CWI Tract 20

Order dependence

B.F. Schriever

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

1980 Mathematics Subject Classification: 62H17, 62H20, 62H99.
ISBN 90 6196 294 3

ACKNOWLEDGEMENTS

The research leading to this monograph was carried out at the Department of Mathematical Statistics of the Mathematical Centre at Amsterdam and at the Department of Mathematics and Computer Science of the Free University in Amsterdam. I appreciated the stimulating working environment created by the members of both departments.

I am very grateful to Dr. R.D. Gill and Prof. Dr. J. Oosterhoff from whom I have received so much advice and constructive criticism. I am indebted to Prof. Dr. J. de Leeuw for his valuable suggestions. I also thank Prof. Dr. P. Groeneboom for his useful comments.

I am especially grateful to Mr. K. Snel for typing the manuscript expertly, to the Centre for Mathematics and Computer Science (CWI) for the opportunity to publish this monograph in their series CWI-Tracts and to all others who have contributed towards its realization.

Bert Schriever

Amsterdam, July 1985.

CONTENTS

CHAPTER 1

INTRODUCTION

1.1. THE SCOPE OF THE STUDY

This monograph considers some aspects of the analysis of ordered contingency tables. The main focus is on bivariate tables. Most literature concerning ordered contingency tables is in the context of log-linear models; see for example FIENBERG (1977, pp. 52-58), GOODMAN (1979, 1981, 1985), HABERMAN (1974 b) and McCULLAGH (1978, 1980). We, however, prefer a non-parametric approach. In analysing the dependence between categorical variables one frequently investigates whether there are meaningful or relevant orderings of the categories in order to faciliate the interpretation of the data. An often used method is correspondence analysis. In fact, our study is motivated by some problems arising in this type of analysis.

Let X and Y be two categorical variables. Consider real valued functions φ_1 and ψ_1, defined on the categories of X and Y respectively, such that the correlation between $\varphi_1(X)$ and $\psi_1(Y)$ is maximal. Such transformed variables are called the first pair of canonical variables and describe the most informative part of the association between X and Y. The second pair of canonical variables, $\varphi_2(X)$ and $\psi_2(Y)$, maximizes the correlation among all transformations which are uncorrelated with $\varphi_1(X)$ and $\psi_1(Y)$. Higher order pairs of canonical variables are similarly defined. The association between X and Y is completely described by all pairs of canonical variables together with their canonical correlations. In fact, this defines the technique correspondence analysis. Clearly, for each $t = 1,2,\ldots$ the values of φ_t and ψ_t at the categories of both variables are invariant under permutations of categories. So this technique makes no assumption on, nor uses any ordering of categories. However, the following phenomena are frequently observed in practice; see for example BENZÉCRI (1973, pp. 44-45, 337-338, 482-484), GIFI (1981, p. 145), GREENACRE (1984, pp. 226-232) or HILL (1974). When the categories of X and Y both have an intuitively meaningful order, then this order is

often reflected in the order of their values assigned by φ_1 and ψ_1; i.e. the functions are both increasing in the intuitive ordering. Moreover, the functions φ_2 and ψ_2 of the second canonical pair are frequently both first decreasing and then increasing in the intuitive ordering, exhibiting a kind of convexity property. Since both technique and contingency tables make no claim on the order in which the rows and columns are presented, this ordering property of correspondence analysis gives rise to the following problem. For which contingency tables (forms of dependence between categorical variables) are φ_1 and ψ_1 increasing and is φ_2 convex with respect to φ_1 and ψ_2 convex with respect to ψ_1 and do generalized convexity properties hold for higher order canonical functions?

A related problem is studied by EAGLESON (1964), EAGLESON and LANCASTER (1967) and LANCASTER (1975, 1980). These authors consider the more general situation where X and Y have arbitrary distributions (not just discrete) and are interested in the problem of determining in which cases the functions φ_t and ψ_t for $t = 1,2,\ldots$ form orthogonal polynomial systems. Orthogonal polynomial systems possess the generalized convexity property. (Orthogonality follows from the fact that the t-th pair of canonical variables is uncorrelated with previous pairs.)

It turns out that the ordering properties can be explained by generalizations of Lehmann's notion of regression dependence, which will be called *order dependence*. These generalizations are closely related to the concept of total positivity. Order dependence induces a meaningful (in a probabilistic sense) ordering over the categories of both variables. Therefore, the ordering property is quite relevant for the practice of correspondence analysis. It not only explains phenomena frequently noted in the literature, it also supports the use of the technique and provides an important tool for the analysis of ordered contingency tables.

The question arises whether tests of independence can be given which are more sensitive to these forms of positive dependence than conventional tests of independence. To this end, some well-known tests for the continuous analogon of the problem are adapted to contingency tables. Using an unified and quite general approach, asymptotic properties of these tests and of some more conventional tests are studied.

It is interesting to know whether positive dependence can be ordered (partially) such that tests sensitive to positive dependence become more powerful under "increasing" positive dependence. Typical for such an

ordering is that negative dependence precedes independence. Not much attention has been paid in the literature to orderings of positive dependence; just a few suggestions to compare bivariate distributions have been published, e.g. CAMBANIS, SIMONS and STOUT (1976), KOWALCZYK and PLESZCZYŃSKA (1977), TCHEN (1976) and WHITT (1982). The orderings considered by Whitt are quite strong and are not very often satisfied; the orderings introduced by the other authors are too weak to be reflected in the power functions of tests.

Let's now review the contents of the monograph in more detail. The next section introduces the notations and conventions used throughout this study.

Chapter 2 is more or less preliminary. It starts with some basic properties of correspondence analysis which are essential for a good understanding of the technique. For the sake of completeness, proofs of these elementary results are given. Several conventional tests of independence are based on the canonical correlations produced by correspondence analysis. Asymptotic properties of these tests are derived under the general setup of conditional Poisson sampling rather than the more usual multinomial sampling. Since such generalizations of well-known results for the multinomial case are frequently mentioned and applied without proof, a complete treatment is given in Sections 2.2 and 2.3. Furthermore, the tests are compared by means of approximate Bahadur and Pitman efficiencies. The results of this chapter are used later to derive asymptotic properties of tests sensitive to positive dependence.

In Chapter 3 the notion of order dependence is introduced. It is actually equivalent, up to permutations of categories of both variables, to generalizations of the notion of regression dependence. These generalizations are related to those studied by VAN ZWET (1968). The theory of total positivity is basic to generalized regression dependence and some results of this theory are summarized in Section 3.1. The oscillation diminishing property given in Theorem 3.2.3 (Section 3.2) is typical for generalized regression dependence. Section 3.4 illustrates that order dependence and generalized regression dependence arise quite naturally in practical models. The main result of this chapter, the ordering property of correspondence analysis, is formulated in Section 3.3. This result actually first appeared in SCHRIEVER (1983). Furthermore, tests of independence which are sensitive to ordered alternatives are considered in the last section of Chapter 3. The

tests of YATES (1948) and adaptations of tests based on Kendall's tau and Spearman's rho turn out to perform well in such situations. Their approximate Bahadur efficiencies with respect to some conventional tests are derived. Tests based on canonical variables turn out to have a rather bad performance.

Chapter 4 discusses several suggestions for a partial ordering for positive dependent distributions. The continuous case as well as the discrete case is considered. Such a partial ordering, denoted in this introductory chapter by $\dot{\geq}$, is in fact only useful when it has the following properties. Suppose F and F' are two bivariate distributions such that $F' \dot{\geq} F$, then any reasonable measure of positive dependence should attain a larger value at F' than at F. Moreover, for finite samples drawn from F' and F, one wants these measures computed at samples from F' to be stochastically larger than at samples from F. This last property implies that tests based on such measures have power functions which are monotone in the ordering. The main focus in Chapter 4 is on an ordering called "more associated" which actually has these properties in case F and F' are continuous. This ordering is related to the notion of dependence called "associated" by ESARY, PROSCHAN and WALKUP (1967); special cases of the ordering are related to regression dependence. In case tied observations can occur in samples from F and F', the order preserving properties only hold in very few and simple cases. But it is shown in Section 4.4 that the results carry over to the contingency table case when the definitions of the partial orderings and measures of dependence are modified appropriately.

The last chapter discusses generalizations of the ordering properties of correspondence analysis to the multivariate case (Section 5.2). But first the most common generalization of correspondence analysis, called multiple correspondence analysis, is briefly described in Section 5.1. (A complete system of generalizations is given in GIFI (1981); see also DE LEEUW (1984 b).) Partial orderings similar to those introduced in Chapter 4 could easily be formulated for the multivariate case, but it is somewhat premature to give partial orderings since the corresponding forms and measures of multivariate positive dependence have not yet been sufficiently explored. Chapter 5 closes with an example of multiple correspondence analysis in ordered latent structure models. Such models are frequently used in psychological measurement theory to analyse ability (or attitude) tests, e.g. intelligence tests. MOKKEN (1971) introduced models with orderings related to order dependence and total positivity. It turns out that, under

realistic assumptions, multi correspondence analysis orders the questions (items) in the tests according to their difficulty and orders individuals according to their ability (or attitude).

1.2. GENERAL NOTATIONS AND CONVENTIONS

The following notation will be used in the sequel. Matrices are denoted by capital letters. The (i,j)-th 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 symbol $\otimes$ is used for the Kronecker matrix product (cf. RAO, 1973, p. 29 or MUIRHEAD, 1982, pp. 73-76). Furthermore, for a matrix A, *vec* [A] denotes the vector with elements a_{ij} arranged lexicographically by its indices and *diag* [A] denotes the diagonal matrix with diagonal elements a_{ij}, also arranged in lexicographical order. The identity matrix is denoted by I and the vector having all its components equal to unity by e; the size of this matrix and vector will be clear from the context, but if ambiguity arises the size is given as an index, e.g. I_n or e_m. The square matrix S denotes the upper triangular matrix with unities on and above the diagonal. Its inverse S^{-1} is the matrix with unit elements on the (main) diagonal, with elements -1 on the first super diagonal and with all other elements zero, thus

$$S = \begin{pmatrix} 1 & 1 & \dots & 1 \\ & 1 & \dots & 1 \\ & & \ddots & \vdots \\ & & 1 & 1 \\ & & & 1 \end{pmatrix}, \qquad S^{-1} = \begin{pmatrix} 1 & -1 & & \\ & 1 & -1 & \\ & & \ddots & \ddots \\ & & & 1 & -1 \\ & & & & 1 \end{pmatrix}.$$

The sizes of these square matrices will also be clear from the context or are given as indices.

A real number x is said to be positive when $x \geq 0$ and is called strictly positive when $x > 0$. Similarly a $n \times n$ matrix A is said to be positive definite when $x^TAx \geq 0$ for all n-dimensional vectors x and is said to be strictly positive definite when $x^TAx > 0$ for all non-zero vectors x. A real valued function f defined on $X \subset \mathbb{R}$ is called increasing on X when $x_1 < x_2$ $(x_1, x_2 \in X) \Rightarrow f(x_1) \leq f(x_2)$ and is called strictly increasing on X

if strict inequality is implied. The notions (strictly) negative, (strictly) negative definite and (strictly) decreasing are similarly defined. For subsets A_1 and A_2 of $\mathbb{R}$, the notation $A_1 < A_2$ is used for $x_1 \in A_1, x_2 \in A_2 \Rightarrow x_1 < x_2$.

The expectation and variance of a random variable Z are denoted by EZ and $Var(Z)$. Similarly, Cov and $Corr$ stand for covariance and correlation of two random variables. The symbol $\sim$ is short for "is distributed as". Convergence in distribution of random variables is denoted by $\to_D$ and convergence in probability by $\to_P$. The n-dimensional normal distribution is abbreviated by N_n.

The bivariate distribution function of the pair of random variables (X,Y) is usually denoted by F. The variable corresponding to the first argument of a bivariate distribution is called the row variable and the second one the column variable. The marginal distributions of the row and column variable of F are denoted by F_r and F_c respectively. A version of the conditional distribution function of the row variable given the column variable is denoted by $F_{r|c}$. A density of F with respect to some product measure $\sigma_r \times \sigma_c$ is usually denoted by p. In case X and Y have discrete distributions (contingency table case), the matrix P contains the cell probabilities. Such a matrix is called a *probability table*. Occasionally, the same notation F or P is used for empirical as well as underlying distributions. In case empirical versions are explicitly meant, the upper script $\hat{}$ is added usually in combination with the sample size (N) as an upper index. So an empirical probability table has relative frequencies as elements. The term *contingency table* is used for the table with (integer valued) frequencies, which add to the sample size.

CHAPTER 2

CORRESPONDENCE ANALYSIS

2.1. BASIC PROPERTIES

Correspondence analysis (abbreviated CA) is the now well-established name of a technique for analysing the association which is present in a contingency table. CA is actually mathematically equivalent to a number of other techniques independently (re-) introduced by several authors, e.g. *reciprocal averaging* (RICHARDSON and KUDER, 1933 and HORST, 1935), *simultaneous linear regression* (HIRSCHFELD, 1935), *Fisher's contingency table analysis* (FISHER, 1940 and MAUNG, 1941), *optimal scaling* (GUTTMAN, 1941), *principal component analysis of qualitative data* (BURT, 1950), *gradient analysis* (WHITTAKER, 1967), *biplot* (GABRIEL, 1971), *dual scaling* (NISHISATO, 1980) and of course *analyse factorielle des correspondances* (BENZÉCRI, 1973). The origin of the principle behind these techniques is hard to trace, because it can be described from several points of view. Fisher and Guttman are often regarded as the inventors, but earlier references are those of Richardson and Kuder, Horst and Hirschfeld. (Hirschfeld later became better known under the name Hartley.) The history of CA and related techniques is described in BENZÉCRI (1982), DE LEEUW (1973,1983) and NISHISATO (1980).

Fisher's and Maung's approaches to CA are perhaps the most appealing from a statistical point of view. In 1940 Fisher considered a contingency table of the hair and eye colour of Scottish school children in the context of discriminant analysis. In his analysis he assigned scores to the categories of the nominal variable eye colour in such a way that the distributions given the hair colour classes were most distinct. Fisher stated that such scores may be found in a variety of ways. He didn't notice that his method to find these scores was actually the core of the reciprocal averaging algorithm described by RICHARDSON and KUDER (1933) and HORST (1935). Furthermore, Fisher noted that in his example each derived (scored) variable has a linear regression on the other. But seeking derived variables

with simultaneous linear regressions was in fact the problem considered by HIRSCHFELD (1935). In 1941 Maung worked out Fisher's contingency table analysis and showed that the derived variables maximize the product moment correlation.

Our approach to CA is also based on maximizing correlations. This section summarizes some basic properties of the CA solution. Furthermore, some attention is paid to the graphical display of the CA solution. Such a graphical representation of contingency tables is quite common in practice and Benzécri's interpretation is briefly explained. This section closes with some remarks on a generalization of CA to continuous variables. Such a generalization is, however, of limited practical value.

This section is concerned with population properties of CA only and does not deal with the problem of how the solution is affected by sampling variation. Some asymptotic sampling results are given in the next section.

Let X and Y be two nominal variables with n and m categories respectively. It is assumed throughout this chapter that $n \leq m$. For notational convenience the categories of X and Y are labeled by $1,2,\ldots,n$ and $1,2,\ldots,m$ respectively; but it should be emphasized that these are just labels, neither their value nor their ordering need be (and generally is not) meaningful for the corresponding categories. Furthermore, let the joint distribution of the two (nominal) variables be given by the $n \times m$ probability table P with elements

$$p_{ij} = P\{X = i,\ Y = j\} \qquad \text{for } i = 1,\ldots,n;\ \ j = 1,\ldots,m.$$

Thus all elements of P are positive and add to unity. The marginal distributions of the row variable X and the column variable Y are given by the vectors $r = (r_1,\ldots,r_n)^T$ and $c = (c_1,\ldots,c_m)^T$ respectively, where

$$r_i = \sum_{j=1}^{m} p_{ij} = P\{X = i\} \quad \text{and} \quad c_j = \sum_{i=1}^{n} p_{ij} = P\{Y = j\} \qquad \begin{matrix} i = 1,\ldots,n; \\ j = 1,\ldots,m. \end{matrix}$$

Diagonal matrices R and C are defined by $R = diag\,[r_1,\ldots,r_n]$ and $C = diag\,[c_1,\ldots,c_m]$ and are assumed to be non-singular.

Correspondence analysis seeks two real valued functions φ_1 and ψ_1, defined on the categories of X and Y respectively, such that the correlation of the derived (numerical) variables $\varphi_1(X)$ and $\psi_1(Y)$ is maximal. Clearly, it is no restriction to assume that the derived variables have mean zero

and unit variance. Such variables $\varphi_1(X)$ and $\psi_1(Y)$ are called the (first pair of) *canonical variables* and their correlation is called the *canonical correlation*. This first pair of canonical variables describes the most informative part of the dependence between X and Y. The remaining part of the dependence can be analysed in a similar way and so CA seeks a second pair of canonical variables $\varphi_2(X)$ and $\psi_2(Y)$ which has maximum correlation but is uncorrelated with the first pair. This procedure is continued with a third pair, a fourth pair etc. until no new pair which is uncorrelated with the previous pairs can be found.

DEFINITION 2.1.1. The t-*th pair of canonical variables* of the distribution of (X,Y) is a pair of transformed variables $\varphi_t(X)$ and $\psi_t(Y)$ for which

$$\lambda_t = Corr\,(\varphi_t(X),\psi_t(Y)) \tag{2.1.1}$$

is maximal subject to

$$\begin{aligned} &E\varphi_t(X) = 0, && E\psi_t(Y) = 0, \\ &Var\,(\varphi_t(X)) = 1, && Var\,(\psi_t(Y)) = 1, \\ &Corr\,(\varphi_t(X),\varphi_s(X)) = 0, && Corr\,(\psi_t(Y),\psi_s(Y)) = 0 \quad \text{for } s = 1,\ldots,t-1. \end{aligned} \tag{2.1.2}$$

The correlation λ_t is called the t-*th canonical correlation* and the functions φ_t and ψ_t are called the t-*th canonical (row and column) functions*. The *CA solution* consists of all triplets $(\lambda_t,\varphi_t,\psi_t)$ for t = 1,2,... .

