the conditions of this theorem the roles of the rows and columns can of course be interchanged. In the proof we use the theorem of Frobenius:

An irreducible TP_1 *matrix always has a positive distinct eigenvalue which is not smaller than the moduli of other eigenvalues. To this maximal eigenvalue there corresponds an eigenvector with positive coordinates.* A proof of the theorem of Frobenius can be found in GANTMACHER (1977), vol. II, p.52-64. An n×n matrix $A = (a_{ij})$ is called *reducible* if the index set $\{1,2,\ldots,n\}$ can be split into two complementary sets $\{i_1,i_2,\ldots,i_{n_1}\}$ and $\{i'_1,i'_2,\ldots,i'_{n_2}\}$, $n_1 + n_2 = n$, such that

$$a_{i_k i'_\ell} = 0 \qquad (k = 1,\ldots,n_1;\ \ell = 1,\ldots,n_2).$$

Otherwise, A is called *irreducible*.

PROOF OF THEOREM 2.6. Q_C is STP_1, Q_R is TP_1 and has no zero row, hence $Q = Q_R Q_C$ is STP_1. Applying the theorem of Frobenius to the matrix Q yields that $\lambda_2^2 > \lambda_3^2$ and $x_i^{(2)} > 0$ $(i = 1,\ldots,n-1)$ or $x_i^{(2)} < 0$ $(i = 1,\ldots,n-1)$. Hence $\lambda_2 > \lambda_3$ and $u^{(2)}$ is strictly increasing. Define

$$y^{(t)} = (y_1^{(t)},\ldots,y_{m-1}^{(t)})^T,$$

where

$$y_i^{(t)} = v_i^{(t)} - v_{i+1}^{(t)} \qquad (i = 1,\ldots,m-1).$$

Then by (2.1) we have $\lambda_t y^{(t)} = Q_C x^{(t)}$. Since Q_C is STP_1, $x^{(2)} < 0$ implies $y^{(2)} < 0$ and hence $v^{(2)}$ is strictly increasing also.

In order to prove that $\lambda_1 > \lambda_2$ we can apply the theorem of Frobenius to the matrix $M := R^{-1}PC^{-1}P^T$, so it is sufficient to prove that M is irreducible. First suppose that $m > 2$. P is SCO implies that $p_{1j} > 0$ $(j = 1,\ldots,m-1)$ and $p_{nj} > 0$ $(j = 2,\ldots,m)$. Since P is also RO it follows that $p_{i1} > 0$ or $p_{i2} > 0$ $(i = 1,\ldots,n)$. Thus the elements $M_{i1} > 0$ and $M_{1i} > 0$ $(i = 1,\ldots,n)$ and hence M is irreducible. In the case that $m = 2$ and $n > 2$ we have that $p_{i1} > 0$ and $p_{i2} > 0$ $(i = 2,\ldots,n-1)$ because P is RO and no rows of P are proportional. Hence only the elements M_{1n} and M_{n1} can be zero and therefore M is irreducible. In the case $m = n = 2$ it is seen that M is reducible iff P is diagonal. □

REMARK. By applying the weaker version of Frobenius' theorem (cf. GANTMACHER (1977), vol. II, p.66-68) we can also prove that

P *is* DO $\Rightarrow u^{(2)}$ *and* $v^{(2)}$ *exist with non-decreasing components.*

Note that under permutations of rows and columns of P the components of $u^{(t)}$ and $v^{(t)}$ undergo the same permutation. Hence it follows from Theorem 2.6 and the above remark that when P satisfies the conditions of Theorem 2.6, there exists only one ordering of rows and columns of P such that P is DO. This result shows that correspondence analysis tries to discover an ordering of the rows and columns of a frequency table; it supports the use of the factors $u^{(2)}$ and $v^{(2)}$ as scores for the categories of the variables I and J, respectively.

Theorem 2.6 can be generalized to stronger forms of dependence than order dependence. In the case that I and J are order dependent, the families of (empirical) conditional distributions of $I|J=j$ and $J|I=i$ are both stochastically increasing. A somewhat stronger form of dependence is obtained when these conditional distributions have monotone likelihood ratio, i.e.

$$j_1 < j_2 \Rightarrow Pr\{I=i|J=j_1\}/Pr\{I=i|J=j_2\} \text{ is non-increasing in } i$$

and

$$i_1 < i_2 \Rightarrow Pr\{J=j|I=i_1\}/Pr\{J=j|I=i_2\} \text{ is non-increasing in } j.$$

Note that these two statements are equivalent and can be written as

$$i_1<i_2, j_1<j_2 \Rightarrow Pr\{I=i_1,J=j_1\}Pr\{I=i_2,J=j_2\} \geq Pr\{I=i_1,J=j_2\}Pr\{I=i_2,J=j_1\},$$

which in turn is equivalent to

$$P \text{ is } TP_2.$$

An even stronger form of dependence is obtained when P is TP_3 or TP_r $(r \geq 3)$. When P is TP_r we shall say that I and J are *totally positive dependent* of order r (TP_r-dependent). LEHMANN (1966) speaks of positive likelihood dependence in the case of TP_2-dependence. Before we generalize Theorem 2.6 to TP_r-dependence, we prove that TP_2-dependence is stronger than order dependence.