The present definition of CA is equivalent to definitions based on the reciprocal averaging algorithm or based on simultaneous linear regression, as is shown by the following result of HIRSCHFELD (1935).

PROPOSITION 2.1.1. *The transformed variables* $\varphi_t(X)$ *and* $\psi_t(Y)$ *are a* t-*th pair of canonical variables iff they satisfy*

$$\begin{aligned} \lambda_t\varphi_t(X) &= E(\psi_t(Y)\,|\,X) && a.s. \\ \lambda_t\psi_t(Y) &= E(\varphi_t(X)\,|\,Y) && a.s. \end{aligned} \tag{2.1.3}$$

with λ_t *maximal subject to* (2.1.2).

PROOF. First suppose that $\varphi_t(X)$ and $\psi_t(Y)$ satisfy (2.1.3) with λ_t maximal subject to (2.1.2). Thus, since both transformed variables are normalized

with mean zero and unit variance, $\lambda_t = \lambda_t E\varphi_t(X)\varphi_t(X) = E\varphi_t(X)E(\psi_t(Y)|X) = Corr\,(\varphi_t(X),\psi_t(Y))$ is maximal subject to (2.1.2).

Conversely, suppose that $\varphi_t(X)$ and $\psi_t(Y)$ is a pair maximizing (2.1.1) subject to (2.1.2). Let $g(Y)$ and μ be such that $Var\,(g(Y)) = 1$ and $\mu g(Y) = E(\varphi_t(X)|Y)$ with probability one. Clearly, $E\varphi_t(X) = 0$ implies $Eg(Y) = 0$. Since all transformed variables are normalized, it follows by Cauchy-Schwarz that $\lambda_t = Corr\,(\varphi_t(X),\psi_t(Y)) = E\varphi_t(X)\psi_t(Y) = E\psi_t(Y)E(\varphi_t(X)|Y) = \mu E\psi_t(Y)g(Y) \leq \mu(E\psi_t^2(Y)Eg^2(Y))^{\frac{1}{2}} = \mu$ with equality only if $\psi_t(Y) = g(Y)$ with probability one. Thus for each $\varphi_t(X)$, $\psi_t(Y)$ maximizes $Corr\,(\varphi_t(X),\psi_t(Y))$ only if $\lambda_t\psi_t(Y) = E(\varphi_t(X)|Y)$ a.s.. So it remains to show that $E(\varphi_t(X)|Y)$ is uncorrelated with $\psi_s(Y)$ for $s = 1,\ldots,t-1$. This can be done by induction over t. For $t = 1$ there is no such condition. Now suppose that $\lambda_s\varphi_s(X) = E(\psi_s(Y)|X)$ a.s. for $s = 1,\ldots,t-1$. Then $Cov(\psi_s(Y),E(\varphi_t(X)|Y)) = E\psi_s(Y)\varphi_t(X) = E\varphi_t(X)E(\psi_s(Y)|X) = \lambda_s E\varphi_t(X)\varphi_s(X) = 0$. Similarly, it follows that $\lambda_t\varphi_t(X) = E(\psi_t(Y)|X)$ a.s.. □

COROLLARY 2.1.2. *Corr* $(\varphi_t(X),\psi_s(Y)) = 0$ *for* $s \neq t$.

Fisher's approach, in the context of discriminant analysis, was to find the transformation of the row variable which yields greatest variance between categories of the column variable relative to its variance (FISHER, 1936). That is to find $\varphi_t(X)$ which maximizes $Var\,(E(\varphi_t(X)|Y))$ subject to $Var\,(\varphi_t(X)) = 1$ and $Corr\,(\varphi_s(X),\varphi_t(X)) = 0$ for $s = 1,\ldots,t-1$. Let λ_t and $\psi_t(Y)$ be such that $\lambda_t\psi_t(Y) = E(\varphi_t(X)|Y)$ a.s., then it is easily verified that Fisher's approach produces another equivalent formulation of CA.

The functions φ_t and ψ_t are only defined on the categories of X and Y. Therefore, only the values $\varphi_t(i)$ for $i = 1,\ldots,n$ and $\psi_t(j)$ for $j = 1,\ldots,m$, called *category scores* of the t-th canonical pair, are of interest for $t = 1,2,\ldots$. In the sequel the functions φ_t and ψ_t are identified with vectors $\varphi_t = (\varphi_{1t},\ldots,\varphi_{nt})^T$ and $\psi_t = (\psi_{1t},\ldots,\psi_{mt})^T$, where $\varphi_{it} = \varphi_t(i)$ for $i = 1,\ldots,n$ and $\psi_{jt} = \psi_t(j)$ for $j = 1,\ldots,m$ and $t = 1,2,\ldots$. Although φ_t and ψ_t are vectors, they still will frequently be called canonical functions. Thus by Proposition 2.1.1, the vectors of category scores satisfy

$$\begin{aligned} \lambda_t\varphi_t &= R^{-1}P\psi_t \\ \lambda_t\psi_t &= C^{-1}P^T\varphi_t \end{aligned} \tag{2.1.4}$$

where λ_t is maximal subject to

$$
\begin{aligned}
& r^T\varphi_t = 0, \qquad c^T\psi_t = 0, \\
& \varphi_t^T R\varphi_t = 1, \qquad \psi_t^T C\psi_t = 1, \\
& \varphi_t^T R\varphi_s = 0, \qquad \psi_t^T C\psi_s = 0, \qquad \text{for } s = 1,\ldots,t-1.
\end{aligned} \tag{2.1.5}
$$

These equations show an important property of CA: *if two rows of* P *are proportional then the CA category scores for these rows are equal and, moreover, the category scores for other rows and columns remain unaltered when such rows are pooled. A similar property holds for proportional columns.*

HIRSCHFELD (1935) showed that the vectors with CA row and column category scores can be obtained by solving a matrix eigenvalue problem. Recall that a set of n-dimensional vectors is said to be complete when they are linearly independent and span $\mathbb{R}^n$.

THEOREM 2.1.3. *The vectors* φ_t *and* ψ_t *satisfying* (2.1.4), *where* λ_t *is maximal subject to* (2.1.5), *exist for* t = 1,2,...,n-1. *Moreover, the vectors* $e,\varphi_1,\ldots,\varphi_{n-1}$ *can be any complete set of eigenvectors of the matrix* $R^{-1}PC^{-1}P^T$ *corresponding to the eigenvalues* $1 \geq \lambda_1^2 \geq \ldots \geq \lambda_{n-1}^2$ *and the vectors* $e,\psi_1,\ldots,\psi_{n-1}$ *is any set of eigenvectors of* $C^{-1}P^TR^{-1}P$ *corresponding to the same eigenvalues.*

PROOF. It is easily seen that u is an eigenvector of $R^{-1}PC^{-1}P^T$ corresponding to an eigenvalue μ iff $R^{\frac{1}{2}}u$ is an eigenvector of $R^{-\frac{1}{2}}PC^{-1}P^TR^{-\frac{1}{2}}$ corresponding to the eigenvalue μ. Since the latter matrix is symmetric and positive definite, a complete set of eigenvectors $u_1,u_2,\ldots,u_n$, normalized such that $u_t^TRu_t = 1$ and $u_t^TRu_s = 0$ for $t \neq s$, corresponding to real eigenvalues $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n \geq 0$ of the matrix $R^{-1}PC^{-1}P^T$ exist. Furthermore, this matrix has positive elements and its row sums are unity (Markov or stochastic matrix) and hence by GANTMACHER (1977, vol II, p. 83) its largest eigenvalue $\mu_1 = 1$ and the corresponding eigenvector can be taken $u_1 = e$. Furthermore, φ_t and ψ_t satisfy (2.1.4) with λ_t maximal subject to (2.1.5) iff

$$
\begin{aligned}
\lambda_t^2\varphi_t &= R^{-1}PC^{-1}P^T\varphi_t \\
\lambda_t\psi_t &= C^{-1}P^T\varphi_t
\end{aligned}
$$

with λ_t maximal subject to (2.1.5). Hence φ_t can be taken u_{t+1} and $\lambda_t^2 = \mu_{t+1}$ for t = 1,...,n-1. □

Note that the vectors φ_t and ψ_t are uniquely determined, up to a change of sign, when the corresponding eigenvalue λ_t^2 is simple.

Fisher pointed out (cf. MAUNG, 1941) that the following *canonical decomposition* (which shows similarity to a singular value decomposition) of a probability table P always exists. It provides more insight in the way category scores describe the dependence structure between the nominal variables X and Y.

PROPOSITION 2.1.4. *Let* $\varphi_1,\ldots,\varphi_{n-1}$ *and* $\psi_1,\ldots,\psi_{n-1}$ *be vectors satisfying* (2.1.5). *Then* (2.1.4) *holds iff*

$$p_{ij} = r_i c_j (1 + \sum_{t=1}^{n-1} \lambda_t \varphi_{it} \psi_{jt}) \qquad \text{for } i = 1,\ldots,n;\ j = 1,\ldots,m. \tag{2.1.6}$$

PROOF. Let Φ denote the $n \times n$ matrix with $e,\varphi_1,\ldots,\varphi_{n-1}$ as columns and Ψ the $m \times n$ matrix with $e,\psi_1,\ldots,\psi_{n-1}$ as columns. Then (2.1.5) for $t = 1,\ldots,n-1$ is equivalent to

$$\Phi^T R \Phi = I \quad \text{and} \quad \Psi^T C \Psi = I.$$

Furthermore, write $\Lambda = \mathit{diag}\,[1,\lambda_1,\ldots,\lambda_{n-1}]$. Note that since the columns of Φ span $\mathbb{R}^n$, $\Phi\Phi^T = R^{-1}$. Therefore, (2.1.4) for $t = 1,\ldots,n-1$ implies

$$\Psi\Lambda\Phi^T = C^{-1}P^T\Phi\Phi^T = C^{-1}P^TR^{-1}$$

and hence $P = R\Phi\Lambda\Psi^TC$. The converse is obvious. □

The canonical correlations $\lambda_1,\ldots,\lambda_{n-1}$ measure the strength of the dependence in the probability table. Consequently, tests of independence can be based on empirical canonical correlations (cf. Section 2.3). For example, it is easily seen from (2.1.6) and (2.1.5) that Pearson's χ^2 equals $\Sigma_1^{n-1} \lambda_t^2$. The category scores (canonical functions) on the other hand, give more insight into the structure or form of dependence. Structure and strength are, however, not completely separated by these two concepts. The canonical variables can be used in further analysis as numerical versions of the row and column variables, in particular when these nominal variables possess an ordinal relation (cf. Chapter 3).

The following theorem shows how an intuitive concept of "stronger" dependence is reflected in the canonical correlations. Consider two $n \times m$ probability tables P and P' such that $P = N^TP'M$, where N and M are square matrices with all elements positive and row sums unity (Markov matrices).

The existence of such matrices means that the distribution P of (X,Y) can be derived from the distribution P' of (X',Y') by independent randomizations of the row and column variable: there exist independent random variables U and V both distributed on [0,1], say, and independent of (X',Y') and there exist functions $\kappa_1 : \{1,\ldots,n\} \times [0,1] \to \{1,\ldots,n\}$ and $\kappa_2 : \{1,\ldots,m\} \times [0,1] \to \{1,\ldots,m\}$ such that

$$(X,Y) \sim (\kappa_1(X',U),\kappa_2(Y',V)).$$

Intuitively, P' is more strongly dependent than P. The following result holds for their canonical correlations $\lambda_1' \geq \ldots \geq \lambda_{n-1}'$ and $\lambda_1 \geq \ldots \geq \lambda_{n-1}$.

THEOREM 2.1.5. *Suppose that* $P = N^T P' M$, *where* N *and* M *are square Markov matrices, then*

$$(2.1.7) \qquad \sum_{s=1}^{t} \lambda_s \leq \sum_{s=1}^{t} \lambda_s' \qquad \text{for } t = 1,\ldots,n-1.$$

PROOF. First make P' square by replacing the last row by m-n+1 copies of $(m-n+1)^{-1}$ times this last row and make N of size $m \times m$ by applying the same operation to both the last row and the last column. This does not affect the canonical correlations of P and P'. Thus it is no loss of generality to assume that P and P' are square.

By Theorem 2.1.3, $1 \geq \lambda_1 \geq \ldots \geq \lambda_{n-1}$ are the square roots of the eigenvalues of the matrix $R^{-1}PC^{-1}P^T$ and hence of $R^{-\frac{1}{2}}PC^{-1}P^TR^{-\frac{1}{2}}$. Therefore, $1 \geq \lambda_1 \geq \ldots \geq \lambda_{n-1}$ are by definition (see MARSHALL and OLKIN, 1979, p. 498) the singular values of $R^{-\frac{1}{2}}PC^{-\frac{1}{2}}$. Similarly, $1 \geq \lambda_1' \geq \ldots \geq \lambda_{n-1}'$ are the singular values of $R'^{-\frac{1}{2}}P'C'^{-\frac{1}{2}}$.

Furthermore, N and M are Markov matrices, so $Ne = e$ and $Me = e$ and it follows from $P = N^TP'M$ that $M^Tc' = c$ and $N^Tr' = r$. Hence $R^{-1}N^TR'N$ and $C^{-1}M^TC'M$ are Markov matrices and (by GANTMACHER, 1977, vol II, p. 83) have largest eigenvalue equal to unity. Thus all singular values of $R^{-\frac{1}{2}}N^TR'^{\frac{1}{2}}$ and $C'^{\frac{1}{2}}MC^{-\frac{1}{2}}$ are less than or equal to unity. Since $P = N^TP'M$, it follows that

$$R^{-\frac{1}{2}}PC^{-\frac{1}{2}} = R^{-\frac{1}{2}}N^TR'^{\frac{1}{2}}R'^{-\frac{1}{2}}P'C'^{-\frac{1}{2}}C'^{\frac{1}{2}}MC^{-\frac{1}{2}}$$

and application of the first inequality on p. 250 of MARSHALL and OLKIN (1979) to this product of (square) matrices yields the desired result. □

If (2.1.7) holds, then the canonical correlations of P' are said to *weakly submajorize* the canonical correlations of P. Hence for all

increasing Schur-convex functions ϕ of the canonical correlations,

$$\phi(\lambda_1,\ldots,\lambda_{n-1}) \leq \phi(\lambda'_1,\ldots,\lambda'_{n-1})$$

(cf. MARSHALL and OLKIN, 1979, p. 59). Examples of such increasing Schur-convex functions are

the first canonical correlation: $\phi_{CC}(\lambda_1,\ldots,\lambda_{n-1}) = \lambda_1$,

Pearson's χ^2 : $\phi_{CS}(\lambda_1,\ldots,\lambda_{n-1}) = \Sigma_{t=1}^{n-1} \lambda_t^2$,

the geometric mean : $\phi_{GM}(\lambda_1,\ldots,\lambda_{n-1}) = \Pi_{t=1}^{n-1} \lambda_t$.

(cf. MARSHALL and OLKIN, 1979, p. 63 and p. 73 respectively). This result for instance implies that pooling row (column) categories decreases the first canonical correlation and Pearson's χ^2. Another example of two probability tables P and P' satisfying the conditions of Theorem 2.1.5 is when P is a convex combination of P' and independence; i.e. $P = \alpha rc^T + (1-\alpha)P'$ for $0 \leq \alpha \leq 1$.

BENZÉCRI (1973) approaches CA from a geometrical viewpoint. He describes CA as a technique for displaying the rows and columns of a probability table as points in a q-dimensional vector space; where q is an integer less than or equal to n-1, usually q = 2 or q = 3. The i-th category of the row variable X is represented by a point ξ_i with coordinates $(\lambda_1\varphi_1(i),\ldots,\lambda_q\varphi_q(i))$, for $i = 1,\ldots,n$, and the j-th category of Y is represented by a point η_j with coordinates $(\lambda_1\psi_1(j),\ldots,\lambda_q\psi_q(j))$, for $j = 1,\ldots,m$. This yields $n+m$ points in a q-dimensional space. Benzécri's interpretation of this graphical representation is based on the following properties of the Euclidian distance, denoted by $\|.\|$, between row points and between column points.

PROPOSITION 2.1.6. *If* $q = n-1$ *or when* $\lambda_{q+1} = \ldots = \lambda_{n-1} = 0$, *then*

$$\|\xi_i\|^2 = \sum_{j=1}^{m} (P\{Y=j|X=i\} - P\{Y=j\})^2/P\{Y=j\} \qquad \textit{for } i = 1,\ldots,n,$$

$$\|\eta_j\|^2 = \sum_{i=1}^{n} (P\{X=i|Y=j\} - P\{X=i\})^2/P\{X=i\} \qquad \textit{for } j = 1,\ldots,m,$$

$$\|\xi_i - \xi_{i'}\|^2 = \sum_{j=1}^{m} (P\{Y=j|X=i\} - P\{Y=j|X=i'\})^2/P\{Y=j\} \qquad \textit{for } i,i' = 1,\ldots,n,$$

$$\|\eta_j - \eta_{j'}\|^2 = \sum_{i=1}^{n} (P\{X=i|Y=j\} - P\{X=i|Y=j'\})^2/P\{X=i\} \qquad \textit{for } j,j' = 1,\ldots,m.$$

PROOF. Follows by straightforward computations using (2.1.5) and (2.1.6). □

Thus in the complete graphical representation the distance between row point ξ_i and the origin indicates the extent to which the conditional distribution of $Y|X=i$ is similar to the marginal distribution of Y and thus can be interpreted as the contribution of the i-th category of the row variable to the dependence between X and Y. Analogously, the distance between two row points ξ_i and $\xi_{i'}$ indicates similarity of the conditional distributions of $Y|X=i$ and $Y|X=i'$. There exists, however, no clear interpretation of $\|\xi_i - \eta_j\|$. In the case that $q < n-1$ and $\lambda_{q+1} > 0$, the expressions for the distances in Proposition 2.1.6 only hold approximately and should be used with care.