THEOREM 2.7. *P is (S)TP$_2$ ⇒ P is (S)DO.*

PROOF. By assumption

$$\begin{matrix} 1\le i_1<i_2\le n \\ \Rightarrow \\ 1\le j_1<j_2\le m \end{matrix} \begin{vmatrix} p_{i_1j_1} & p_{i_1j_2} \\ p_{i_2j_1} & p_{i_2j_2} \end{vmatrix} \ge 0, \text{ i.e. } \begin{matrix} 1\le i_1<i_2\le n \\ \Rightarrow \\ 1\le j_1<j_2\le m \end{matrix} \begin{vmatrix} \sum_{j=1}^{j_1} p_{i_1j} & p_{i_1j_2} \\ \sum_{j=1}^{j_1} p_{i_2j} & p_{i_2j_2} \end{vmatrix} \ge 0$$

and it follows that

$$\begin{matrix} 1\le i_1<i_2\le n \\ \Rightarrow \\ i\le j_1<j_2\le m \end{matrix} \begin{vmatrix} \sum_{j=1}^{j_1} p_{i_1j} & \sum_{j=j_1+1}^{j_2} p_{i_1j} \\ \sum_{j=1}^{j_1} p_{i_2j} & \sum_{j=j_1+1}^{j_2} p_{i_2j} \end{vmatrix} = r_{i_1}r_{i_2} \begin{vmatrix} \sum_{j=1}^{j_1} p_{i_1j}/r_{i_1} & \sum_{j=1}^{j_2} p_{i_1j}/r_{i_1} \\ \sum_{j=1}^{j_1} p_{i_2j}/r_{i_2} & \sum_{j=1}^{j_2} p_{i_2j}/r_{i_2} \end{vmatrix} \ge 0.$$

Choosing $j_2 = m$ yields

$$i_1<i_2 \Rightarrow \sum_{j\le j_1} p_{i_1j}/r_{i_1} \ge \sum_{j\le j_1} p_{i_2j}/r_{i_2} \qquad (j_1 = 1,\dots,m-1).$$

Similarly, it follows that P is CO. In the case that P is STP$_2$ strict inequalities hold. □

REMARK. Generally, it can be proved that

$$P \textit{ is } (S)TP_r \Rightarrow Q_R \textit{ and } Q_C \textit{ are } (S)TP_{r-1}.$$

Note that the converse of Theorem 2.7 does not hold.

THEOREM 2.8. *Let the frequency table P be*

(i) TP$_r$ $(r \ge 2)$,

(ii) *such that every r consecutive rows and every r consecutive columns of P are linearly independent,*

(iii) *not of the blockform* $P = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix}$, *where P_1 and P_2 are not empty.*

Then correspondence analysis applied to P yields

(a) *eigenvalues* $1 = \lambda_1 > \lambda_2 > \dots > \lambda_r > \lambda_{r+1}$, *and*

(b) *row factors $u^{(t)}$ such that for arbitrary real numbers $c_k, c_{k+1},\dots,c_\ell$ $(1\le k\le \ell\le r,\ \sum_{t=k}^{\ell} c_t^2 > 0)$, the number of changes of sign of the vector $u = \sum_{t=k}^{\ell} c_t u^{(t)}$ satisfies $k-1 \le S_u^- \le S_u^+ \le \ell-1$ and column factors $v^{(t)}$ with similar properties.*

PROOF. Consider the matrix $M = R^{-1}PC^{-1}P^T$. By Lemma 2.2 M is TP_r. From the TP_2 character of M and the fact that P is not of blockform and has no zero rows it follows that M is STP_1. Furthermore, condition (ii) and non-singularity of $C^{-\frac{1}{2}}$ imply

$$(PC^{-1}P^T)\begin{pmatrix} i & i+1 & \dots & i+p-1 \\ i & i+1 & \dots & i+p-1 \end{pmatrix} > 0 \qquad (i = 1,\dots,n-p+1;\ p = 1,\dots,r)$$

and hence

$$M\begin{pmatrix} i & i+1 & \dots & i+p-1 \\ i & i+1 & \dots & i+p-1 \end{pmatrix} > 0 \qquad (i = 1,\dots,n-p+1;\ p = 1,\dots,r).$$

Now, by Lemma 2.3 M is OS_r. Application of Lemma 2.4 to matrix M verifies the desired results for the row factors.
The same arguments hold for the matrix $C^{-1}P^TR^{-1}P$. □

REMARK. For a STP_r frequency table P the conditions (i), (ii) and (iii) of Theorem 2.8 are satisfied.

From the result (b) we can derive many properties of the row and column factors. We formulate the most important properties of the row factors in the following corollaries. In formulating these corollaries it is assumed that the conditions (i), (ii) and (iii) of Theorem 2.8 are satisfied for some $r \geq 2$. Furthermore, without loss of generality, it is also assumed that the first non-zero component of each row factor is negative.

COROLLARY 2.9. *The row factor* $u^{(t)}$ *has exactly* $t-1$ *changes of sign* $(t = 1,\dots,r)$.

COROLLARY 2.10. *The components of the first non-trivial row factor* $u^{(2)}$ *are strictly increasing.*

PROOF. Suppose $u^{(2)}$ is not strictly increasing, then since $u^{(1)} = e$ there exists a constant c for which the vector $u = u^{(2)} + cu^{(1)}$ satisfies $S_u^+ \geq 2$. □

Although the conditions of Theorem 2.8 with r = 2 do not quite imply the conditions of Theorem 2.6, we have the same result.

In the next corollary the first and last component are not considered as a maximum or minimum.

<u>COROLLARY 2.11</u>. *When* $r \geq 3$, *the components of* $u^{(3)}$ *have exactly one maximum, no minimum, and equal values of consecutive coordinates can only occur at the maximum.*

<u>PROOF</u>. It follows from Corollary 2.9 that the components of $u^{(3)}$ must have a maximum. Suppose $u^{(3)}$ has a maximum and a minimum, then since $u^{(1)}$ is constant and $u^{(2)}$ is increasing, there exist constants c_1 and c_2 such that the vector $u = u^{(3)} + c_2u^{(2)} + c_1u^{(1)}$ satisfies $S_u^+ \geq 3$. Now suppose that $u^{(3)}$ has consecutive coordinates with equal values (not at the maximum). Then since $u^{(2)}$ is strictly increasing, the vectors $u = u^{(3)} + cu^{(2)}$ would have a maximum and a minimum, for all constants c with a proper choice of sign, which is again impossible. □

In the usual graphical representation of correspondence analysis each row and column of P is represented as a point; row i has coordinates $(\lambda_2u_i^{(2)},\lambda_3u_i^{(3)})$ and column j has coordinates $(\lambda_2v_j^{(2)},\lambda_3v_j^{(3)})$. When both these row points and column points lie on a convex or concave curve, we speak of a *horseshoe* in the graphical representation.

<u>COROLLARY 2.12</u>. *When* $r \geq 2$, *the points in the plot of the first against the second non-trivial row factor lie on a strictly concave curve.*

<u>PROOF</u>. Suppose that this curve is not strictly concave, then

$$(u_{i+1}^{(3)}-u_i^{(3)})/(u_{i+1}^{(2)}-u_i^{(2)}) \text{ is not strictly decreasing in i.}$$

Hence there exist an index i and a constant c such that

$$\frac{u_{i+1}^{(3)} - u_i^{(3)}}{u_{i+1}^{(2)} - u_i^{(2)}} + c \leq 0,\ \frac{u_i^{(3)} - u_{i-1}^{(3)}}{u_i^{(2)} - u_{i-1}^{(2)}} + c \geq 0,\ \frac{u_{i-1}^{(3)} - u_{i-2}^{(3)}}{u_{i-1}^{(2)} - u_{i-2}^{(2)}} + c \leq 0.$$

The vector $u := u^{(3)} + cu^{(2)}$ now satisfies

$$u_{i+1} - u_i = u_{i+1}^{(3)} - u_i^{(3)} + cu_{i+1}^{(2)} - cu_i^{(2)} \leq 0$$

$$u_i - u_{i-1} = u_i^{(3)} - u_{i-1}^{(3)} + cu_i^{(2)} - cu_{i-1}^{(2)} \geq 0$$

$$u_{i-1} - u_{i-2} = u_{i-1}^{(3)} - u_{i-2}^{(3)} + cu_{i-1}^{(2)} - cu_{i-2}^{(2)} \leq 0.$$

Hence, the vector u does not have the property of Corollary 2.11. By the same arguments as in Corollary 2.11 this leads to a contradiction. □

These results show that the TP_3-ordering is reflected in the first two non-trivial correspondence analysis scores. In general, similar results can be derived for TP_r-ordering.

3. SOME PROBABILITY MODELS FOR TP-DEPENDENT FREQUENCY TABLES

3.1. Discretisations of TP functions

A class of probability models for frequency tables is obtained by making a discretisation of bivariate density functions. In this section we extend the TP and the weaker ordering properties DO, RO and CO to density functions. Furthermore, we prove that these properties are preserved by discretisation.

Let f(x,y) be defined on X×Y, where X and Y are subsets of $\mathbb{R}$. Note that when X and Y are both finite sets of discrete values, f can be considered as a matrix. We assume that f is a bivariate density function w.r.t. a product measure $\sigma_1 \times \sigma_2$ on X×Y and that

$$\int_Y f(x,t)d\sigma_2(t) > 0 \quad \text{for all } x \in X$$

and

$$\int_X f(t,y)d\sigma_1(t) > 0 \quad \text{for all } y \in Y.$$

Define the transpose of f as $f^T(x,y) := f(y,x)$.

DEFINITION 3.1. The function f defined on X×Y is called *(strictly) row ordered* (abbreviated (S)RO) if

$$F_x(y) := \int_{\inf(Y)}^{y} f(x,t)d\sigma_2(t) \Big/ \int_Y f(x,t)d\sigma_2(t)$$

is strictly decreasing in $x \in X$, for all $y \in Y$, $y < \sup(Y)$.
The function f defined on X×Y is called *(strictly) column ordered* (abbreviated (S)CO) if f^T is (S)RO.
The function f is called *(strictly) doubly ordered*, (S)DO, if f is (S)RO and (S)CO.

A subset $E \subset X$ is said to be *relatively convex* if

$$\forall x,x_1,x_2 \in X \; (x_1,x_2 \in E,\; x_1 \le x \le x_2 \;\Rightarrow\; x \in E).$$

In the next lemma it is shown that grouping of a relative convex subset in the set X does not affect the RO property of densities.

<u>LEMMA 3.1</u>. *Let* f *be* RO *on* X×Y *and let* E *be a relatively convex subset of* X. *Define for arbitrary* $\xi \in E$ *the set* $\tilde{X} := (X-E) \cup \{\xi\}$. *Then the function*

$$\tilde{f}(x,y) = \begin{cases} f(x,y) & \text{if } x \in X-E,\; y \in Y \\ \int_E f(t,y)d\sigma_1(t) & \text{if } x = \xi,\quad y \in Y \end{cases}$$

defined on $\tilde{X}\times Y$ *is* RO. *If in addition* f *is* SRO *and* $\sigma_1(E) > 0$, $\tilde{f}$ *is even* SRO.