Further results and properties and, in particular, applications of CA can be found in GIFI (1981), GREENACRE (1981, 1984), HILL (1974), LEBART et al. (1977) and NISHISATO (1980).

In some cases the reciprocal averaging formula's (2.1.4) or Benzécri's geometrical approach motivate applications of CA to matrices other than probability tables. An example is application to so called incidence matrices which are matrices consisting of unities and zero's specifying presence-absence of joint occurences of the row and column categories. The application of CA to incidence matrices is equivalent to Whittaker's gradient analysis (cf. HILL, 1973, 1974 and GAUCH, 1982).

Definition 2.1.1 of CA is also appropriate for bivariate distributions in general. Note that the canonical functions then not always exist (cf. RÉNYI, 1959). LANCASTER (1958, 1963) and CHESSON (1976) generalized some of the results in this section of which the following canonical decomposition is needed for later reference. Let $L^2(\mathbb{R},d\sigma)$ be the set of all real functions, square integrable with respect to a measure σ. *Suppose that* F *is a bivariate distribution function with marginals* F_r *and* F_c *and suppose that Pearson's* χ^2 *for* F, *defined by*

$$(2.1.8) \qquad \chi^2 = \int\int (dF/(dF_r dF_c))^2 dF_r dF_c - 1,$$

is finite. Then complete sets of orthonormal functions $\{1,\varphi_1,\varphi_2,\ldots\}$ *on* $L^2(\mathbb{R},dF_r)$ *and* $\{1,\psi_1,\psi_2,\ldots\}$ *on* $L^2(\mathbb{R},dF_c)$ *exist such that each member of a set of canonical functions appears as a member of the complete set of orthonormal functions. Furthermore, the canonical decomposition of* F *is given by*

$$(2.1.9) \qquad dF(x,y) = (1 + \sum_{t=1}^{\infty} \lambda_t \varphi_t(x)\psi_t(y))dF_r(x)dF_c(y) \qquad a.e.$$

where λ_t *is the* t-th *canonical correlation. Moreover,*

$$\chi^2 = \sum_{t=1}^{\infty} \lambda_t^2.$$

A converse of this result also holds. Such a canonical decomposition for the standard bivariate normal distribution is the well-known Mehler's expansion in which the canonical functions are Hermite polynomials and the canonical correlations are powers of the correlation parameter in absolute value. EAGLESON (1964) gives such decompositions for other well-known bivariate distributions (cf. Section 3.4). The canonical functions of continuous distributions give insight into its dependence structure but are of less practical value than in the contingency table case.

It can be proved (cf. DAUXOIS and POUSSE, 1977; MAUNG, 1941 and NAOURI, 1970) that the CA solution of discretisized continuous distributions approximates the continuous CA solution, provided the discretization is fine enough.

2.2. ASYMPTOTIC RESULTS

This section considers asymptotic distributions of the sample canonical correlations and functions. In particular, the limiting distributions under the null-hypothesis of independence and under local alternatives are considered. The sample canonical correlations $\hat{\lambda}_1^{(N)},\ldots,\hat{\lambda}_{n-1}^{(N)}$ and the vectors of the sample category scores $\hat{\varphi}_1^{(N)},\ldots,\hat{\varphi}_{n-1}^{(N)}$ and $\hat{\psi}_1^{(N)},\ldots,\hat{\psi}_{n-1}^{(N)}$ are the canonical correlations and canonical functions of the empirical probability table $\hat{P}^{(N)}$, which contains the relative frequencies of a sample of size N drawn from an underlying table P. Attention is not restricted to samples of N i.i.d. observations from P (multinomial sampling), but the larger class of *conditional Poisson distributions* for $\hat{P}^{(N)}$ is considered. A conditional Poisson distribution for instance arises when sampling is conditioned on (both) marginals.

Several results presented in this section are well-known under multinomial distributions for $\hat{P}^{(N)}$. This section gives generalizations to conditional Poisson sampling. It turns out that results concerning the problem of testing independence are the same as under multinomial sampling, provided the conditioning is only on marginals.

The conditional Poisson model for $\hat{P}^{(N)}$ as a generalization of the multinomial model is also considered in HABERMAN (1974, pp. 14-27). In conditional Poisson models the underlying probability table may be different for each sample size, so it is denoted by $P^{(N)}$, but it is assumed that

$$(2.2.1) \qquad P^{(N)} \to P \qquad \text{as } N \to \infty.$$

(The underlying table $P^{(N)}$ should not be confused with its empirical estimator $\hat{P}^{(N)}$). Let $Z^{(N)}$ be a $n \times m$ random matrix of independent Poisson variables with expectation $NP^{(N)}$. Furthermore let A be a matrix of size $k \times mn$ consisting of integer elements, which do not depend on the sample size, and such that the vector e is in the linear subspace spanned by the rows of A. Then $\hat{P}^{(N)}$ is said to have a *conditional Poisson distribution* when

$$(2.2.2) \qquad \hat{P}^{(N)} \sim (N^{-1}Z^{(N)} \mid A\,vec\,[Z^{(N)} - NP^{(N)}] = 0),$$

provided $P\{A\,vec\,[Z^{(N)} - NP^{(N)}] = 0\} > 0$. The assumption that e can be written as a linear combination of the rows of A implies that the sample size equals N. Some well-known specific cases of conditional Poisson models are given in Examples 2.2.1-2.2.3 below. HABERMAN (1974, p. 18) derived the following asymptotic result for conditional Poisson distributions.

LEMMA 2.2.1. *Suppose that the underlying probability table* $P^{(N)}$ *satisfies* (2.2.1) *with* $p_{ij} > 0$ *for* $i = 1,\ldots,n$ *and* $j = 1,\ldots,m$ *and that* $\hat{P}^{(N)}$ *has the conditional Poisson distribution* (2.2.2), *then*

$$N^{\frac{1}{2}}\,vec\,[\hat{P}^{(N)} - P^{(N)}] \to_D N_{nm}(0,\Sigma) \qquad as\ N \to \infty$$

with covariance matrix

$$(2.2.3) \qquad \Sigma = D - DA^{T}(ADA^{T})^{-}AD$$

where $D = diag\,[P]$ *and* $(ADA^{T})^{-}$ *is a reflexive generalized inverse of* ADA^{T}.

The asymptotic sampling properties of the canonical correlations will be derived from asymptotic distributions of the $n \times m$ random matrix $\hat{W}^{(N)}$ with elements

$$(2.2.4) \qquad \hat{w}_{ij}^{(N)} = (\hat{p}_{ij}^{(N)} - \hat{r}_{i}^{(N)}\hat{c}_{j}^{(N)})/(\hat{r}_{i}^{(N)}\hat{c}_{j}^{(N)})^{\frac{1}{2}} \qquad i = 1,\ldots,n;\ \ j = 1,\ldots,m,$$

where $\hat{w}_{ij}^{(N)}$ is defined zero when $\hat{r}_i^{(N)}$ or $\hat{c}_j^{(N)}$ equals zero. The matrix $\hat{W}^{(N)}$ describes the deviation from independence in the sample. The matrix $W^{(N)}$ corresponding to the underlying table $P^{(N)}$ is similarly defined and satisfies $W^{(N)} \to W$ as $N \to \infty$. Furthermore, it is assumed (throughout) that the marginals r and c of the limiting underlying probability table P are strictly positive. The following lemma shows that the canonical correlations $\lambda_1,\ldots,\lambda_{n-1}$ and the canonical functions $\varphi_1,\ldots,\varphi_{n-1}$ and $\psi_1,\ldots,\psi_{n-1}$ can be obtained from the matrix W.

LEMMA 2.2.2. *The squared canonical correlations* $\lambda_1^2 \geq \ldots \geq \lambda_{n-1}^2$ *and* 0 *are the eigenvalues of the matrix* $WW^\top$. *Furthermore, the vectors* $\varphi_1,\ldots,\varphi_{n-1},e$ *are eigenvectors of* $R^{-\frac{1}{2}}WW^\top R^{\frac{1}{2}}$ *corresponding to eigenvalues* $\lambda_1 \geq \ldots \geq \lambda_{n-1} \geq 0$ *and the vectors* $\psi_1,\ldots,\psi_{n-1},e$ *are eigenvectors of* $C^{-\frac{1}{2}}W^\top WC^{\frac{1}{2}}$ *corresponding to the same eigenvalues.*

PROOF. Follows from Theorem 2.1.3 and the fact that $WW^\top = R^{-\frac{1}{2}}PC^{-1}P^\top R^{-\frac{1}{2}} - R^{\frac{1}{2}}ee^\top R^{\frac{1}{2}}$. □

Of course similar results hold for the canonical correlations and functions of $P^{(N)}$ and the empirical $\hat{P}^{(N)}$ provided inverses are replaced by generalized inverses.

LEMMA 2.2.3. *Under the assumptions of Lemma* 2.2.1,

$$(2.2.5) \qquad N^{\frac{1}{2}}\, vec\,[\hat{W}^{(N)} - W^{(N)}] \to_D N_{nm}(0, G\Xi G^\top) \qquad as\ N \to \infty$$

where Ξ *is given by* (2.2.3), $D = diag\,[P]$ *and*

$$(2.2.6) \qquad G = (R^{-\frac{1}{2}} \otimes C^{-\frac{1}{2}})(I - \tfrac{1}{2}(D + R \otimes C)(R^{-1} \otimes ee^\top + ee^\top \otimes C^{-1})).$$

PROOF. Consider the nm-dimensional function g such that $g(vec\,[P]) = vec\,[W]$ and thus also $g(vec\,[\hat{P}^{(N)}]) = vec\,[\hat{W}^{(N)}]$. Since the marginals of P are strictly positive, this function is continuously differentiable in a neighborhood of P. Routine differentiation shows that the $nm \times nm$ matrix G of first derivatives of g at P,

$$g_{ij,kl} = \left.\frac{\partial g_{ij}(Z)}{\partial z_{kl}}\right|_{Z=P},$$

is given by (2.2.6). Application of the ∂-method (cf. RAO, 1973, p. 388) to this function g and the limiting distribution of $N^{\frac{1}{2}}\, vec\,[\hat{P}^{(N)} - P^{(N)}]$ in

Lemma 2.2.1 gives the desired result. (Note that the ∂-method, as formulated in Rao, also holds when the parameter depends on N but converges to a limit, and the function g is continuously differentiable in a neighbourhood of this limit point.) □

Since eigenvalues are continuous functions of the matrix elements and are even differentiable when the eigenvalue is simple (cf. WILKINSON, 1965, pp. 67-69), the ∂-method can again be applied to obtain the joint distribution of $N^{\frac{1}{2}}(\hat{\lambda}_t^{(N)2} - \lambda_t^{(N)2})$ for $t = 1,\ldots,n-1$. This produces the following result which is, in a less transparent notation, also given in O'NEILL (1978) for the special case of multinomial sampling.

THEOREM 2.2.4. *Suppose that the conditions of Lemma* 2.2.1 *hold and that all canonical correlations of* P *are simple, i.e.* $\lambda_1 > \lambda_2 > \ldots > \lambda_{n-1} > 0$, *then*

$$(2.2.7) \qquad N^{\frac{1}{2}}\begin{pmatrix} \hat{\lambda}_1^{(N)2} - \lambda_1^{(N)2} \\ \vdots \\ \hat{\lambda}_{n-1}^{(N)2} - \lambda_{n-1}^{(N)2} \end{pmatrix} \to_D N_{nm}(0, 4\Lambda(R^{\frac{1}{2}}\Phi \otimes C^{\frac{1}{2}}\Psi)^\top G \ddagger G^\top (R^{\frac{1}{2}}\Phi \otimes C^{\frac{1}{2}}\Psi)\Lambda)$$

as $N \to \infty$, *where the* $n \times (n-1)$ *matrix* Φ *has the row category scores* $\varphi_1,\ldots,\varphi_{n-1}$ *as columns, the* $m \times (n-1)$ *matrix* Ψ *has the column category scores* $\psi_1,\ldots,\psi_{n-1}$ *as columns,* $\Lambda = diag\,[\lambda_1,\ldots,\lambda_{n-1}]$ *and* G *and* $\ddagger$ *are given by* (2.2.6) *and* (2.2.3).

PROOF. O'NEILL (1978) derives (2.2.7) from (2.2.5). Alternatively, one can again apply the ∂-method as indicated above. □

This result gives information about asymptotic sampling variation of the canonical correlations. (Second order asymptotics are given in O'NEILL, 1978 b.) It follows from Lemma 2.2.3 that $\hat{W}^{(N)} \to_P W$ as $N \to \infty$ and hence

$$(2.2.8) \qquad \hat{\lambda}_t^{(N)} \to_P \lambda_t \qquad \text{as } N \to \infty$$

and, in case that λ_t is simple, also

$$(2.2.9) \qquad \hat{\varphi}_t^{(N)} \to_P \varphi_t, \quad \hat{\psi}_t^{(N)} \to_P \psi_t \qquad \text{as } N \to \infty$$

for $t = 1,\ldots,n-1$ (because eigenvectors are continuous functions of the matrix elements when the corresponding eigenvalue is simple; cf. WILKINSON, 1965, p. 67). Thus the sample canonical correlations are (weakly) consistent estimators for λ_t, $t = 1,\ldots,n-1$. But it follows from MARSHALL and

OLKIN (1979, p. 357) that for each N,

$$\sum_{s=1}^{t} E\hat{\lambda}_s^{(N)} \geq \sum_{s=1}^{t} \lambda_s^{(N)} \qquad \text{for } t = 1,\ldots,n-1,$$

which means that these partial sums of sample canonical correlations are positively biased estimators in finite samples (see also O'NEILL, 1978 b).

The remaining part of this chapter considers asymptotic properties of tests of independence based on canonical correlations or canonical variables. Such tests are considered in the next section, but the limiting distributions of canonical correlations and functions under the hypothesis of independence and local alternatives needed for this are derived in this section.

These limiting distributions are considerably simplified when the conditional Poisson distribution of $\hat{P}^{(N)}$ is only conditioned on some (linear combinations) of the marginals $\hat{r}^{(N)}$ and $\hat{c}^{(N)}$. So it is assumed in the sequel that the conditioning matrix A is of the form

$$A^T = (I_n \otimes e_m \,\vdots\, e_n \otimes I_m)B \tag{2.2.10}$$

for some $(n+m) \times k$ matrix B. The following three examples demonstrate that the conditional Poisson model (2.2.2) with conditioning matrix A of the form (2.2.10) covers most practical sampling models.

EXAMPLE 2.2.1. Commonly, the contingency table $N\hat{P}^{(N)}$ arises from N independent and identical samples from the probability table P and thus $N\hat{P}^{(N)}$ has a multinomial distribution with parameters N and P. In order to show that this multinomial distribution can be considered as a conditional Poisson distribution let $Z^{(N)}$ be a matrix of independent Poisson variables with expectation NP and let $A^T = e$. Then since $e^T vec\,[Z^{(N)}]$ is a Poisson variable with expectation N, evaluation of the density function (with respect to the counting measure) of $(Z^{(N)} \mid e^T vec\,[Z^{(N)}] = N)$ produces the desired multinomial.

EXAMPLE 2.2.2. In some cases the contingency table $N\hat{P}^{(N)}$ arises from a sample in which the column sums are kept fixed in advance at say $N_1, N_2, \ldots, N_m$ where $\Sigma_1^m N_j = N$. This situation corresponds to independent multinomial samples for each column; the j-th column arises from a multinomial sample of size N_j and probability vector $(p_{1|j},\ldots,p_{n|j})$ with $\Sigma_{i=1}^n p_{i|j} = 1$. Making the natural assumption that the underlying probability table $P^{(N)}$ has elements $p_{ij}^{(N)} = p_{i|j}N_j/N$, it follows, as in the example

above, that sampling conditional on column sums leads to a conditional Poisson distribution with conditioning matrix $A^T = (e_n \otimes I_m)$.

EXAMPLE 2.2.3. Suppose that sampling is conditional on both row and column sums, i.e. both marginals $N\hat{r}^{(N)}$ and $N\hat{c}^{(N)}$ of the contingency table are kept fixed. Again assume that the marginals $r^{(N)}$ and $c^{(N)}$ of the underlying table $P^{(N)}$ are equal to those fixed values of $\hat{r}^{(N)}$ and $\hat{c}^{(N)}$. The distribution of $N\hat{P}^{(N)}$ can be considered as a conditional multinomial distribution with parameters N and $P^{(N)}$, conditioned on both marginals. Therefore, it follows from Example 2.2.1 that $\hat{P}^{(N)}$ has a conditional Poisson distribution with conditioning matrix $A^T = (I_n \otimes e_m \vdots e_n \otimes I_m)$. Under the hypothesis of independence $P^{(N)} = r^{(N)}c^{(N)T}$, the distribution of $N\hat{P}^{(N)}$ is the generalized multiple hypergeometric distribution (cf. LEHMANN, 1975, p. 384). In classical contingency table analysis it is quite common to perform the analysis conditionally on the marginals (even when sampling is unconditional). This gives extra motivation for considering this specific conditional Poisson distribution.

Let's turn to the asymptotic distributions of the canonical correlations and functions under the null-hypothesis of independence and under local alternatives. Let Π_0 be a set of $n \times m$ null-hypothesis probability tables with marginals bounded away from zero, e.g. for arbitrary $\varepsilon > 0$ define $\Pi_0 = \{P : p_{ij} = r_i c_j \geq \varepsilon \text{ for } i = 1,\ldots,n;\ j = 1,\ldots,m\}$. Introduction of such Π_0 is unnatural, but only for such restricted hypotheses can the tests be shown to have the proper asymptotic size (cf. Section 2.3). Denote by $P_0 = \{P_0^{(N)}\}_{N=1}^{\infty}$ the sequence of *null-hypothesis conditional Poisson distributions* with conditioning matrix A of the form (2.2.10) drawn from underlying tables

$$(2.2.11) \qquad P^{(N)} = r^{(N)}c^{(N)T} \in \Pi_0$$

where $r^{(N)} \to r$ and $c^{(N)} \to c$ as $N \to \infty$. Consequently, $r > 0$ and $c > 0$. Furthermore, consider alternative probability tables $P'^{(N)}$ with marginals $r^{(N)}$ and $c^{(N)}$ such that $P'^{(N)} \to P'$. Assume that the elements of $P'^{(N)}$ are bounded away from zero. Let $\lambda_1^{(N)},\ldots,\lambda_{n-1}^{(N)}$ denote the canonical correlations and $\varphi_1^{(N)},\ldots,\varphi_{n-1}^{(N)}$ and $\psi_1^{(N)},\ldots,\psi_{n-1}^{(N)}$ the corresponding canonical functions of $P'^{(N)}$. Since $P'^{(N)} \to P'$, they can be chosen such that $\lambda_t^{(N)} \to \lambda_t$, $\varphi_t^{(N)} \to \varphi_t$ and $\psi_t^{(N)} \to \psi_t$ for $t = 1,\ldots,n-1$ as $N \to \infty$, where $(\lambda_t,\varphi_t,\psi_t)$ is the t-th canonical triplet of P'. Let $W'^{(N)}$ be the $n \times m$ matrix with elements

$$w'^{(N)}_{ij} = (p'^{(N)}_{ij} - r^{(N)}_i c^{(N)}_j)/(r^{(N)}_i c^{(N)}_j)^{\frac{1}{2}} \qquad i = 1,\ldots,n;\ j = 1,\ldots,m.$$

Clearly, $W'^{(N)} \to W'$ as $N \to \infty$. Denote by $P_L = \{P_L^{(N)}\}_{N=1}^{\infty}$ the sequence of *local alternative conditional Poisson distributions* (with conditioning matrix A of the form (2.2.10)) drawn from the local contamination alternative tables

$$(2.2.12) \qquad P^{(N)} = r^{(N)}c^{(N)\top} + \theta N^{-\frac{1}{2}}(P'^{(N)} - r^{(N)}c^{(N)\top}), \qquad \theta > 0.$$

The matrix $W^{(N)}$, defined similarly to (2.2.4), for this local alternative table equals $\theta N^{-\frac{1}{2}}W'^{(N)}$. Therefore, the canonical correlations of the table (2.2.12) are $\theta N^{-\frac{1}{2}}\lambda_1^{(N)},\ldots,\theta N^{-\frac{1}{2}}\lambda_{n-1}^{(N)}$ and its canonical functions are $\varphi_1^{(N)},\ldots,\varphi_{n-1}^{(N)}$ and $\psi_1^{(N)},\ldots,\psi_{n-1}^{(N)}$. Furthermore, let $P_A = \{P_A^{(N)}\}_{N=1}^{\infty}$ denote a sequence of *fixed alternative conditional Poisson distributions* drawn from the underlying tables $P'^{(N)}$.

It can be shown by standard methods that the sequences P_0 and P_L are *contiguous*.

The following theorem gives the limiting distributions of the canonical correlations and functions under the null-hypothesis and under local alternative conditional Poisson distributions. The first part of this theorem generalizes results of O'NEILL (1978) (asymptotic null distribution of canonical correlations under multinomial sampling), CORSTEN (1976) (asymptotic null distribution of $N\hat{\lambda}_1^{(N)2}$ under multiple generalized hypergeometric sampling) and of HABERMAN (1981) (asymptotic distribution of $N\hat{\lambda}_1^{(N)2}$ under local alternative multinomial sampling; his proof is rather sketchy). Let $W_{n-1}(m-1,I)$ and $W_{n-1}(m-1,I,\theta^2\Lambda\Lambda^\top)$ denote the central and non-central standard Wishart distributions of dimension n-1 and with m-1 degrees of freedom. Thus when X denotes a $(n-1)\times(m-1)$ standard normally distributed matrix, i.e. $vec\,[X] \sim N_{(n-1)(m-1)}(0,I)$, and Λ denotes the $(n-1)\times(m-1)$ "diagonal" matrix with diagonal elements $\lambda_1,\ldots,\lambda_{n-1}$, then $XX^\top \sim W_{n-1}(m-1,I)$ and $(X+\theta\Lambda)(X+\theta\Lambda)^\top \sim W_{n-1}(m-1,I,\theta^2\Lambda\Lambda^\top)$. Furthermore, in order to make the sample canonical functions asymptotically unique with probability one, it is assumed in the theorem below that the sign of each vector $\hat{\varphi}_t^{(N)}$, for t = 1,...,n-1, is such that the first non-zero component is strictly negative. The sign of $\hat{\psi}_t^{(N)}$ is determined by

$$\hat{\lambda}_t^{(N)}\hat{C}^{(N)}\hat{\psi}_t^{(N)} = \hat{P}^{(N)\top}\hat{\varphi}_t^{(N)}$$

for t = 1,...,n-1.

THEOREM 2.2.5. (i) *The asymptotic joint distribution of* $N\hat{\lambda}_1^{(N)2},\ldots,N\hat{\lambda}_{n-1}^{(N)2}$ *under* P_0 *is the same as the joint distribution of the eigenvalues of a* $W_{n-1}(m-1,I)$ *distributed matrix and their asymptotic joint distribution under* P_L *is the same as the joint distribution of the eigenvalues of a* $W_{n-1}(m-1,I,\theta^2\Lambda\Lambda^T)$ *distributed matrix.*

(ii) *The asymptotic marginal distribution of each* $\hat{\varphi}_t^{(N)}$, *for* $t = 1,\ldots,n-1$, *under* P_0 *is the uniform distribution on the* $(n-1)$*-dimensional half ellipsoid* $\{u \in \mathbb{R}^n : r^Tu = 0,\ u^TRu = 1,\ u_1 \le 0\}$. *The asymptotic marginal distribution of each* $\hat{\psi}_t^{(N)}$, *for* $t = 1,\ldots,n-1$, *under* P_0 *is the uniform distribution on the* $(m-1)$*-dimensional ellipsoid* $\{v \in \mathbb{R}^m : c^Tv = 0,\ v^TCv = 1\}$.

(iii) *Moreover, denoting* $\hat{\psi}_t^{(N)*} = \pm\hat{\psi}_t^{(N)}$ *where the sign is randomly chosen for* $t = 1,\ldots,n-1$, *the sets* $\{\hat{\varphi}_1^{(N)},\ldots,\hat{\varphi}_{n-1}^{(N)}\},\{\hat{\psi}_1^{(N)*},\ldots,\hat{\psi}_{n-1}^{(N)*}\}$ *and* $\{N\hat{\lambda}_1^{(N)2},\ldots,N\hat{\lambda}_{n-1}^{(N)2}\}$ *are under* P_0 *pairwise asymptotically stochastically independent.*

PROOF. First consider the random matrix $\hat{W}^{(N)}$ defined by (2.2.4). It follows by Lemma 2.2.3 that under P_0

$$N^{\frac{1}{2}}\,vec\,[\hat{W}^{(N)}] \to_D N_{nm}(0,G\ddagger G^T) \qquad \text{as } N \to \infty$$

and under P_L

$$N^{\frac{1}{2}}\,vec\,[\hat{W}^{(N)} - \theta N^{-\frac{1}{2}}W'^{(N)}] \to_D N_{nm}(0,G\ddagger G^T) \qquad \text{as } N \to \infty$$

where in both cases

$$G = (R^{-\frac{1}{2}} \otimes C^{-\frac{1}{2}})(I - (R \otimes C)(R^{-1} \otimes ee^T + ee^T \otimes C^{-1}))$$

$$\ddagger = (R \otimes C) - (R \otimes C)A^T(A(R \otimes C)A^T)^{-}A(R \otimes C)$$

since in both cases $diag\,[P] \to R \otimes C$ as $N \to \infty$. Furthermore, since the sample size equals N, $e^T\,vec\,[\hat{P}^{(N)} - P^{(N)}] = 0$ a.s. and hence $e^T\ddagger = 0^T$ and $\ddagger e = 0$. Therefore the covariance matrix equals $G\ddagger G^T = \Gamma(R^{-\frac{1}{2}} \otimes C^{-\frac{1}{2}})\ddagger(R^{-\frac{1}{2}} \otimes C^{-\frac{1}{2}})\Gamma^T$ where

$$\Gamma = (G + (R^{\frac{1}{2}} \otimes C^{\frac{1}{2}})ee^T)(R^{\frac{1}{2}} \otimes C^{\frac{1}{2}}) = (I - R^{\frac{1}{2}}ee^TR^{\frac{1}{2}}) \otimes (I - C^{\frac{1}{2}}ee^TC^{\frac{1}{2}}). \tag{2.2.13}$$

Since A is of the form (2.2.10), it follows that

$$\Gamma(R^{\frac{1}{2}} \otimes C^{\frac{1}{2}})A^T = \Gamma(R^{\frac{1}{2}} \otimes C^{\frac{1}{2}}e \,\vdots\, R^{\frac{1}{2}}e \otimes C^{\frac{1}{2}})B = 0.$$

Hence the covariance matrix equals $G\ddagger G^T = \Gamma\Gamma^T = \Gamma$. Thus under P_0

(2.2.14) $\quad N^{\frac{1}{2}} vec\,[\hat{W}^{(N)}] \to_D N_{nm}(0,\Gamma) \qquad$ as $N \to \infty$

and since $W'^{(N)} \to W'$, under P_L

(2.2.15) $\quad N^{\frac{1}{2}} vec\,[\hat{W}^{(N)}] \to_D N_{nm}(\theta\, vec\,[W'],\Gamma)$ as $N \to \infty$.

Next consider the probability table P'. Recall that its canonical functions $\varphi_1,\ldots,\varphi_{n-1}$ together with the vector e span $\mathbb{R}^n$. Let Φ denote the $n \times (n-1)$ matrix with $\varphi_1,\ldots,\varphi_{n-1}$ as columns, then $R^{\frac{1}{2}}\Phi\Phi^\top R^{\frac{1}{2}} = I - R^{\frac{1}{2}}ee^\top R^{\frac{1}{2}}$. The vectors $e,\psi_1,\ldots,\psi_{n-1}$ do not span $\mathbb{R}^m$ when $m > n$, therefore define $\psi_n,\ldots,\psi_{m-1}$ such that the $m \times (m-1)$ matrix Ψ with $\psi_1,\ldots,\psi_{n-1},\psi_n,\ldots,\psi_{m-1}$ as columns satisfies $\Psi^\top C\Psi = I_{m-1}$ and $\Psi^\top c = 0$. The columns of Ψ together with e now span $\mathbb{R}^m$ and hence $C^{\frac{1}{2}}\Psi\Psi^\top C^{\frac{1}{2}} = I - C^{\frac{1}{2}}ee^\top C^{\frac{1}{2}}$. Furthermore, since $\lambda_1,\ldots,\lambda_{n-1}$ are its canonical correlations, it follows from the canonical decomposition (2.1.6) that

$$W' = R^{\frac{1}{2}}\left(\sum_{t=1}^{n-1} \lambda_t \varphi_t \psi_t^\top\right) C^{\frac{1}{2}} = R^{\frac{1}{2}}\Phi\Lambda\Psi^\top C^{\frac{1}{2}},$$

where Λ is the $(n-1) \times (m-1)$ "diagonal" matrix with $\lambda_1,\ldots,\lambda_{n-1}$ as diagonal elements. Hence for a standard normally distributed $(n-1) \times (m-1)$ matrix X,

$$vec\,[R^{\frac{1}{2}}\Phi X\Psi^\top C^{\frac{1}{2}}] \sim N_{nm}(0,R^{\frac{1}{2}}\Phi\Phi^\top R^{\frac{1}{2}} \otimes C^{\frac{1}{2}}\Psi\Psi^\top C^{\frac{1}{2}}) = N_{nm}(0,\Gamma)$$

and

$$vec\,[R^{\frac{1}{2}}\Phi(X + \theta\Lambda)\Psi^\top C^{\frac{1}{2}}] \sim N_{nm}(\theta\, vec\,[W'],\Gamma).$$

Combining this with (2.2.14) and (2.2.15) yields that under P_0,

(2.2.16) $\quad N^{\frac{1}{2}}\hat{W}^{(N)} \to_D R^{\frac{1}{2}}\Phi X\Psi^\top C^{\frac{1}{2}} \qquad$ as $N \to \infty$

and under P_L

(2.2.17) $\quad N^{\frac{1}{2}}\hat{W}^{(N)} \to_D R^{\frac{1}{2}}\Phi(X + \theta\Lambda)\Psi^\top C^{\frac{1}{2}} \qquad$ as $N \to \infty$.

(i) By Lemma 2.2.2, $N\hat{\lambda}_1^{(N)2},\ldots,N\hat{\lambda}_{n-1}^{(N)2}$ are the non-zero eigenvalues of $N\hat{W}^{(N)}\hat{W}^{(N)\top}$ and since eigenvalues are continuous functions of the matrix elements, it follows by the continuous mapping theorem that $N\hat{\lambda}_1^{(N)2},\ldots,N\hat{\lambda}_{n-1}^{(N)2}$ converge under P_0 in distribution to the non-zero eigenvalues of $R^{\frac{1}{2}}\Phi X\Psi^\top C\Psi X^\top\Phi^\top R^{\frac{1}{2}}$, that is to the eigenvalues of $XX^\top$. The result under P_L follows similarly.

(iii) Since an eigenvector corresponding to a simple eigenvalue is a

continuous function of the matrix elements (cf. WILKINSON, 1965, p. 67) and since $P\{\text{all eigenvalues of } \Phi XX^T\Phi^T R \text{ are simple}\} = 1$, it follows from Lemma 2.2.2 and the continuous mapping theorem that the joint distribution of $\hat{\varphi}_1^{(N)},\ldots,\hat{\varphi}_{n-1}^{(N)}$ converges under P_0 in distribution to the joint distribution of the (suitably normalized) eigenvectors of $\Phi XX^T\Phi^T R$ corresponding to non-zero eigenvalues. Similarly, $\hat{\psi}_1^{(N)},\ldots,\hat{\psi}_{n-1}^{(N)}$ converge under P_0 in distribution to the joint distribution of the (suitably normalized) eigenvectors of $\Psi X^T X\Psi^T C$ corresponding to non-zero eigenvalues. Furthermore, it is easily seen that $h_1,\ldots,h_{n-1}$ form a complete set of orthonormal eigenvectors of XX^T iff $\Phi h_1,\ldots,\Phi h_{n-1}$ and e form a complete set of eigenvectors of $\Phi XX^T\Phi^T R$. Let H^* denote the $(n-1)\times(n-1)$ orthonormal matrix with the vectors $h_1^*,\ldots,h_{n-1}^*$ as columns, where $h_t^* = \pm h_t$ with random choice of sign for $t = 1,\ldots,n-1$. Then for any $(n-1)\times(n-1)$ orthonormal matrix Q, the columns of QH^* form a complete set of orthonormal eigenvectors of QXX^TQ^T and, moreover, $QXX^TQ^T \sim XX^T$. This implies that the joint distribution of the eigenvectors of QXX^TQ^T does not depend on Q and thus $QH^* \sim H^*$. Therefore, H^* has the so called *Haar invariant distribution* (cf. ANDERSON, 1958, p. 321). It also follows that
$(QXX^TQ^T \mid \text{all eigenvalues of } QXX^TQ^T) \sim (XX^T \mid \text{all eigenvalues of } XX^T)$, but since the eigenvalues of QXX^TQ^T and XX^T are the same,
$(QH^* \mid \text{eigenvalues of } XX^T) \sim (H^* \mid \text{eigenvalues of } XX^T)$. Thus
$(H^* \mid \text{all eigenvalues of } XX^T)$ also has the unique Haar invariant distribution, i.e. $(H^* \mid \text{all eigenvalues of } XX^T) \sim H^*$. So H^* and the eigenvalues of XX^T are independently distributed (see also ANDERSON, 1958, p. 322). Similarly, let K^* denote the $(m-1)\times(n-1)$ orthogonal matrix with the eigenvectors (random choice of sign) of X^TX corresponding to non-zero eigenvalues. Then $(QXX^TQ^T \mid \text{all eigenvectors of } X^TQ^TQX) \sim (XX^T \mid \text{all eigenvectors of } X^TX)$ which implies that H^* and K^* are independently distributed. This proves pairwise independence of ΦH^*, ΨK^* and the eigenvalues of XX^T and hence proves pairwise asymptotic independence of $\{\hat{\varphi}_1^{(N)*},\ldots,\hat{\varphi}_{n-1}^{(N)*}\}$, $\{\hat{\psi}_1^{(N)*},\ldots,\hat{\psi}_{n-1}^{(N)*}\}$ and $\{N\hat{\lambda}_1^{(N)2},\ldots,N\hat{\lambda}_{n-1}^{(N)2}\}$, where the superscript * denotes random choice of sign. Since the signs of the vectors $\hat{\varphi}_t^{(N)}$ and $\hat{\psi}_t^{(N)}$ are related, the * can only be removed in one of the two sets.