<u>PROOF</u>. Let $x \in X-E$ and $y \in Y$. Then

$$\tilde{F}_\xi(y) - \tilde{F}_x(y) = \Big(\int_Y\int_E f(s,t)d\sigma_1(s)d\sigma_2(t)\Big)^{-1}\Big(\int_{\inf(Y)}^{y}\int_E f(s,t)d\sigma_1(s)d\sigma_2(t)\Big) - F_x(y) =$$

$$\Big(\int_Y\int_E f(s,t)d\sigma_1(s)d\sigma_2(t)\Big)^{-1}\int_E\Big\{\int_{\inf(Y)}^{y} f(s,t)d\sigma_2(t) - F_x(y)\int_Y f(s,t)d\sigma_2(t)\Big\}d\sigma_1(s) =$$

$$\Big(\int_Y\int_E f(s,t)d\sigma_1(s)d\sigma_2(t)\Big)^{-1}\int_E\Big\{F_s(y)-F_x(y)\Big\}\int_Y f(s,t)d\sigma_2(t)d\sigma_1(s) = \begin{cases} \ge 0 \text{ if } x > \xi \\ \le 0 \text{ if } x < \xi \end{cases}.$$

If f is SRO and $\sigma_1(E) > 0$, the inequalities are strict. □

It can easily be verified that grouping in the set Y does not affect the RO property either.

Let $\{E_i\}_{i=1}^{n}$ and $\{F_j\}_{j=1}^{m}$ be finite ordered partitions of X and Y, respectively, i.e.

$$i_1<i_2 \Rightarrow x_1<x_2 \text{ for all } x_1 \in E_{i_1},\ x_2 \in E_{i_2}$$

and

$$j_1<j_2 \Rightarrow y_1<y_2 \text{ for all } y_1 \in F_{j_1},\ y_2 \in F_{j_2}.$$

Note that the subsets E_i $(i = 1,\dots,n)$ and F_j $(j = 1,\dots,m)$ are relatively convex. We shall say that a frequency table P is a *discretisation* of the bivariate density f if there exist ordered finite partitions $\{E_i\}_{i=1}^{n}$ and $\{F_j\}_{j=1}^{m}$ such that

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

THEOREM 3.2. *If* f *is* (S)RO, (S)CO *or* (S)DO, *the discretisation* P *of* f *is* (S)RO, (S)CO *or* (S)DO, *respectively.*

PROOF. The proof follows by repeated application of Lemma 3.1 and similar results. □

We now turn to TP functions.

DEFINITION 3.2. The function f defined on X×Y is called *totally positive* of order r (TP_r) if for p = 1,2,...,r

$$\left.\begin{array}{l} x_1 < x_2 < \dots < x_p,\ x_i \in X\ (i = 1,\dots,p) \\ y_1 < y_2 < \dots < y_p,\ y_j \in Y\ (j = 1,\dots,p) \end{array}\right\} \Rightarrow$$

$$f\begin{pmatrix} x_1 & x_2 & \dots & x_p \\ y_1 & y_2 & \dots & y_p \end{pmatrix} := \begin{vmatrix} f(x_1,y_1) & f(x_1,y_2) & \dots & f(x_1,y_p) \\ f(x_2,y_1) & f(x_2,y_2) & \dots & f(x_2,y_p) \\ \vdots & \vdots & & \vdots \\ f(x_p,y_1) & f(x_p,y_2) & \dots & f(x_p,y_p) \end{vmatrix} \geq 0.$$

If strict inequality holds everywhere, f is said to be STP_r.

Just as in the case of matrices it can be proved that f is (S)TP_2 implies that f is (S)DO.

<u>LEMMA 3.3</u>. *Let* f *be* TP_r *on* X×Y *and let* E *be a relatively convex subset of* X. *The function* $\tilde{f}$ *defined in Lemma* 3.1 *is* TP_s, *where* $s = \min(r, \#\text{ points in } \tilde{X})$. *If in addition* f *is* STP_r *and* $\sigma_1(E) > 0$, $\tilde{f}$ *is even* STP_s.

<u>PROOF</u>. Consider for $x_1 < x_2 < \ldots < x_{i-1} < \xi < x_{i+1} < \ldots < x_p$, $x_i \in X$; $y_1 < y_2 < \ldots < y_p$, $y_i \in Y$, the expansion of the determinant w.r.t. the i-th row

$$\tilde{f}\begin{pmatrix} x_1 \ldots x_{i-1}\xi x_{i+1} \ldots x_p \\ y_1 \ldots\ldots\ldots\ldots\ldots y_p \end{pmatrix} = \begin{vmatrix} f(x_1,y_1) & \ldots & f(x_1,y_p) \\ \vdots & & \vdots \\ \int_E f(t,y_1)d\sigma_1(t) & \ldots & \int_E f(t,y_p)d\sigma_1(t) \\ \vdots & & \vdots \\ f(x_p,y_1) & \ldots & f(x_p,y_p) \end{vmatrix} =$$

$$\sum_{k=1}^{p} \Delta_k \int_E f(t,y_k)d\sigma_1(t) = \int_E \sum_{k=1}^{p} \Delta_k\, f(t,y_k)d\sigma_1(t) =$$

$$\int_E \begin{vmatrix} f(x_1,y_1) & \ldots & f(x_1,y_p) \\ \vdots & & \vdots \\ f(t,y_1) & \ldots & f(t,y_p) \\ \vdots & & \vdots \\ f(x_p,y_1) & \ldots & f(x_p,y_p) \end{vmatrix} d\sigma_1(t) \geq 0,$$

because the determinants on the right-hand side are non-negative for all $t \in E$. Here Δ_k are signed minors. Hence $\tilde{f}$ is TP_s. In the case that f is STP_r and $\sigma_1(E) > 0$ it follows that f is STP_s. □

<u>THEOREM 3.4</u>. *If* f *is* (S)TP_r, *any discretisation of* f *into an* n×m *frequency table* P *is* (S)$TP_{\min(r,m,n)}$.

<u>PROOF</u>. The proof follows by repeated application of Lemma 3.4. □

It follows from this theorem that any discretisation of a STP_k density will satisfy the conditions of Theorem 2.8 for appropriate $r \leq k$. However, some bivariate densities are TP_k but not STP_k. The next theorem shows that for these densities the conditions of Theorem 2.8 may also be satisfied in special cases.