(ii) Each column h_t^* of H^* is uniformly distributed on the unit sphere in $\mathbb{R}^{n-1}$ (cf. MUIRHEAD, 1982, p. 37) and therefore Φh_t^* is uniformly distributed on $\{u \in \mathbb{R}^n : r^Tu = 0,\ u^TRu = 1\}$. Since it is assumed that vectors of row category scores have first non-zero component strictly negative, it

follows that the limiting distribution under P_0 of $\hat{\varphi}_t^{(N)}$ is the uniform distribution on $\{u \in \mathbb{R}^n : r^\top u = 0,\ u^\top R u = 1,\ u_1 \le 0\}$. Similarly, it follows that under P_0, $\hat{\psi}_t^{(N)*}$ is asymptotically uniformly distributed on $\{v \in \mathbb{R}^m : c^\top v = 0,\ v^\top C v = 1\}$. Since under $P_0^{(N)}$, $N^{\frac{1}{2}}\hat{W}^{(N)} \sim -N^{\frac{1}{2}}\hat{W}^{(N)}$ it follows that under $P_0^{(N)}$, $\hat{\psi}_t^{(N)} \sim \hat{\psi}_t^{(N)*}$. □

2.3. TESTS OF INDEPENDENCE

This section briefly discusses some properties of tests of independence based on canonical correlations and canonical variables. The limiting distributions of the test statistics considered under null-hypothesis and local alternative conditional Poisson distributions easily follow from Theorem 2.2.5. It is shown that tests based on canonical correlations, at least the ones considered, are generally consistent and asymptotically unbiased. Tests based on canonical variables are generally not consistent. Furthermore, tests are compared by means of approximate Bahadur and Pitman efficiencies.

Tests most common in contingency table analysis are the *chi-square test*, which rejects for large values of

$$T_{CS}^{(N)} = N \sum_{i=1}^{n} \sum_{j=1}^{m} \hat{w}_{ij}^{(N)2} = N \sum_{t=1}^{n-1} \hat{\lambda}_t^{(N)2},$$

and the *likelihood ratio test*, which rejects for large values of

$$T_{LR}^{(N)} = 2N \sum_{i=1}^{n} \sum_{j=1}^{m} \hat{p}_{ij}^{(N)} \log\left(\hat{p}_{ij}^{(N)} / (\hat{r}_i^{(N)} \hat{c}_j^{(N)})\right).$$

(This test is not based on canonical correlations, but is nevertheless considered in this section). Less familiar is the *canonical correlation test* which rejects for large values of

$$T_{CC}^{(N)} = N\hat{\lambda}_1^{(N)2}.$$

It follows from Theorem 2.2.5 (i) that the asymptotic testing problem (against local alternatives) is equivalent to the problem of testing the hypothesis $EY = 0$ against the alternative $EY = \theta\Lambda$, where the columns of Y are independently and normally distributed with common covariance matrix (identity matrix). This is exactly the MANOVA problem with known covariance matrix. So one might expect that well-known tests for the MANOVA problem have a good performance in contingency table analysis. The likelihood

ratio statistic for the MANOVA problem corresponds to the *geometric mean* of the squared canonical correlations,

$$T_{GM}^{(N)} = N\left(\prod_{t=1}^{n-1} \hat{\lambda}_t^{(N)}\right)^{2/(n-1)}.$$

The statistic

$$T_P^{(N)} = N \sum_{t=1}^{n-1} \hat{\lambda}_t^{(N)2}/(1+\hat{\lambda}_t^{(N)2})$$

is based on *Pillai's* test for the MANOVA problem. A variation on this is

$$T_Q^{(N)} = N \sum_{t=1}^{n-1} \hat{\lambda}_t^{(N)2}/(1-\hat{\lambda}_t^{(N)2}).$$

COROLLARY 2.3.1. *The asymptotic distribution of each of the statistics* $T_{CS}^{(N)}$, $T_{LR}^{(N)}$, $T_P^{(N)}$ *and* $T_Q^{(N)}$ *under the hypothesis* P_0 *is chi-square with* $(n-1)(m-1)$ *degrees of freedom; their asymptotic distribution under the local alternative* P_L *is non-central chi-square with* $(n-1)(m-1)$ *degrees of freedom and non-centrality parameter* $\theta^2 \Sigma_1^{n-1} \lambda_t^2$. *The asymptotic distribution of* $T_{CC}^{(N)}$ *under* P_0 *and* P_L *is the distribution of the largest eigenvalue of a* $W_{n-1}(m-1,I)$ *and a* $W_{n-1}(m-1,I,\theta^2\Lambda\Lambda^T)$ *distributed matrix, respectively. The asymptotic distribution of* $T_{GM}^{(N)}$ *under* P_0 *is equal to the distribution of the geometric mean of* $n-1$ *independent chi-square variables with degrees of freedom* $m-1, m-2, \ldots, m-n+1$.

PROOF. The limiting distributions of $T_{CS}^{(N)}$ and $T_{CC}^{(N)}$ follow directly from Theorem 2.2.5. The asymptotic distribution of $T_{GM}^{(N)}$ under P_0 follows from the result (×) on p. 540 in RAO (1973). Furthermore, it is easily seen that $\hat{\lambda}_t^{(N)}$ for $t = 1,\ldots,n-1$ tends to zero in probability under P_0 as well as under P_L as $N \to \infty$. Therefore, $T_{CS}^{(N)}$, $T_P^{(N)}$ and $T_Q^{(N)}$ are asymptotically equivalent under null-hypothesis and local alternatives. Expansion of the logarithm shows that $T_{LR}^{(N)}$ and $N \Sigma_{i=1}^{n} \Sigma_{j=1}^{m} \hat{w}_{ij}^{(N)2}$ are, under P_0 and under P_L, asymptotically equivalent. □

The limiting distributions of $T_{CS}^{(N)}$ and $T_{LR}^{(N)}$ under P_0 and P_L also follow from HABERMAN (1974, pp. 109-110).

In practice tests based on these six statistics can be performed by rejecting the null-hypothesis when the test statistic exceeds the upper α-point of the corresponding limiting (null-hypothesis) distribution. Since these limiting distributions are valid under any convergent sequence $\{P^{(N)} = r^{(N)}c^{(N)T}\}_{N=1}^{\infty}$ in the null-hypothesis Π_0, the null-hypothesis dis-

tribution of these statistics converges *uniformly* in the elements of Π_0. Hence for each statistic $T^{(N)}$

$$\lim_{N\to\infty} \sup_{P^{(N)}\in\Pi_0} P\{T^{(N)} \geq t_{1-\alpha}\} = \alpha.$$

So all tests have asymptotic size α. In case that the marginals of the tables in Π_0 are not bounded away from zero, the sequence of probability tables in Π_0 for which the supremum is attained could tend to a limiting table with one or more zero marginals. For such sequences the derived limiting distributions are not valid and therefore the limiting tail probability may differ from α. In fact, in 2×2 tables this tail probability exceeds α.

The following theorem shows that these tests have the desirable properties of consistency and asymptotic unbiasedness; extending earlier results of HABERMAN (1981).

THEOREM 2.3.2. (i) *Tests based on the statistics* $T_{CS}^{(N)}$, $T_{LR}^{(N)}$, $T_{CC}^{(N)}$, $T_{P}^{(N)}$ *and* $T_{Q}^{(N)}$ *are consistent against all fixed alternatives. The test based on* $T_{GM}^{(N)}$ *is consistent against fixed alternatives with strictly positive smallest canonical correlation, i.e. with* $\lambda_{n-1} > 0$.

(ii) *The asymptotic power functions, for local alternatives* P_L, *of these tests are strictly increasing in each* λ_t *for* $t = 1,\ldots,n-1$; *in particular these tests are asymptotically unbiased.*

PROOF. (i) Since eigenvalues are continuous functions of the matrix elements, the first result follows from Lemma 2.2.3.

(ii) The asymptotic testing problem is equivalent to the MANOVA testing problem that the expectation of Y is zero, where the columns of the matrix Y are independently and normally distributed. Therefore, (ii) follows from DAS GUPTA, ANDERSON and MUDHOLKAR (1964) (see also Theorem 1.1 in PERLMAN and OLKIN, 1980), provided the acceptance region of the test considered is convex in each column of Y when the remaining columns are held fixed. By asymptotic equivalence of $T_{CS}^{(N)}$, $T_{LR}^{(N)}$, $T_{P}^{(N)}$ and $T_{Q}^{(N)}$ it is sufficient to prove that the trace (sum of eigenvalues), largest eigenvalue and determinant (geometric mean of eigenvalues) of YY^T are convex functions in each column of Y. To this end let Y_1 and Y_2 be two arbitrary (fixed) matrices of size $(n-1) \times (m-1)$ which only differ in the j_0-th column and let $\bar{\alpha} = 1-\alpha$ for some $0 \leq \alpha \leq 1$. By Cauchy-Schwarz,

$$\textit{trace}\ (\alpha Y_1 + \bar{\alpha} Y_2)(\alpha Y_1 + \bar{\alpha} Y_2)^T = \alpha^2\, \textit{trace}\, Y_1 Y_1^T + 2\alpha\bar{\alpha}\, \textit{trace}\, Y_1 Y_2^T + \bar{\alpha}^2\, \textit{trace}\, Y_2 Y_2^T$$
$$\leq \alpha\, \textit{trace}\, Y_1 Y_1^T + \bar{\alpha}\, \textit{trace}\, Y_2 Y_2^T$$

and

$$\textit{largest eigenvalue}\,((\alpha Y_1 + \bar{\alpha} Y_2)(\alpha Y_1 + \bar{\alpha} Y_2)^T) =$$
$$\max_{u^T u = 1} u^T (\alpha Y_1 + \bar{\alpha} Y_2)(\alpha Y_1 + \bar{\alpha} Y_2)^T u \leq \max_{u^T u = 1} (\alpha u^T Y_1 Y_1^T u + \bar{\alpha} u^T Y_2 Y_2^T u)$$
$$\leq \alpha \max_{u^T u = 1} u^T Y_1 Y_1^T u + \bar{\alpha} \max_{u^T u = 1} u^T Y_2 Y_2^T u.$$

The desired convexity property of the trace and largest eigenvalue functions follows. Convexity of the determinant is proved as follows. First consider the case that Y_1 and Y_2 are square, then $\textit{det}\,(Y_1 Y_1^T) = \textit{det}^2\,(Y_1)$ and similar equalities for Y_2 and $\alpha Y_1 + \bar{\alpha} Y_2$. Expansion with respect to the j_0-th column yields $\textit{det}\,(\alpha Y_1 + \bar{\alpha} Y_2) = \alpha\, \textit{det}\,(Y_1) + \bar{\alpha}\, \textit{det}\,(Y_2)$. Hence

$$\textit{det}^2\,(\alpha Y_1 + \bar{\alpha} Y_2) \leq \alpha\, \textit{det}^2\,(Y_1) + \bar{\alpha}\, \textit{det}^2\,(Y_2). \tag{2.3.1}$$

In case that Y_1 and Y_2 are not square, write $\textit{det}\,(Y_1 Y_1^T)$ as the sum of squares of all determinants of $(n-1) \times (n-1)$ submatrices of Y_1 (cf. RAO, 1973, p. 32) and make similar expansions for $\textit{det}\,(Y_2 Y_2^T)$ and $\textit{det}\,((\alpha Y_1 + \bar{\alpha} Y_2)(\alpha Y_1 + \bar{\alpha} Y_2)^T)$. Using inequality (2.3.1) for each $(n-1) \times (n-1)$ subdeterminant involving the j_0-th column in the expansion of $\textit{det}\,((\alpha Y_1 + \bar{\alpha} Y_2)(\alpha Y_1 + \bar{\alpha} Y_2)^T)$ yields $\textit{det}\,((\alpha Y_1 + \bar{\alpha} Y_2)(\alpha Y_1 + \bar{\alpha} Y_2)^T) \leq \alpha\, \textit{det}\,(Y_1 Y_1^T) + \bar{\alpha}\, \textit{det}\,(Y_2 Y_2^T)$. This completes the proof. □

Although these tests are asymptotically unbiased, Example 4.4.3 shows that they are biased for finite samples.

The relative performance of two tests can be investigated by the *finite sample relative efficiency*: the ratio of the minimal sample sizes required to guarantee a fixed power against an alternative. Its computation is generally not feasible, but the asymptotic efficiencies of Bahadur and Pitman can be interpreted as approximations of the finite sample relative efficiency.

BAHADUR (1960) introduced the *approximate Bahadur efficiency* for two tests based on so called standard sequences of statistics. The sequence of statistics $\{T^{(N)}\}_{N=1}^{\infty}$ is said to be a *standard sequence* (for testing the hypothesis of independence) if the following three conditions are satisfied.

(i) There exists a continuous distribution function H such that under each P_0,

$$\lim_{N\to\infty} P\{T^{(N)} \le t\} = H(t) \qquad \text{for every } t.$$

(ii) There exists a constant a, $0 < a < \infty$ such that

$$\log(1 - H(t)) = -\tfrac{1}{2}at^2(1 + o(1)) \qquad \text{as } t \to \infty.$$

(iii) There exists a function b defined on all probability tables such that under each P_A

$$\lim_{N\to\infty} P\{|N^{-\frac{1}{2}}T^{(N)} - b(P')| > \varepsilon\} = 0 \qquad \text{for every } \varepsilon > 0$$

with $0 < b(P') < \infty$.

The present conditions (i) and (iii) are stronger than Bahadur's conditions since arbitrary convergent sequences of underlying probability tables are considered here, i.e. uniform convergence is required in (i) and (iii).

The *approximate slope* of a standard sequence is defined as $s(P') = ab^2(P')$ and the *approximate Bahadur efficiency* of the standard sequence $\{T_1^{(N)}\}_{N=1}^{\infty}$ with respect to the standard sequence $\{T_2^{(N)}\}_{N=1}^{\infty}$ is defined as the ratio of slopes

$$ABE(T_1,T_2;P') = s_1(P')/s_2(P').$$

A value larger than unity indicates that T_1 is superior to T_2 at the (limiting) alternative P'.

THEOREM 2.3.3. *The sequences of statistics* $\{T_{CS}^{(N)\frac{1}{2}}\}_{N=1}^{\infty}$, $\{T_{LR}^{(N)\frac{1}{2}}\}_{N=1}^{\infty}$, $\{T_{CC}^{(N)\frac{1}{2}}\}_{N=1}^{\infty}$ *and* $\{T_{P}^{(N)\frac{1}{2}}\}_{N=1}^{\infty}$ *are standard sequences with approximate Bahadur slopes*

$$s_{CS}(P') = \sum_{t=1}^{n-1} \lambda_t^2,$$

$$s_{LR}(P') = 2 \sum_{i=1}^{n} \sum_{j=1}^{m} p'_{ij} \log(p'_{ij}/(r_i c_j)),$$

$$s_{CC}(P') = \lambda_1^2,$$

$$s_{P}(P') = \sum_{t=1}^{n-1} \lambda_t^2/(1 + \lambda_t^2).$$

Furthermore, if attention is restricted to alternatives P' *with* $\lambda_{n-1} > 0$, *then* $\{T_{GM}^{(N)\frac{1}{2}}\}_{N=1}^{\infty}$ *is a standard sequence with approximate slope*

$$s_{GM}(P') = (n-1)\Big(\prod_{t=1}^{n-1} \lambda_t\Big)^{2/(n-1)},$$

and if attention is restricted to alternatives P' *with* $\lambda_1 < 1$, *then* $\{T_Q^{(N)^{\frac{1}{2}}}\}_{N=1}^{\infty}$ *is a standard sequence with approximate slope*

$$s_Q(P') = \sum_{t=1}^{n-1} \lambda_t^2/(1-\lambda_t^2).$$

PROOF. The three conditions for standard sequences are verified as follows.

(i) Immediately from Corollary 2.3.1.

(iii) It follows from (2.2.8) that under fixed alternatives P_A, $\hat{\lambda}_t^{(N)}$ converges in probability to λ_t for $t = 1,\ldots,n-1$. Application of the continuous mapping theorem produces the desired result.

(ii) First consider a slightly more general set-up. Let $T^{(N)}$ be an univariate statistic of the form $T^{(N)} = \tau(N\hat{\lambda}_1^{(N)^2},\ldots,N\hat{\lambda}_{n-1}^{(N)^2})$, where τ is a continuous positive function such that $\tau(c\ell_1,\ldots,c\ell_{n-1}) = c\tau(\ell_1,\ldots,\ell_{n-1})$ for all $c \geq 0$ and all $\ell_1 \geq \ldots \geq \ell_{n-1} \geq 0$. It follows from Theorem 2.2.5 that, under P_0, $T^{(N)}$ converges in distribution to $\tau(\ell_1,\ldots,\ell_{n-1})$, where $\ell_1 \geq \ldots \geq \ell_{n-1} \geq 0$ are the eigenvalues of a $W_{n-1}(m-1,I)$ distributed matrix. Let H denote the limiting null-distribution function of $T^{(N)^{\frac{1}{2}}}$. Then by MUIRHEAD (1982, p. 107),

$$1 - H(t) =$$

$$K_n \int\cdots\int_{\substack{\ell_1 \geq \ldots \geq \ell_{n-1} \geq 0 \\ \tau(\ell_1,\ldots,\ell_{n-1}) \geq t^2}} \exp\Big(-\tfrac{1}{2}\sum_{i=1}^{n-1} \ell_i\Big)\Big(\prod_{i=1}^{n-1} \ell_i\Big)^{(m-n-1)/2} \prod_{i<j} (\ell_i - \ell_j)\, d\ell_1 \ldots d\ell_{n-1}.$$

Changing the variables, $t^2 u_i = \ell_i$ for $i = 1,\ldots,n-1$, produces

$$1 - H(t) = K_n t^{(n-1)(m-1)} \int\cdots\int_{\substack{u_1 \geq \ldots \geq u_{n-1} \geq 0 \\ \tau(u_1,\ldots,u_{n-1}) \geq 1}} \exp\Big(-\tfrac{1}{2}t^2\sum_{i=1}^{n-1} u_i\Big)\Big(\prod_{i=1}^{n-1} u_i\Big)^{(m-n-1)/2} \times$$

$$\times \prod_{i<j} (u_i - u_j)\, du_1 \ldots du_{n-1}.$$

By Laplace's method of asymptotic expansions of integrals (cf. ERDÉLYI, 1956, pp. 36-39) it follows that

$$\lim_{t\to\infty} -t^{-2} \log(1 - H(t)) = \tfrac{1}{2}a,$$

where a is the minimum of $\Sigma_1^{n-1} u_i$ on the integration area $B = \{u \in \mathbb{R}^{n-1} : u_1 \geq u_2 \geq \ldots \geq u_{n-1} \geq 0,\ \tau(u_1,\ldots,u_{n-1}) \geq 1\}$.