THEOREM 3.5. *Let the triangular density*

$$f(x,y) \begin{cases} > 0 & \text{if } x \geq y, \\ = 0 & \text{if } x < y, \end{cases}$$

defined on $X \times Y$*, where* $X = Y$ *and* $\sigma = \sigma_1 = \sigma_2$*, be* TP_r. *Then the* $n \times m$ *frequency table* P *which is a discretisation of* f *satisfies*

(i) P *is* $TP_{\min(m,n,r)}$;

(ii) *every* $\min(m,n,r)$ *consecutive rows and every* $\min(m,n,r)$ *consecutive columns of* P *are linearly independent*;

(iii) P *is not of blockform.*

PROOF. The result (i) follows from Theorem 3.4. In order to prove (ii) consider the finite ordered partition $\{G_k\}_{k=1}^{\ell}$, which is the intersection of $\{E_i\}_{i=1}^{n}$ and $\{F_j\}_{j=1}^{m}$ deleting elements $E_i \cap F_j$ with σ-measure zero. Discretisation of f with the partition $\{G_k\}_{k=1}^{\ell}$ on both X and Y yields a right-lower-triangular frequency table P^* which is $TP_{\min(r,\ell)}$. The elements $p^*_{i\ell-i+1} > 0$ $(i = 1,\ldots,\ell)$ and hence P^* is non-singular. The frequency table P can be obtained from P^* by grouping consecutive rows and columns. It follows that (ii) must hold. It is trivial that P is not of blockform. □

3.2. Properties and examples of TP functions

With the properties and examples given in this section, the TP character of many bivariate densities can be verified.

Definition 3.2 has two obvious consequences for an $(S)TP_r$ function $f(x,y)$ defined on $X \times Y$:

(i) $h(x)g(y)f(x,y)$ is $(S)TP_r$ on $X \times Y$, for all functions $h(x)$ and $g(y)$ which are non-negative (positive) on X and Y, respectively;

(ii) $f(\phi(s),\psi(t))$ is $(S)TP_r$ on $\phi^{-1}(X) \times \psi^{-1}(Y)$, for all functions ϕ and ψ which are both (strictly) increasing or both (strictly) decreasing on $\phi^{-1}(X)$ and $\psi^{-1}(Y)$, respectively.

Furthermore, we state the following lemmas.

LEMMA 3.6. *If* f *is* (S)TP_r *on* X×Y*,* g *is* (S)TP_s *on* Y×Z *and* σ *is a σ-finite measure on* Y*, then the convolution*

$$h(x,z) = \int_Y f(x,y)g(y,z)d\sigma(y)$$

is (S)$TP_{\min(r,s)}$.

PROOF. See KARLIN (1968), p.17. □

LEMMA 3.7. *If* f *is defined on* X×Y*, where* Y *is an open interval and the derivative*

$$\frac{\partial^{r-1}}{\partial y^{r-1}} f(x,y)$$

exists and is continuous for all $x \in X$*, then*

(i) f *is* TP_r *and* $x_1 < x_2 < \dots < x_k$, $x_i \in X$ $(i = 1,\dots,k) \Rightarrow$

$$f^*\begin{pmatrix} x_1 & x_2 & \dots & x_k \\ y & y & \dots & y \end{pmatrix} := \begin{vmatrix} f(x_1,y) & \frac{\partial}{\partial y} f(x_1,y) & \dots & \frac{\partial^{k-1}}{\partial y^{k-1}} f(x_1,y) \\ \vdots & \vdots & & \vdots \\ f(x_k,y) & \frac{\partial}{\partial y} f(x_k,y) & \dots & \frac{\partial^{k-1}}{\partial y^{k-1}} f(x_k,y) \end{vmatrix} \geq 0$$

$$k = 1,\dots,r;$$

(ii) $f^*\begin{pmatrix} x_1 & x_2 & \dots & x_k \\ y & y & \dots & y \end{pmatrix} > 0$ *for all* $x_1 < \dots < x_k$, $x_i \in X$; $y \in Y$ *and for* $k = 1,\dots,r \Rightarrow f(x,y)$ *is* STP_r *on* X×Y.

PROOF. The assertions (i) and (ii) are particular cases of the results in KARLIN (1968), p.50 and p.52, respectively. □

EXAMPLE 1. The function $f(x,y) = e^{xy}$, $-\infty < x,y < \infty$ is STP_∞.

EXAMPLE 2. The function $f(x,y) = (x+y)^{-\alpha}$, $0 < x, y < \infty$, $\alpha > 0$ is STP_∞. To verify this we consider for p = 1,2,... the determinant

$$f^*\begin{pmatrix} x_1 & x_2 & \dots & x_p \\ y & y & \dots & y \end{pmatrix} = \begin{vmatrix} \frac{1}{(x_1+y)^{\alpha}} & \frac{-\alpha}{(x_1+y)^{\alpha+1}} & \dots & \frac{(-1)^{p-1}\alpha(\alpha+1)\dots(\alpha+p-2)}{(x_1+y)^{\alpha+p-1}} \\ \vdots & \vdots & & \vdots \\ \frac{1}{(x_p+y)^{\alpha}} & \frac{-\alpha}{(x_p+y)^{\alpha+1}} & \dots & \frac{(-1)^{p-1}\alpha(\alpha+1)\dots(\alpha+p-2)}{(x_p+y)^{\alpha+p-1}} \end{vmatrix} =$$

$$\frac{(-1)^{p(p-1)}\alpha^{p-1}(\alpha+1)^{p-2}\dots(\alpha+p-2)}{\prod_{k=1}^{p}(x_k+y)^{\alpha+p-1}} \begin{vmatrix} 1 & (x_1+y) & \dots & (x_1+y)^{p-2} & (x_1+y)^{p-1} \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & (x_p+y) & \dots & (x_p+y)^{p-2} & (x_p+y)^{p-1} \end{vmatrix} > 0$$

for $0 < x_1 < \dots < x_k$, $y > 0$. The last determinant is the Vandermonde determinant. For the case $\alpha = 1$, the result can be found in KARLIN (1968), p.149.