The chi-square, canonical correlation and geometric mean statistic

fit this set-up. It can be verified that the minima of $\Sigma_1^{n-1} u_i$ on the corresponding integration area's are $a_{CS} = 1$, $a_{CC} = 1$ and $a_{GM} = n-1$ respectively. Therefore, the square roots of these statistics satisfy condition (ii) of standard sequences with these constants a_{CS}, a_{CC} and a_{GM}. Since $T_{LR}^{(N)}$, $T_P^{(N)}$ and $T_Q^{(N)}$ have the same limiting null-distribution as $T_{CS}^{(N)}$, the square roots of these statistics also satisfy this condition with constants $a_{LR} = a_P = a_Q = 1$. □

Note that tests based on $T^{(N)^{\frac{1}{2}}}$ are equivalent to tests based on $T^{(N)}$ and therefore produce the same relative efficiencies. The following inequalities are easily verified for each alternative P',

$$s_{CC}(P') \leq s_{CS}(P') \leq s_Q(P'),$$
$$s_{GM}(P') \leq s_{CS}(P') \leq s_Q(P'),$$
$$s_P(P') \leq s_{CS}(P') \leq s_Q(P').$$

This suggests that the test based on $T_Q^{(N)}$ is superior to the other four and that the chi-square test is second best. The slope of the likelihood ratio test cannot be ordered with respect to the other slopes for all alternatives. These orderings of tests should be interpreted with care because the approximate Bahadur efficiency does not always give a good approximation to the finite sample relative efficiency. In many cases Pitman efficiencies give better approximations.

The *Pitman efficiency* is based on comparisons of powers of tests at local alternatives. Using results of ROTHE (1981) it can be verified that *the Pitman efficiencies between tests based on* $T_{CS}^{(N)}$, $T_{LR}^{(N)}$, $T_P^{(N)}$ *and* $T_Q^{(N)}$ *are all unity*, which means that Pitman efficiencies do not discriminate between these tests. Furthermore, in a neighbourhood of Π_0 the distributions of the six statistics converge uniformly in the alternative parameter and, therefore, it can be shown using the Lemma and the Theorem in WIEAND (1976) that *the limit of the approximate Bahadur efficiency* (*as the alternative approaches the hypothesis*) *equals the limit of the Pitman efficiency* (*as the size* α *of the test tends to zero*).

Of interest is the Pitman efficiency between the chi-square (or likelihood ratio) test and the canonical correlation test. Using the arguments above it can be shown that

$$\lim_{\alpha \downarrow 0} PE(T_{CS}, T_{CC}; P') = \lim_{\theta \to 0} ABE(T_{CS}, T_{CC}; rc^\top + \theta(P' - rc^\top)) = \sum_{t=1}^{n-1} \lambda_t^2/\lambda_1^2$$

(where PE is short for Pitman efficiency), suggesting that the chi-square test always dominates the canonical correlation test. The computation of the exact Pitman efficiency between these two tests is rather intractable. Even its numerical computation is quite a difficult problem, but in the case n = m = 3 it can be carried out. Such numerical computations indicate that the canonical correlation test is slightly better than the chi-square test for alternatives with λ_2 very close to zero, but that the chi-square test beats the canonical correlation test rapidly as the alternative moves away from the axis $\lambda_2 = 0$. Furthermore, it is noted in HABERMAN (1981) that tests based on $T_{LR}^{(N)}$, $T_{CS}^{(N)}$, $T_{P}^{(N)}$ and $T_{Q}^{(N)}$ are asymptotically locally best (maximize the derivative of the asymptotic power function at the null-hypothesis).

The considerations given above recommend tests based on $T_{CS}^{(N)}$, $T_{LR}^{(N)}$ or $T_{Q}^{(N)}$.

The tests above are based on canonical correlations and are tests against omnibus alternatives. Intuitively, one might think that tests based on canonical variables are sensitive to certain *forms* of dependence. In view of the results of Section 3.3, certain (normalized) linear combinations of the row or column category scores might produce tests with a good performance against certain alternatives.

Let a be any n-dimensional vector which is not proportional to e and consider

$$(2.3.2) \qquad T_{R,t}^{(N)} = (a^{\top}\hat{R}^{(N)}\hat{\varphi}_t^{(N)})^2/(a^{\top}\hat{R}^{(N)}a - (a^{\top}\hat{r}^{(N)})^2).$$

This statistic can be interpreted as the square of the sample correlation between the t-th sample canonical variable $\hat{\varphi}_t^{(N)}(X)$ and the transformed row variable a(X), where the function a is defined by $a(i) = a_i$ for $i = 1,\ldots,n$. Similarly, consider for any m-dimensional vector b, which is not proportional to e, the statistic

$$(2.3.3) \qquad T_{C,t}^{(N)} = (b^{\top}\hat{C}^{(N)}\hat{\psi}_t^{(N)})^2/(b^{\top}\hat{C}^{(N)}b - (b^{\top}\hat{c}^{(N)})^2).$$

The following proposition shows that the limiting distributions of these statistics under the hypothesis is a beta distribution.

PROPOSITION 2.3.4. *Under the hypothesis* P_0,

$$(2.3.4) \qquad T_{R,t}^{(N)} \to_D B(\tfrac{1}{2},\tfrac{1}{2}(n-1)) \qquad \text{as } N \to \infty,$$

(2.3.5) $T_{C,t}^{(N)} \to_D B(\frac{1}{2},\frac{1}{2}(m-1))$ *as* $N \to \infty$,

for $t = 1,\ldots,n-1$. *Moreover, under fixed alternatives* P_A,

(2.3.6) $T_{R,t}^{(N)} \to_P \tau_{R,t}$, $T_{C,t}^{(N)} \to_P \tau_{C,t}$ *as* $N \to \infty$,

for $t = 1,\ldots,n-1$, *where*

$$\tau_{R,t} = (a^T R\varphi_t)^2/(a^T Ra - (a^T r)^2),$$

$$\tau_{C,t} = (b^T C\psi_t)^2/(b^T Cb - (b^T c)^2),$$

provided the t*-th canonical correlation* λ_t *is simple.*

PROOF. Using the same notation as in the proof of Theorem 2.2.5, it follows under P_0 that

$$T_{R,t}^{(N)} \to_D (a^T R\Phi h_t)^2/a^T(R - rr^T)a \sim (\alpha^T h_t^*)^2 \quad \text{as } N \to \infty$$

where $\alpha^T = a^T R\Phi/(a^T(R - rr^T)a)^{\frac{1}{2}}$. Note that $\alpha^T\alpha = 1$ since $R\Phi\Phi^T R = R - rr^T$. Since h_t^* is uniformly distributed on the unit sphere in $\mathbb{R}^{n-1}$, the result (2.3.4) follows from MUIRHEAD (1982, p. 39). Similarly, (2.3.5) follows. Furthermore, (2.3.6) is a trivial consequence of (2.2.9). □

Thus, unfortunately, tests based on canonical variables are generally not consistent. Combining them with tests based on canonical correlations produces consistent tests but with a rather bad performance.

CHAPTER 3

ORDERING PROPERTIES IN CORRESPONDENCE ANALYSIS

3.1. TOTAL POSITIVITY

This section summarizes results of the theory of total positivity which will be needed in the present and following chapters. Important references on this subject are GANTMACHER and KREIN (1950) and KARLIN (1968).

Let X and Y be subsets of $\mathbb{R}$ and let K be a real valued function (kernel) defined on $X \times Y$. In the case that $X = \{1,\ldots,n\}$ and $Y = \{1,\ldots,m\}$, the kernel K can be considered as a $n \times m$ matrix K with elements $k_{ij} = K(i,j)$ for $i = 1,\ldots,n$ and $j = 1,\ldots,m$.

DEFINITION 3.1.1. The kernel K is called *totally positive of order* s (abbreviated TP_s) on $X \times Y$ if for every $t = 1,\ldots,s$, all $x_1 < x_2 < \ldots < x_t$ and all $y_1 < y_2 < \ldots < y_t$ ($x_i \in X$, $y_i \in Y$ for $i = 1,\ldots,t$) the determinant

$$(3.1.1) \qquad K\begin{pmatrix} x_1,x_2,\ldots,x_t \\ y_1,y_2,\ldots,y_t \end{pmatrix} = \begin{vmatrix} K(x_1,y_1) & K(x_1,y_2)\ldots K(x_1,y_t) \\ K(x_2,y_1) & K(x_2,y_2)\ldots K(x_2,y_t) \\ \vdots & \vdots \\ K(x_t,y_1) & K(x_t,y_2)\ldots K(x_t,y_t) \end{vmatrix}$$

is positive. When the determinant (3.1.1) is even strictly positive for every $t = 1,\ldots,s$ and all $x_1 < \ldots < x_t$, $y_1 < \ldots < y_t$, then K is said to be *strictly totally positive of order* s (STP_s).

This definition has the following two obvious consequences.

PROPOSITION 3.1.1. *Let* K *be* $(S)TP_s$ *on* $X \times Y$, *then*

(i) $f(x)g(y)K(x,y)$ *is* $(S)TP_s$ *on* $X \times Y$ *whenever the functions* f *and* g *are (strictly) positive on* X *and* Y *respectively,*

(ii) $K(f(u),g(v))$ *is* $(S)TP_s$ *on* $f^{-}(X) \times g^{-}(Y)$ *when the functions* f *and* g *are both (strictly) increasing or both (strictly) decreasing on* $f^{-}(X)$ *and*

$g^{-}(Y)$ *respectively.* (*Here* $f^{-}(X) = \{u \in \mathbb{R} : f(u) \in X\}$ *and* $g^{-}(Y)$ *is similarly defined*).

PROOF. Trivial. □

The following important result in the theory of total positivity is often called the *basic composition formula.*

LEMMA 3.1.2. *Let* σ *be a sigma-finite measure and suppose that the integral*

$$M(x,z) = \int_Y K(x,y)L(y,z)d\sigma(y) \tag{3.1.2}$$

converges absolutely, then

$$M\begin{pmatrix} x_1,\ldots,x_t \\ z_1,\ldots,z_t \end{pmatrix} = \int\ldots\int_{y_1<\ldots<y_t} K\begin{pmatrix} x_1,\ldots,x_t \\ y_1,\ldots,y_t \end{pmatrix} L\begin{pmatrix} y_1,\ldots,y_t \\ z_1,\ldots,z_t \end{pmatrix} d\sigma(y_t)\ldots d\sigma(y_1).$$

PROOF. See KARLIN (1968, p. 17). □

COROLLARY 3.1.3. *Let* K *be* TP_s *on* $X \times Y$ *and let* L *be* TP_r *on* $Y \times Z$, *then* M *defined by* (3.1.2) *is* $TP_{\min(s,r)}$ *on* $X \times Z$. *Moreover, when* L *is* STP_r *and* K *is* TP_s *and satisfies the rank condition that for all* $t = 1,\ldots,s$ *and all* $x_1 < \ldots < x_t$ $(x_i \in X;\ i = 1,\ldots,t)$ *there exist sets* $B_1 < B_2 < \ldots < B_t$ *with* $\sigma(B_j) > 0$, $B_j \subset Y$ *for* $j = 1,\ldots,t$ *such that det* $|(K(x_i,B_j))| \neq 0$, *where*

$$K(x_i,B_j) = \int_{B_j} K(x_i,y)d\sigma(y) \qquad \text{for } i,j = 1,\ldots,t, \tag{3.1.3}$$

then M *is actually* $STP_{\min(s,r)}$.

Using Proposition 3.1.1 and Corollary 3.1.3 many examples of (S)TP kernels can be constructed from a few basic examples. In Section 3.4 (strict) total positivity of well-known bivariate probability densities is proved starting from the following numerous examples.

EXAMPLE 3.1.1. The function

$$K(x,y) = \begin{cases} 1 & \text{if } x \leq y \\ 0 & \text{if } x > y \end{cases}$$

is TP_∞ on $\mathbb{R}^2$ (cf. KARLIN, 1968, p. 16).

EXAMPLE 3.1.2. The function $K(x,y) = e^{xy}$ is STP_∞ on $\mathbb{R}^2$ (cf. KARLIN, 1968, p. 15).

EXAMPLE 3.1.3. The function $K(x,y) = (x+y)^{-\alpha}$, $\alpha > 0$ is STP_∞ for $0 < x,y < \infty$. To verify this consider for $s = 1,2,\ldots$ the "derived" determinant

$$K^*\begin{pmatrix} x_1,x_2,\ldots,x_s \\ y,y,\ldots,y \end{pmatrix} \equiv \left|\left(\frac{\partial^{j-1}}{\partial y^{j-1}} K(x_i,y)\right)\right|,$$

which in the present example for $0 < x_1 < x_2 < \ldots < x_s$ and $y > 0$ equals

$$K^*\begin{pmatrix} x_1,x_2,\ldots,x_s \\ y,y,\ldots,y \end{pmatrix} = \begin{vmatrix} \frac{1}{(x_1+y)^{\alpha}} & \frac{-\alpha}{(x_1+y)^{\alpha+1}} & \cdots & \frac{(-1)^{s-1}\alpha(\alpha+1)\ldots(\alpha+s-2)}{(x_1+y)^{\alpha+s-1}} \\ \vdots & \vdots & & \vdots \\ \frac{1}{(x_s+y)^{\alpha}} & \frac{-\alpha}{(x_s+y)^{\alpha+1}} & \cdots & \frac{(-1)^{s-1}\alpha(\alpha+1)\ldots(\alpha+s-2)}{(x_s+y)^{\alpha+s-1}} \end{vmatrix} =$$

$$\frac{(-1)^{s(s-1)}\alpha^{s-1}(\alpha+1)^{s-2}\ldots(\alpha+s-2)}{\prod_{t=1}^{s}(x_t+y)^{\alpha+s-1}} \begin{vmatrix} 1 & (x_1+y) & \ldots & (x_1+y)^{s-2} & (x_1+y)^{s-1} \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & (x_s+y) & \ldots & (x_s+y)^{s-2} & (x_s+y)^{s-1} \end{vmatrix} > 0.$$

The last determinant is the Vandermonde determinant. By KARLIN (1968, pp. 49-50), K is STP_∞. (For $\alpha = 1$ this example is also given in KARLIN, 1968, p. 149.)

EXAMPLE 3.1.4. The function $K(x,y) = \Gamma(x+y+1)$ is STP_∞ for $0 \le x,y < \infty$. By definition $\Gamma(x+y+1) = \int_0^\infty e^{x\log(z)} e^{y\log(z)} e^{-z} dz$ and hence the result follows from Example 3.1.2, Proposition 3.1.1 and Corollary 3.1.3.

EXAMPLE 3.1.5. The function

$$K(x,y) = \begin{cases} (x-y)^m & \text{if } x > y \\ 0 & \text{if } x \le y \end{cases}, \quad m \in \mathbb{N}$$

is TP_∞ on $\mathbb{R}^2$ (cf. KARLIN and STUDDEN, 1966, p. 17).

EXAMPLE 3.1.6. The function

$$K(x,y) = \begin{cases} 1/(x-y)! & \text{if } x \ge y \\ 0 & \text{if } x < y \end{cases}$$

is TP_∞ on $\mathbb{Z}^2$. (Follows from KARLIN (1968, p. 137) and Proposition 3.1.1.)

EXAMPLE 3.1.7. The function

$$K(x,y) = \begin{cases} \binom{n}{x-y} & \text{if } x \ge y \text{ and } x-y \le n \\ 0 & \text{if } x < y \text{ or } x-y > n \end{cases}$$

is TP_∞ on $\mathbb{Z}^2$ (cf. KARLIN, 1968, p. 44).

EXAMPLE 3.1.8. Let $\{\phi_t(x)\}_{t=0}^{\infty}$ be an orthogonal polynomial system (where ϕ_t is of exact degree t) with respect to a measure σ on $[a,\infty)$, where $a > -\infty$. Assume that $\phi_t(0) > 0$ for $t = 0,1,\ldots$. Then $K(x,t) = \phi_t(-x)$ is STP_∞ for $a \leq x < \infty$ and $t = 0,1,\ldots$ (cf. KARLIN and McGREGOR, 1959, p. 1115 or KARLIN, 1968, p. 438).

EXAMPLE 3.1.9. An important class of (S)TP functions are of the form $K(x,y) = f(x-y)$ for $x,y \in \mathbb{R}$. Functions f for which $f(x-y)$ is $(S)TP_s$ are called (*strictly*) *Pólya frequency functions of order* s (abbreviated $(S)PF_s$). It can be proved that f is PF_2 iff f is positive and log concave on $\mathbb{R}$.

The interest of totally positive functions is largly due to the *variation diminishing property*. Let f be a function defined on a subset Z of $\mathbb{R}$. The *minimum number of sign changes* of f on Z is defined by

$$S^-(f) = \sup S^-[f(z_1),f(z_2),\ldots,f(z_t)]$$

where the supremum is over all sets $z_1 < z_2 < \ldots < z_t$ $(z_i \in Z;\ i = 1,\ldots,t)$, t is arbitrary but finite, and $S^-[u_1,u_2,\ldots,u_t]$ is the number of sign changes of the sequence $u_1,u_2,\ldots,u_t$ discarding zero terms. The *maximum number of sign changes* of f on Z is defined by

$$S^+(f) = \sup S^+[f(z_1),f(z_2),\ldots,f(z_t)]$$

where $S^+[u_1,u_2,\ldots,u_t]$ denotes the maximum number of sign changes of the sequence, assigning arbitrary signs to zero terms. The *relevant number of sign changes* of f with respect to a sigma-finite measure σ on Z is defined by

$$S_\sigma^-(f) = \inf S^-(h)$$

where the infimum is taken over all functions h which σ-a.e. equal f.

LEMMA 3.1.4. (Variation Diminishing Property). *Let* K *be a kernel defined on* $X \times Y$, *where* X *and* Y *are Borel-measurable sets, and let* σ *be a sigma-finite measure on* X. *Suppose that* $\int_X K(x,y)d\sigma(x)$ *exists and is finite for every* $y \in Y$. *Consider the transformation*

$$(3.1.4) \qquad g(y) = \int_X K(x,y)f(x)d\sigma(x) \qquad \text{for } y \in Y$$

where f *is a bounded Borel-measurable function.*

(i) *If* K *is* TP_s, *then*

$$(3.1.5) \qquad S_\sigma^-(f) \le s-1 \Rightarrow S^-(g) \le S_\sigma^-(f).$$

Moreover, if $S^-(g) = S_\sigma^-(f) \le s-1$, *then* g *and* f *exhibit the same arrangement of signs.*

(ii) *If* K *is* STP_s *and* $f \neq 0$ σ*-a.e., then*

$$(3.1.6) \qquad S_\sigma^-(f) \le s-1 \Rightarrow S^+(g) \le S_\sigma^-(f).$$

(iii) *Under slight non-degeneracy conditions, converses of* (i) *and* (ii) *also hold.*

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

Another related property is the following lemma on the number of sign changes of eigenfunctions of TP kernels. This result is formulated for the matrix case, because its general version still seems to be unproved. (A related result for continuous (symmetric) kernels is given in GANTMACHER and KREIN, 1950, pp. 268-269; see also KARLIN, 1968, pp. 35-38.)

Let $u_1,\ldots,u_s$ denote the eigenvectors of the $n \times n$ matrix K corresponding to the s "largest" eigenvalues $|\mu_1| \ge |\mu_2| \ge \ldots \ge |\mu_s|$ and let U denote the $n \times s$ matrix with $u_1,\ldots,u_s$ as columns.

LEMMA 3.1.5. *Suppose that the* $n \times n$ *matrix* K *is* TP_s *and that there exists an integer* $q \ge 1$ *such that the* q*-th iterate* K^q *is* STP_s. *Then the* s *largest eigenvalues of* K *are strictly positive and distinct, i.e.* $\mu_1 > \mu_2 > \ldots > \mu_s > |\mu_{s+1}|$. *Furthermore, the signs of the eigenvectors* $u_1,\ldots,u_s$ *can be chosen such that*

$$(3.1.7) \qquad U\begin{pmatrix} i_1, i_2, \ldots, i_t \\ 1, 2, \ldots, t \end{pmatrix} > 0$$

for all $1 \le i_1 < i_2 < \ldots < i_t \le n$ *and every* $t = 1,\ldots,s$ (*i.e. the columns of* U *form a so called complete Tchebycheff system*) *and hence for arbitrary real numbers* $a_k, a_{k+1}, \ldots, a_\ell$ $(1 \le k \le \ell \le s, \sum_{t=k}^{\ell} a_t^2 > 0)$ *the number of changes of sign of the linear combination*

$$u = \sum_{t=k}^{\ell} a_t u_t$$

satisfies $k-1 \le S^-(u) \le S^+(u) \le \ell-1$.

PROOF. The lemma is a weaker version of a result in GANTMACHER and KREIN (1950, p. 349). □

Furthermore, the following lemma will be used.

LEMMA 3.1.6. *The set of all* $n \times m$ STP_s *matrices is dense in the set of all* $n \times m$ TP_s *matrices.*

PROOF. In KARLIN (1968, p. 88) the proof is given for matrices of size $s \times m$, but the same method applies for matrices of size $n \times m$. □

3.2. GENERALIZATIONS OF REGRESSION DEPENDENCE

This section first discusses generalizations of regression dependence in the context of arbitrary bivariate distributions. These generalizations are basic to the notion *order dependence* which is introduced at the end of this section.

Let (X,Y) have an arbitrary bivariate distribution F which has density p with respect to a product measure $\sigma_r \times \sigma_c$ on $\mathcal{X} \times \mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are subsets of $\mathbb{R}$. Recall that $F_{r|c}(\cdot|y)$ denotes the distribution function of $X|Y = y$ for $y \in \mathcal{Y}$.

DEFINITION 3.2.1. The pair of variables (X,Y) is said to be *(S)TP_s-dependent* when a suitable version of their joint density p is (S)TP_s on $\mathcal{X} \times \mathcal{Y}$.

TP_2-dependence is sometimes called positive likelihood ratio dependence (cf. LEHMANN, 1966 and TONG, 1980, p. 79). The interpretation of its definition

$$x_1 < x_2,\ y_1 < y_2 \quad (x_1,x_2 \in \mathcal{X};\ y_1,y_2 \in \mathcal{Y}) \Rightarrow p(x_1,y_1)p(x_2,y_2) \geq p(x_1,y_2)p(x_2,y_1)$$

obviously indicates a strong form of positive dependence between X and Y. Even stronger forms of dependence are obtained when p is TP_s-dependent for $s > 2$. TP_2-dependence means that the family of conditional distributions of $X|Y = y$ has the monotone likelihood ratio property. A weaker form of dependence arises when this family is only stochastically increasing, i.e.

$$(3.2.1) \qquad y_1 < y_2 \quad (y_1,y_2 \in \mathcal{Y}) \Rightarrow F_{r|c}(x|y_1) \geq F_{r|c}(x|y_2) \quad \text{for all } x \in \mathcal{X}.$$

In this case LEHMANN (1966) calls X *regression dependent on* Y. It is seen

from (3.2.1) that the mass of the conditional distributions of $X|Y = y$ concentrates on higher values in X as y increases. Condition (3.2.1) says that the differences $F_{r|c}(x|y_1) - F_{r|c}(x|y_2)$ are positive whenever $y_1 < y_2$. Assuming further TP properties for these differences yields the following generalization of regression dependence. The actual motivation for this generalization is Theorem 3.2.3 below. Let $X \setminus \sup\{X\}$ denote the set X with its supremum deleted.

DEFINITION 3.2.2. The pair of variables (X,Y) is said to be *column regression dependent of order* s on $X \times Y$ (abbreviated CR_s-dependent) if there exists a suitable version of the conditional distribution of $X|Y$ such that for every $t = 2,\ldots,s+1$ and all $x_1 < x_2 < \ldots < x_{t-1}$ ($x_i \in X$, $i = 1,\ldots,t-1$) and all $y_1 < y_2 < \ldots < y_t$ ($y_j \in Y$, $j = 1,\ldots,t$) the determinant

$$(3.2.2) \qquad \begin{vmatrix} F_{r|c}(x_1|y_1) & \ldots & F_{r|c}(x_1|y_t) \\ \vdots & & \vdots \\ F_{r|c}(x_{t-1}|y_1) & \ldots & F_{r|c}(x_{t-1}|y_t) \\ 1 & \ldots & 1 \end{vmatrix}$$

is positive. When the determinant (3.2.2) is even strictly positive for every $t = 2,\ldots,s+1$ and all $x_1 < \ldots < x_{t-1}$ ($x_i \in X \setminus \sup\{X\}$, $i=1,\ldots,t-1$) and all $y_1 < \ldots < y_t$ ($y_j \in Y$, $j = 1,\ldots,t$), then (X,Y) is said to be *strictly column regression dependent of order* s (SCR_s-dependent). Similarly, (X,Y) is said to be (*strictly*) *row regression dependent of order* s (abbreviated (S)RR_s-dependent) when (Y,X) is (strictly) column regression dependent of order s (i.e. (3.2.2) holds when the roles of rows and columns and of x's and y's are interchanged). Furthermore, (X,Y) is said to be (*strictly*) *double regression dependent of order* s (denoted (S)DR_s-dependent) if (X,Y) is both (S)RR_s and (S)CR_s-dependent.

In the sequel we shall not bother about the choice of different versions of conditional distributions.

CR_1-dependence corresponds with the usual definition (3.2.1) of regression dependence of X on Y. The family of conditional distributions of $X|Y = y$ which arises when (X,Y) is CR_2-dependent is in fact the *invariant convexity preserving family* considered by VAN ZWET (1968).

PROPOSITION 3.2.1. *(S)TP_{s+1}-dependence $\Rightarrow$ (S)DR_s-dependence.*

PROOF. Let (X,Y) be TP_{s+1}-dependent, then their joint density p and hence

the conditional density of $X|Y = y$ is TP_{s+1} on $X \times Y$. Application of Corollary 3.1.3 to this conditional density and the triangular kernel of Example 3.1.1 yields that $F_{r|c}$ is TP_{s+1}. Hence by Definition 3.1.1 with $x_t = \sup\{X\}$, (X,Y) is CR_s-dependent. RR_s-dependence follows similarly. In the case that (X,Y) is STP_{s+1}-dependent the same arguments apply because the triangular kernel satisfies the rank condition of Corollary 3.1.3. □

In practice the $(S)TP_{s+1}$ property is often easier verified than the $(S)RR_s$ or $(S)CR_s$ property. The following proposition shows that in the two most common cases $(S)RR_s$ and $(S)CR_s$-dependence actually involve a $(S)TP_s$ property. The first part of this result is, for the case s = 2, also proved by VAN ZWET (1968).

PROPOSITION 3.2.2. (i) *Let* Y *be an interval and suppose that* $q_c(x,y) = -\partial F_{r|c}(x|y)/\partial y$ *exists, then* (X,Y) *is* $(S)CR_s$*-dependent iff* $q_c(x,y)$ *is* $(S)TP_s$ *on* $X \setminus \sup\{X\} \times Y$.

(ii) *Let* $Y = \{1,\ldots,m\}$ *and define* $q_c(x,y) = F_{r|c}(x|y) - F_{r|c}(x|y+1)$ *for* $x \in X$ *and* $y = 1,\ldots,m-1$, *then* (X,Y) *is* $(S)CR_s$*-dependent iff* $q_c(x,y)$ *is* $(S)TP_s$ *on* $X \setminus \sup\{X\} \times \{1,\ldots,m-1\}$.

PROOF. (i) Let $x_1 < \ldots < x_{t-1}$ $(x_i \in X \setminus \sup\{X\})$ and $y_1 < \ldots < y_t$ $(y_j \in Y)$ be arbitrary. Consider the function $g(\eta)$ defined for $y_1 \le \eta \le y_2$ by

$$g(\eta) \equiv \begin{vmatrix} F_{r|c}(x_1|\eta) & F_{r|c}(x_1|y_2) & \ldots & F_{r|c}(x_1|y_t) \\ \vdots & \vdots & & \vdots \\ F_{r|c}(x_{t-1}|\eta) & F_{r|c}(x_{t-1}|y_2) & \ldots & F_{r|c}(x_{t-1}|y_t) \\ 1 & 1 & \ldots & 1 \end{vmatrix}$$

$$= \sum_{i=1}^{t-1} F_{r|c}(x_i|\eta)M_i + M_t$$

where M_i for $i = 1,\ldots,t$ are signed minors. By the mean value theorem there exists $y_1 < \eta_1 < y_2$ such that $g'(\eta_1)(y_2 - y_1) = g(y_2) - g(y_1)$. Since $g(y_2) = 0$, it follows that

$$g(y_1) = \sum_{i=1}^{t-1} q_c(x_i,\eta_1)M_i(y_2 - y_1)$$

$$= \begin{vmatrix} q_c(x_1,\eta_1) & F_{r|c}(x_1|y_2) & \ldots & F_{r|c}(x_1|y_t) \\ \vdots & \vdots & & \vdots \\ q_c(x_{t-1},\eta_1) & F_{r|c}(x_{t-1}|y_2) & \ldots & F_{r|c}(x_{t-1}|y_t) \\ 0 & 1 & \ldots & 1 \end{vmatrix} (y_2 - y_1).$$

Repeated application of the mean value theorem yields the existence of $y_1 < \eta_1 < y_2 < \eta_2 < \ldots < \eta_{t-1} < y_t$ such that

$$\text{sign} \begin{vmatrix} F_{r|c}(x_1|y_1) & \cdots & F_{r|c}(x_1|y_t) \\ \vdots & & \vdots \\ F_{r|c}(x_{t-1}|y_1) & \cdots & F_{r|c}(x_{t-1}|y_t) \\ 1 & \cdots & 1 \end{vmatrix} =$$

$$= \text{sign} \begin{vmatrix} q_c(x_1,\eta_1) & \cdots & q_c(x_1,\eta_{t-1}) & F_{r|c}(x_1|y_t) \\ \vdots & & \vdots & \vdots \\ q_c(x_{t-1},\eta_1) & \cdots & q_c(x_{t-1},\eta_{t-1}) & F_{r|c}(x_{t-1}|y_t) \\ 0 & \cdots & 0 & 1 \end{vmatrix}$$

which proves the first part of the proposition.

(ii) It is easily verified that for $x_1 < \ldots < x_{t-1}$ ($x_i \in X \setminus \sup\{X\}$) and $y = 1,\ldots,m-t+2$

$$(3.2.3) \qquad q_c\begin{pmatrix} x_1,x_2,\ldots,x_{t-1} \\ y,y+1,\ldots,y+t-2 \end{pmatrix} = \begin{vmatrix} F_{r|c}(x_1|y) & \cdots & F_{r|c}(x_1|y+t-1) \\ \vdots & & \vdots \\ F_{r|c}(x_{t-1}|y) & \cdots & F_{r|c}(x_{t-1}|y+t-1) \\ 1 & \cdots & 1 \end{vmatrix}.$$

By KARLIN (1968, p. 60), $q_c\begin{pmatrix} x_1,x_2,\ldots,x_{t-1} \\ y,y+1,\ldots,y+t-2 \end{pmatrix} > 0$ for $y = 1,\ldots,m-t+1$ implies $q_c\begin{pmatrix} x_1,x_2,\ldots,x_{t-1} \\ y_1,y_2,\ldots,y_{t-1} \end{pmatrix} > 0$ for $1 \le y_1 < y_2 < \ldots < y_{t-1} \le m-1$. Similarly, strict positive sign of the determinant on the right hand side of (3.2.3) for all t-1 consecutive elements of y implies strict positive sign of this determinant for all ordered t-1 tuples. It now easily follows that (X,Y) is SCR_s-dependent iff q_c is STP_s. The restriction that all determinants are *strictly* positive can be dropped by appealing to Lemma 3.1.6 and continuity. □

Consider the conditional expectations

$$(3.2.4) \qquad g(y) = E(f(X)|Y=y) \qquad \text{for } y \in y.$$

It follows from the variation diminishing property that the function g changes sign at most t times when the number of sign changes of f equals t and (X,Y) is TP_s-dependent with $s \ge t+1$. Although this result is useful in particular situations, the following related weaker result gives a better

intuitive grasp of the notions TP_{s+1} and CR_s-dependence. It is well-known that a stochastically increasing family preserves monotonicity of functions. Thus when (X,Y) is CR_1-dependent, the function g is increasing whenever f is. A stronger form of dependence is needed if one also wants the regression of convex functions of X on Y = y to be monotone of order 2, i.e. first decreasing in y and then increasing. It turns out that CR_2-dependent distributions preserve monotonicity of order 2. To illustrate that this stronger ordering of (conditional) distributions is natural, consider the problem of testing hypotheses about the parameter y of the one parameter family of distributions based on one observation of the variable X. If the family of distributions is stochastically increasing, then any one-sided test has a monotone power function, as one would hope. When this family also preserves order 2 monotonicity of functions, then any two-sided test has the desirable property that its power function is first decreasing and then increasing.

The following theorem formalizes this property of CR_s-dependent distributions. A function f defined on a subset Z of $\mathbb{R}$ is called *monotone of order* s, denoted by $M(f) = s$, when

$$\sup S^-[f(z_2) - f(z_1), f(z_3) - f(z_2), \ldots, f(z_t) - f(z_{t-1})] = s-1 \tag{3.2.5}$$

where t is arbitrary but finite and $z_1 < z_2 < \ldots < z_t$ ($z_i \in Z$; $i = 1,\ldots,t$), and f is said to be *strictly monotone of order* s, denoted $SM(f) = s$, if in addition to (3.2.5) also

$$\sup S^+[f(z_2) - f(z_1), f(z_3) - f(z_2), \ldots, f(z_t) - f(z_{t-1})] = s-1.$$

Furthermore, the function f is said to be (*strictly*) *monotone of order* s *with respect to a sigma-finite measure* σ, denoted $(S)M_\sigma(f) = s$, when $s = \inf (S)M(h)$, where the infimum is taken over all functions h which σ-a.e. equal f. Note that monotonicity is a kind of oscillatory property. In the case that Z is an interval and f is differentiable with derivative f', it follows from the mean value theorem that $M(f) = S^-(f') + 1$. Furthermore, when $Z = \{1,\ldots,n\}$ and d_i denotes the difference $d_i = f(i+1) - f(i)$ for $i = 1,\ldots,n-1$, then $M(f) = S^-[d_1,\ldots,d_{n-1}] + 1$.

THEOREM 3.2.3. *Consider the conditional expectations*

$$g(y) = E(f(X)|Y = y) \qquad \textit{for } y \in Y$$

where f *is a bounded measurable function.*

(i) *If* (X,Y) *is* CR_s*-dependent, then*

$$M_{\sigma_r}(f) \le s \Rightarrow M_{\sigma_c}(g) \le M_{\sigma_r}(f).$$

Moreover, when $M_{\sigma_r}(f) = M_{\sigma_c}(g) \le s$, *then* f *and* g *exhibit the same arrangement of oscillations.*

(ii) *If* (X,Y) *is* SCR_s*-dependent,* f *is monotone of order* $t \le s$ *with respect to the measure* σ_r *and* g *is monotone of order* t *with respect to* σ_c, *then* g *is even strictly monotone of order* t *with respect to* σ_c.