EXAMPLE 3. The function $f(x,y) = \Gamma(x+y+1)$, $0 \le x, y < \infty$ is STP_∞. By definition $\Gamma(x+y+1) = \int_0^\infty e^{x\log(t)}\, e^{y\log(t)}\, e^{-t}dt$. The result follows from Example 1 and Lemma 3.6.

EXAMPLE 4. The function

$$f(x,y) = \begin{cases} (x-y)^m\, e^{-\beta(x-y)} & \text{if } x > y \\ 0 & \text{if } x \le y \end{cases}, \text{ where } -\infty<x,y<\infty,\ \beta\ge 0,\ m \in \mathbb{N},$$

is TP_∞. This result can be found in KARLIN & STUDDEN (1966), p.17.

EXAMPLE 5. The function

$$f(x,y) = \begin{cases} \frac{1}{(x-y)!} & \text{if } x \ge y \\ 0 & \text{if } x < y \end{cases}, \quad x,y = 0,\pm 1,\pm 2,\dots$$

is TP_∞. In KARLIN (1968), p.137, it is proved that $\binom{x}{y}$ is TP_∞; the result of Example 5 follows immediately.

EXAMPLE 6. The function

$$f(x,y) = \begin{cases} \binom{n}{x-y} & \text{if } x \ge y \text{ and } x-y \le n \\ 0 & \text{if } x < y \text{ or } x-y > n \end{cases}, \quad x,y = 0,\pm 1,\pm 2,\dots$$

is TP_∞. This result can be found in KARLIN (1968), p.44.

EXAMPLE 7. Let $\{Q_n(x)\}$, $n = 0,1,2,\ldots$ be an orthogonal polynomial system (where Q_n is of exact degree n) w.r.t. a measure μ on $[a,\infty)$, $a > -\infty$. Assume $Q_n(0) > 0$. The function $f(x,n) = Q_n(-x)$, $a \leq x < \infty$ and $n = 0,1,2,\ldots$ is STP_∞. This result is proved in KARLIN & McGREGOR (1959), p.1115.

EXAMPLE 8. An important class of (S)TP functions are those which have the form $f(x,y) = h(x-y)$, $-\infty < x,y < \infty$. The functions $h(x)$ for which $h(x-y)$ is $(S)TP_r$ are called (strictly) Pólya frequency functions of order r (abbreviated $(S)PF_r$).

3.3. Bivariate densities and total positivity

Let W_1, W_2 and W_3 be independent random variables with distributions from a common family which is closed under convolutions. A bivariate distribution is obtained by considering the joint distribution of W_1+W_3 and W_2+W_3. This method of generating bivariate distributions is called *trivariate reduction*. It follows from Lemma 3.6 that bivariate distributions generated by trivariate reduction are $(S)TP_r$, when the corresponding univariate family consists of $(S)PF_r$ distributions. It is seen from the Examples 1, 6, 5 and 4, respectively, that the univariate normal, binomial, Poisson and gamma distributions, which are closed under convolutions, are PF_∞. Hence the *bivariate normal* (with correlation parameter $\rho > 0$), the *bivariate binomial*, the *bivariate Poisson* (cf. HOLGATE (1964)) and the *bivariate gamma* (cf. CHERIAN (1941)) generated by trivariate reduction, are TP_∞. In fact, an alternative proof shows that these bivariate densities are even STP_∞. We give a sketch of this proof. EAGLESON (1964) proves that for these four bivariate densities a canonical expansion exists

$$f(x,y) = \left\{\sum_{r=0}^{\infty} \rho_r Q_r(x) Q_r(y)\right\} \psi_1(x)\psi_2(y),$$

where $\sum_{r=0}^{\infty} \rho_r^2 < \infty$ and $\{Q_r\}$, $r = 0,1,2,\ldots$ is an orthogonal polynomial system w.r.t. a measure σ. It can now be proved, by using Example 7, that the bivariate binomial, the bivariate Poisson and the bivariate gamma are STP_∞. Example 7 cannot be applied to the bivariate normal distribution, but it can be verified in many other ways that this distribution is STP_∞

when $\rho > 0$. (In the case that $\rho < 0$, the reversed density $f(-x,y)$ is STP_∞.)

It can be proved that other known bivariate densities are $(S)TP_\infty$. To identify various types of bivariate densitites, we give references in which the distributions are derived. It follows from the properties and examples in Section 3.2 that the *negative trinomial*, the *bivariate* F (GHOSH (1955)), the *bivariate Pareto* (MARDIA (1962)) and the *bivariate logistic* distribution (GUMBEL (1961)) are STP_∞. It can also easily be verified that the *bivariate gamma* (McKAY (1934)), the *bivariate beta* (JOHNSON (1960)), the *bivariate hypergeometric* and the *trinomial* distribution are TP_∞ and satisfy the conditions of Theorem 3.5. It should be noted that the latter three distributions show a negative dependence, so that the TP_∞ character is only satisfied for the reversed densities (reversed in one variable). Of course not all bivariate densities are TP_r (for some $r \geq 2$); for instance the bivariate Cauchy distribution with density $f(x,y) = (2\pi)^{-1}(1+x^2+y^2)^{-3/2}$, $-\infty < x,y < \infty$, is not TP_2.

3.4. Some other specific models

In a *latent structure model* it is assumed that there exists one latent variable, L, say. The distributions of the variables I and J of the frequency table conditional on the value of the latent variable are independent. Denoting the conditional densities of I and J given $L = \ell$ by $h_I(i \mid L=\ell)$ and $g_J(j \mid L=\ell)$, respectively, and the density of L w.r.t. a measure σ on $\mathbb{R}$ by $f(\ell)$, the latent structure model for the frequency table P can be written as

$$p_{ij} = \int_{\mathbb{R}} h_I(i|L=\ell)\; g_J(j|L=\ell) f(\ell) d\sigma(\ell) \qquad (i = 1,\ldots,n;\ j = 1,\ldots,m).$$

If the unconditional densities $h(i,\ell)$ and $g(j,\ell)$ are both DO or both TP_r, the frequency table P is also DO or TP_r.

Another model for a frequency table P is the *log-linear model*

$$\log p_{ij} = \mu + a_i + b_j + c_i d_j \qquad (i = 1,\ldots,n;\ j = 1,\ldots,m),$$

where $\sum_i a_i = \sum_j b_j = \sum_i c_i = \sum_j d_j = 0$. Note that P is $STP_{\min(m,n)}$ when

c_i $(i = 1,\ldots,n)$ and d_j $(j = 1,\ldots,m)$ are both strictly increasing or strictly decreasing in their indices. GOODMAN (1981) compares maximum likelihood estimates of c_i $(i = 1,\ldots,n)$ and d_j $(j = 1,\ldots,m)$ in this model with the first non-trivial row and column factor of correspondence analysis. He also discusses the ordering of the rows and columns which is present in this model. Essentially, he makes use of TP_2 and DO as we do; however, his treatment is rather sketchy.

APPENDIX. Proof of Lemma 2.3.

The results in this Appendix are slight extensions of the results in GANTMACHER & KREIN (1950), p.137-141 and KARLIN (1968), p.88-93.

LEMMA A-1. *Let the matrix* A *of order* n×m *be* TP_r. *If the row vectors corresponding to certain indices* $1 = i_1 < i_2 < \ldots < i_{p-1} < i_p = n$ *are linearly dependent, where* $p \le r$, *but the row vectors corresponding to the indices* $1 = i_1 < i_2 < \ldots < i_{p-1}$ *and the row vectors corresponding to* $i_2 < i_3 < \ldots < i_{p-1} < i_p = n$ *are linearly independent, then* A *has rank* p-1.

PROOF. Since the row vectors corresponding to $1 = i_1, i_2, \ldots, i_p = n$ are linearly dependent, it follows that there exist numbers $c_1, c_2, \ldots, c_{p-1}$ such that $a_{i_p k} = \sum_{\ell=1}^{p-1} c_\ell a_{i_\ell k}$ (k = 1,...,n). Because the rows corresponding to $i_2, i_3, \ldots, i_p$ are linearly independent, $c_1 \neq 0$. Now, for p < n there exists an index j such that $i_h < j < i_{h+1}$ (in the case that p = n, the lemma is trivial) and we have

$$\text{(A-1)} \qquad A\begin{pmatrix} i_2 & \cdots & i_h & j & i_{h+1} & \cdots & i_p \\ k_1 & \cdots & \cdots & \cdots & \cdots & \cdots & k_p \end{pmatrix} = (-1)^{p-1} c_1 \, A\begin{pmatrix} i_1 & \cdots & i_h & j & i_{h+1} & \cdots & i_{p-1} \\ k_1 & \cdots & \cdots & \cdots & \cdots & \cdots & k_p \end{pmatrix}$$

and

$$\text{(A-2)} \qquad A\begin{pmatrix} i_2 & \cdots & i_h & i_{h+1} & \cdots & i_p \\ k_1^* & \cdots & \cdots & \cdots & \cdots & k_{p-1}^* \end{pmatrix} = (-1)^{p} c_1 \, A\begin{pmatrix} i_1 & \cdots & i_{p-1} \\ k_1^* & \cdots & k_{p-1}^* \end{pmatrix},$$

where $k_1 < k_2 < \ldots < k_p$ are arbitrarely chosen indices and $k_1^* < \ldots < k_{p-1}^*$ are such that

$$\text{(A-3)} \qquad A\begin{pmatrix} i_1 & \cdots & i_{p-1} \\ k_1^* & \cdots & k_{p-1}^* \end{pmatrix} \neq 0.$$

The minors in the formulas (A-1), (A-2) and (A-3) are all non-negative. It follows from (A-3) and (A-2) that $(-1)^p c_1 > 0$. Hence, by (A-1)

$$A\begin{pmatrix} i_1 & \cdots & i_h & j & i_{h+1} & \cdots & i_{p-1} \\ k_1 & \cdots & \cdots & \cdots & \cdots & \cdots & k_p \end{pmatrix} = 0 \qquad \text{for all } 1 \le k_1 < k_2 < \ldots < k_p \le m.$$

Hence the row corresponding to j is a linear combination of the rows corresponding to $i_1, i_2, \ldots, i_{p-1}$. Since j is an arbitrary index different from

$i_1,\ldots,i_p$, it follows that all the row vectors of A can be written as linear combinations of the row vectors corresponding to $i_1,\ldots,i_{p-1}$. Hence the rank of A is $p-1$. □

COROLLARY A-2. *If the* TP_r *matrix* A *of order* $n\times m$ *has indices* $1 = i_1 < i_2 < \ldots < i_p = n$ *and* $1 = j_1 < j_2 < \ldots < j_p = m$, *where* $p \le r$, *such that*

$$A\begin{pmatrix} i_1 & i_2 & \cdots & i_p \\ j_1 & j_2 & \cdots & j_p \end{pmatrix} = 0 \quad \textit{and} \quad A\begin{pmatrix} i_1 & i_2 & \cdots & i_{p-1} \\ j_1 & j_2 & \cdots & j_{p-1} \end{pmatrix} A\begin{pmatrix} i_2 & i_3 & \cdots & i_p \\ j_2 & j_3 & \cdots & j_p \end{pmatrix} > 0,$$

then A *has rank* $p-1$.

PROOF. Apply Lemma A-1 first to the $n\times p$ submatrix of A formed by the columns corresponding to the indices $j_1,\ldots,j_p$; then apply it to A^T. □

Indices $i_1,\ldots,i_p$ and $j_1,\ldots,j_p$ satisfying

$$1 \le i_1,j_1 < i_2,j_2 < \ldots < i_p,j_p \le n \quad \text{and} \quad |i_k - j_k| \le 1 \quad (k = 1,\ldots,p)$$

are said to be *nearly coincident*. Minors whose indices are nearly coincident are called *quasi-principal minors*.

LEMMA A-3. *All quasi-principal minors of order* $p \le r$ *of an* $n\times n$ TP_r *matrix* A *are positive when* A *satisfies*

(i) $|i-j| \le 1 \Rightarrow a_{ij} > 0$;

(ii) $A\begin{pmatrix} i & i+1 & \ldots & i+p-1 \\ i & i+1 & \ldots & i+p-1 \end{pmatrix} > 0 \quad (i = 1,\ldots,n-p+1;\ p = 1,\ldots,r).$

PROOF. The proof is by induction on p. For $p = 1$ the lemma is trivial. Now suppose that the lemma is true for all quasi-principal minors of order $p-1$, and that there exists a quasi-principal minor of order p which vanishes, i.e. there exist nearly coincident indices $i_1^* < i_2^* < \ldots < i_p^*$ and $j_1^* < j_2^* < \ldots < j_p^*$ such that

$$A\begin{pmatrix} i_1^* & i_2^* & \cdots & i_p^* \\ j_1^* & j_2^* & \cdots & j_p^* \end{pmatrix} = 0.$$

By the induction hypothesis we have that

$$A\begin{pmatrix} i_1^* & i_2^* & \cdots & i_{p-1}^* \\ j_1^* & j_2^* & \cdots & j_{p-1}^* \end{pmatrix} A\begin{pmatrix} i_2^* & \cdots\cdots & i_p^* \\ j_2^* & \cdots\cdots & j_p^* \end{pmatrix} > 0.$$

It follows from Corollary A-2 that the submatrix of A formed by the rows $i_1^*, i_1^*+1, \ldots, i_p^*$ and the columns $j_1^*, j_1^*+1, \ldots, j_p^*$ has rank p-1. Hence

$$A\begin{pmatrix} i_1 & i_2 & \cdots & i_p \\ j_1 & j_2 & \cdots & j_p \end{pmatrix} = 0 \quad \text{for} \quad \begin{matrix} i_1^* \le i_1 < \cdots < i_p \le i_p^* \\ j_1^* \le j_1 < \cdots < j_p \le j_p^* \end{matrix}.$$

In particular, for $i_\ell = j_\ell = k+\ell-1$ ($\ell = 1,\ldots,p$) and suitable $k \ge \min(i_1^*, j_1^*)$ we have

$$A\begin{pmatrix} k & k+1 & \cdots & k+p-1 \\ k & k+1 & \cdots & k+p-1 \end{pmatrix} = 0.$$

This contradicts condition (ii) and hence the lemma is true for all quasi-principal minors of order p. □

LEMMA 2.3. *The* TP_r *matrix* A *of order* n×n *is* OS_r *when* A *satisfies the conditions* (i) *and* (ii) *of Lemma* A-3.

PROOF. It is sufficient to prove that A^{n-1} is STP_r. The proof is exactly that of Theorem 9.3 in KARLIN (1968), p.92-93, replacing TP and STP by TP_r and STP_r, respectively (and the words "Theorem 9.2" by "Lemma A-3"). □

ACKNOWLEDGEMENT. *I am very grateful to Dr. R.D. Gill, Prof.Dr. J. Oosterhoff and Prof.Dr. J. de Leeuw for their suggestions and support of this work.*

REFERENCES

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

[2] CHERIAN, K.C. (1941), *A bivariate correlated gamma-type distribution function*, J. Indian Math. Soc. 5, 133-144.

[3] EAGLESON, G.K. (1964), *Polynomial expansions of bivariate distribution*, A.M.S. 35, 1208-1215.

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

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

[6] GHOSH, M.N. (1955), *Simultaneous tests of linear hypotheses*, Biometrica 42, 441-449.

[7] GIFI, A. (1981), *Non-linear multivariate analysis*, Leiden.

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

[9] GUMBEL, E.J. (1961), *Bivariate logistic distributions*, JASA 55, 698-707.

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

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

[12] HOLGATE, P. (1964), *Estimation for the bivariate Poisson distribution*, Biometrica 51, 241-245.

[13] JOHNSON, N.L. (1960), *An approximation to the multinomial distribution*, Biometrica 47, 93-102.

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

[15] KARLIN, S. & J.L. McGREGOR (1955), *Coincidence properties of birth and death processes*, Pacific J. Math. 9, 1109-1140.

[16] KARLIN, S. & W.J. STUDDEN (1966), *Tchebycheff systems: with application in analysis and statistics*, Wiley.

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

[18] LEHMANN, E.L. (1966), *Some concepts of dependence*, A.M.S. 37, 1137-1153.

[19] McKAY, A.T. (1934), *Sampling from batches*, J.R. Statist. Soc., B, 1, 207-216.

[20] MARDIA, K.V. (1962), *Multivariate Pareto distributions*, A.M.S. 33, 1008-1015.

[21] WILKINSON, J.H. (1965), *The algebraic eigenvalue problem*, Clarendon, Oxford.