PROOF. (i) Without loss of generality it can be assumed that $M(f) = M_{\sigma_r}(f) \le s$ and hence f is of bounded variation. Therefore f can be approximated by step functions on intervals. Consider a partition $x_1 < x_2 < \ldots < x_n$ ($x_1 = \inf\{X\}$, $x_n = \sup\{X\}$, $x_i \in X$ for $i = 2,\ldots,n-1$) of X. (In case f is not properly defined in x_1 or in x_n replace x_1 or x_n by an element of X "close" to $\inf\{X\}$ or $\sup\{X\}$ respectively.) Let f_n be the step function such that $f_n(x) = f(x_{i+1})$ when $x_i < x \le x_{i+1}$. Clearly for all n and all partitions

$$(3.2.6) \qquad M(f_n) = S^-[f(x_2) - f(x_1),\ldots,f(x_n) - f(x_{n-1})] + 1 \le M(f).$$

Furthermore define

$$(3.2.7) \qquad g_n(y) = \int_X f_n(x)\,dF_{r|c}(x|y) \qquad \text{for } y \in Y,$$

for a suitable version of the conditional distribution. By bounded convergence, $g_n(y) \to g(y)$ for $y \in Y$ as $n \to \infty$ provided the partition of X is refined in an appropriate way.

Next fix $y_1 < y_2 < \ldots < y_m$ ($y_j \in Y$). Since (X,Y) is (S)CR_s-dependent it follows from Propositions 3.4.1 and 3.2.2 that the kernel $q_c(i,j) = F_{r|c}(x_{i+1}|y_j) - F_{r|c}(x_{i+1}|y_{j+1})$ is (S)TP_t, where $t = \min(s,n-1,m-1)$, on $\{1,\ldots,n-1\} \times \{1,\ldots,m-1\}$. Furthermore define $u_i = f(x_{i+1}) - f(x_i)$ for $i = 1,\ldots,n-1$ and $v_j = g_n(y_{j+1}) - g_n(y_j)$ for $j = 1,\ldots,m-1$. Then by (3.2.7) and summation by parts

$$(3.2.8) \qquad v_j = \sum_{i=1}^{n-1} u_i q_c(i,j) \qquad \text{for } j = 1,\ldots,m-1.$$

Applying the variation diminishing property to (3.2.8) produces

$$(3.2.9) \qquad S^-[g_n(y_2) - g_n(y_1),\ldots,g_n(y_m) - g_n(y_{m-1})] \le M(f_n) - 1 \le M(f) - 1$$

where the last inequality is obtained from (3.2.6).

Consider the vector $w = (g(y_2) - g(y_1),\ldots,g(y_m) - g(y_{m-1}))^T$. Choose $0 < \varepsilon < \min\{|w_j| : |w_j| > 0,\ 1 \le j \le m-1\}$, then any $(m-1)$-dimensional vector v such that $|w_j - v_j| < \varepsilon$ for $j = 1,\ldots,m-1$ satisfies $S^-(v) \ge S^-(w)$. Therefore, since $g_n \to g$ it follows from (3.2.9) with sufficiently large n that

$$S^-[g(y_2) - g(y_1),\ldots,g(y_m) - g(y_{m-1})] + 1 \le M(f).$$

Since m and $y_1 < y_2 < \ldots < y_m$ are arbitrary it follows that $M(g) \le M(f)$. Furthermore since g is a version of the conditional expectation, $M(g)$ can be replaced by $M_{\sigma_c}(g)$.

(ii) First note that in case of SCR_s-dependence, (3.2.9) also holds with S^- replaced by S^+. Moreover, the method of proof of part (i) also works for any step function $f_n(x) = f(\xi_{i+1})$ for $x_i < x \le x_{i+1}$ where $x_i < \xi_{i+1} \le x_{i+1}$. Instead of approximating f by one step function, now approximate f by upper and lower step functions f_n^u and f_n^ℓ such that $f_n^\ell(x) \le f(x) \le f_n^u(x)$ for all $x \in X$. Denote the corresponding conditional expectations by g_n^u and g_n^ℓ respectively. Then $g_n^\ell(y) \le g(y) \le g_n^u(y)$ and $g_n^\ell(y) - g_n^u(y) \to 0$ as $n \to \infty$ for all $y \in Y$. Moreover, for any $y_1 < y_2 < \ldots < y_m$ $(y_j \in Y)$

$$S^+[g_n^\ell(y_2) - g_n^\ell(y_1),\ldots,g_n^\ell(y_m) - g_n^\ell(y_{m-1})] \le M(f) - 1,$$
$$S^+[g_n^u(y_2) - g_n^u(y_1),\ldots,g_n^u(y_m) - g_n^u(y_{m-1})] \le M(f) - 1.$$

For sufficiently large n this implies

$$S^+[g(y_2) - g(y_1),\ldots,g(y_m) - g(y_{m-1})] \le M(f) - 1,$$

which proves the theorem. □

Under slight non-degeneracy conditions, the converse of Theorem 3.2.3 also holds.

Let's return to the case where X and Y are categorical variables and introduce the notion order dependence. It is clear that TP, RR and CR-dependence can be defined because the elements of the sets X and Y can be ordered. For categorical variables there is generally no clear a priori ordering of categories and hence these forms of dependence are not properly defined. But in case there *exist* labels $1,\ldots,n$ for the categories of the row variable and labels $1,\ldots,m$ for the categories of the column variable

such that the corresponding (integer valued) variables X and Y are CR_1-dependent, then the mass of the conditional distributions of $X|Y = y$ concentrates on row categories with larger indices as the category index y of the column variable increases. Moreover, the conditional distribution functions do not cross each other and can be ordered properly. Hence CR_1-dependence implies that given the labels of the row variable, the ordering of categories corresponding to the labels of the column variable is meaningful. If in addition (for the same labels) the variables X and Y are also RR_1-dependent, then the conditional distributions of $Y|X = x$ imply that the ordering of row categories according to the labels also makes sense. Summarizing, when there exist numerical labels for the categories of the row and column variable such that the labeled variables are DR_1-dependent, then the dependence implies a meaningful ordering over the row and column categories. In this case there is a strong ordinal relation between both categorical variables. Intuitively, it is obvious that an ordering of categories itself is not necessarily meaningful; it often only makes sense with respect to other variables.

DEFINITION 3.2.3. Two categorical variables are called (*strictly*) *order dependent of order* s when there exist labels 1,...,n and 1,...,m for the categories of both variables such that the labeled integer valued variables are (S)DR_s-dependent.

It will follow from Theorems 3.3.1 and 3.3.2 below that the ordering of the labels in Definition 3.2.3 is unique when the probability table of both categorical variables has no proportional rows and columns.

Recall that S denotes the upper triangular matrix with unities on and above the diagonal. For a $n \times m$ probability table P let the $(n-1) \times (m-1)$ matrices Q_r and Q_c^T be obtained by deleting the last row and column of $S^{-1}R^{-1}PS$ and of $(S^{-1}C^{-1}P^TS)^T$ respectively. Then by Proposition 3.2.2 the row and column variable of a $n \times m$ probability table are (strictly) order dependent of order s iff there exists a permutation of rows and columns such that the matrices Q_r and Q_c corresponding to the permuted table P are both (S)TP_s.

The following proposition gives an example of order dependence. Further examples are given in Section 3.4.

PROPOSITION 3.2.4. *The row and column variables of any rank 2 probability table are order dependent of order* 1.

PROOF. Let P be the rank 2 probability table with canonical decomposition

$$P = rc^T + \lambda_1 R\varphi_1\psi_1^T C.$$

Rearrange the rows and columns such that the vectors φ_1 and ψ_1 of row and column category scores are both increasing in their components (note that category scores are invariant under permutations of rows and columns). Straightforward computation of the matrices Q_r and Q_c for the permuted table shows that they are TP_1. □

3.3. CORRESPONDENCE ANALYSIS AND ORDER DEPENDENCE

It is clear that CA makes no assumption on, nor uses any ordering of the categories of the row and column variables. CA considers both variables nominally; under permutation of rows and columns of the probability table, the category scores undergo the same permutation. However, the following phenomenon is frequently observed in practice. When the categories of both variables have an intuitive meaningful order, then this order is often reflected by the order of the category scores of the first canonical pair. A second phenomenon often encountered in practice is the so called "horseshoe" in Benzécri's two-dimensional graphical display of the probability table; i.e. the row points and column points lie on convex or concave curves. The main result given in this section shows that both phenomena occur when the row and column variables are order dependent (of order 2). Moreover, it is shown that the stronger ordinal relations corresponding to higher orders of order dependence are reflected in the higher pairs of canonical functions.

Consider the setting of Section 2.1, where X and Y denote categorical variables with categories labeled by $1,\ldots,n$ and $1,\ldots,m$, respectively, with $n \leq m$. Recall that φ_t and ψ_t denote the vectors of row and column category scores on the t-th canonical pair with canonical correlation λ_t (for $t = 1,\ldots,n-1$).

THEOREM 3.3.1. *Suppose that, for given category labels,* (X,Y) *is* SDR_s*-dependent, then*

(i) *the first* s *canonical correlations are strictly positive and distinct, i.e.* $1 \geq \lambda_1 > \lambda_2 > \ldots > \lambda_s > \lambda_{s+1}$,

(ii) *the canonical functions* φ_t *and* ψ_t *for* $t = 1,\ldots,s$ *are strictly*

monotone of order t *and exhibit the same arrangement of oscillations. Moreover, for arbitrary real numbers* $a_k, a_{k+1}, \ldots, a_\ell$ $(1 \le k \le \ell \le s, \Sigma_{t=k}^{\ell} a_t^2 > 0)$ *the orders of monotonicity of the linear combinations*

$$\varphi = \sum_{t=k}^{\ell} a_t \varphi_t \qquad \textit{and} \qquad \psi = \sum_{t=k}^{\ell} a_t \psi_t$$

satisfy $k \le M(\varphi), M(\psi) \le \ell$.

(iii) *In case* $s \ge 2$, *the sign of* φ_2 *can be chosen such that* φ_2 *is strictly convex with respect to* φ_1, *and* ψ_2 *is strictly convex with respect to* ψ_1 (*i.e. there exist strictly convex functions* f *and* g *such that* $\varphi_2 = f(\varphi_1)$ *and* $\psi_2 = g(\psi_1)$).

PROOF. Let the joint distribution of (X,Y) be given by the $n \times m$ probability table P and let R and C denote the diagonal matrices with row and column sums (as usual). By Theorem 2.1.3 the vectors $e, \varphi_1, \ldots, \varphi_{n-1}$ form a complete set of eigenvectors of $R^{-1}PC^{-1}P^T$ corresponding to the eigenvalues $1 \ge \lambda_1^2 \ge \ldots \ge \lambda_{n-1}^2$. Hence, since the square matrix S is non-singular, $S^{-1}e, S^{-1}\varphi_1, \ldots, S^{-1}\varphi_{n-1}$ form a complete set of eigenvectors of $\bar{Q} = S^{-1}R^{-1}PC^{-1}P^TS$ corresponding to the same eigenvalues. Let Q be obtained by deleting the last row and column of $\bar{Q}$. Since $R^{-1}PC^{-1}P^T$ has row sums unity, it follows that the last column of $\bar{Q}$ equals $(0,\ldots,0,1)^T$. This implies that $S^{-1}e$ is an eigenvector of $\bar{Q}$ corresponding to eigenvalue unity and that for $t = 1,\ldots,n-1$, $\bar{u}_t = (\bar{u}_{1t},\ldots,\bar{u}_{n-1\,t},\bar{u}_{nt})^T$ is an eigenvector of $\bar{Q}$ corresponding to eigenvalue λ_t^2 iff $u_t = (\bar{u}_{1t},\ldots,\bar{u}_{n-1\,t})^T$ is an eigenvector of Q corresponding to the same eigenvalue and $(\lambda_t^2 - 1)\bar{u}_{nt} = \Sigma_{i=1}^{n-1} \bar{u}_{it}$. Therefore, φ_t can be taken such that $S^{-1}\varphi_t = \bar{u}_t$ for $t = 1,\ldots,n-1$. Thus,

$$u_{it} = \bar{u}_{it} = \varphi_{it} - \varphi_{i+1\,t} \qquad \text{for } i = 1,\ldots,n-1 \text{ and } t = 1,\ldots,n-1. \tag{3.3.1}$$

Furthermore, since X and Y are SDR_s-dependent, the matrices Q_r and Q_c which are obtained by deleting the last row and column of $S^{-1}R^{-1}PS$ and of $S^{-1}C^{-1}P^TS$ respectively are both STP_s (cf. Section 3.2). It is easily verified that $Q = Q_rQ_c$ and hence Q is STP_s. Application of Lemma 3.1.5 to the matrix Q yields $\lambda_1^2 > \lambda_2^2 > \ldots > \lambda_s^2 > \lambda_{s+1}^2$ and that for arbitrary real numbers $a_k,\ldots,a_\ell$ $(1 \le k \le \ell \le s, \Sigma_k^\ell a_t^2 > 0)$ the number of sign changes of the linear combination

$$u = \sum_{t=k}^{\ell} a_t u_t$$

satisfies $k-1 \leq S^-(u) \leq S^+(u) \leq \ell-1$. So in particular $S^-(u_t) = S^+(u_t) = t-1$ for $t = 1,\ldots,s$. In view of (3.1.1) this proves part (i) of the theorem and the monotonicity result of the canonical functions $\varphi_1,\ldots,\varphi_s$. The results for the vectors $\psi_1,\ldots,\psi_s$ follow similarly. By Theorem 3.2.3 and the transition formula (2.1.4), φ_t and ψ_t exhibit the same arrangement of oscillations. Moreover, it follows from (3.1.7) for $s = 2$ that the sign of φ_2 can be chosen such that

$$1 \leq i_1 < i_2 \leq n-1 \Rightarrow u_{i_1 2}/u_{i_1 1} < u_{i_2 2}/u_{i_2 1},$$

which by (3.3.1) implies that φ_2 is strictly convex with respect to φ_1. It similarly follows that ψ_2 is strictly convex with respect to ψ_1. □

Note that the conditions of this theorem are somewhat too strong, because it is sufficient that some iterate of Q is STP_s, whereas the present conditions imply that Q itself is STP_s. However, it seems hard to find simple sufficient conditions for Theorem 3.3.1 which are essentially weaker.

THEOREM 3.3.2. *Suppose that, for given labels,* (X,Y) *is* DR_s*-dependent, then there exist, for* $t = 1,\ldots,s$, *canonical functions* φ_t *and* ψ_t *which are monotone of order* t *exhibiting the same arrangement of oscillations. Furthermore, the results* (ii) *and* (iii) *of Theorem* 3.3.1 *are valid in a non-strict sense for these canonical functions.*

PROOF. Follows from Theorem 3.3.1, Lemma 3.1.6 and continuity considerations. □

When φ_2 is (strictly) convex with respect to φ_1, the two-dimensional row points $\xi_1 = (\lambda_1\varphi_{11},\lambda_2\varphi_{12}),\ldots,\xi_n = (\lambda_1\varphi_{n1},\lambda_2\varphi_{n2})$ lie on a (strictly) convex curve. So "horseshoes" in Benzécri's two-dimensional graphical representation are implied by order dependence of order 2.

Theorems 3.3.1 and 3.3.2 are quite relevant for the practice of CA. If the row and column variables are order dependent, then the (meaningful) ordering of categories induced by the dependence is reflected in the order of the category scores on the first canonical pair and hence is easily obtained by inspection of φ_1 and ψ_1. Moreover, since the most important aspect of assigning scores to categories is perhaps the ordering induced by these scores, these results support the use of the canonical variables $\varphi_1(X)$

and $\psi_1(Y)$ as numerical versions of the nominal row and column variables.

The abundance of examples given in the next section demonstrate that DR or order dependence is quite common in practical models for probability tables. Although this does not imply the DR character for random samples from such populations, one may nevertheless expect that empirical probability tables also often show DR-dependence or are close to it, and hence that the ordering properties of the canonical functions remain valid. It is, however, difficult to derive precise and useful statistical properties of such qualitative aspects of CA. A better argument for the practical relevance of these results, when sampling variation has to be taken into account, is the fact mentioned before, that an intuitively expected ordering and the "horseshoe" phenomenon are so often found with small-sample real data. When, 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 it. This makes CA a very useful exploratory tool for checking that an a priori ordering of categories of a nominal variable is correct or for pointing out a breakdown for such an ordering.

One can probably show by using approximation arguments that similar ordering properties hold for the canonical functions of continuous bivariate distributions. But the arguments above, showing the relevance of the results in the contingency table case, do not carry over to the continuous case, which makes such generalizations less useful. In spite of this, such results contribute to a better description of the structure of bivariate distributions and give solutions to a somewhat more general problem than considered by EAGLESON (1964), EAGLESON and LANCASTER (1967), LANCASTER (1975, 1980). These authors were interested in the problem of determining in which cases the canonical functions of the decomposition (2.1.9) are orthogonal polynomials. Since orthogonal polynomials are strictly monotone (with order equal to degree), this should yield a subclass of the DR_∞-dependent distributions. (Note that by Theorem 3.2.3, the class of DR_∞-dependent distributions can be characterized by monotonicity preserving properties; similarly, the class of distributions with polynomial canonical functions preserves linearity, convexity, etc.)

3.4. ORDER DEPENDENCE IN PRACTICE

Order dependence frequently occurs in real data. In fact, the

variables hair and eye colour of the Scottish school children as considered in one of the first applications of CA by Fisher and Maung, are (up to a very slight correction) order dependent of order 1. Another well-studied example in literature is the British social mobility contingency table, which compares the occupational statuses of father and son (cf. GIFI (1981, p. 145), HABERMAN (1974, p. 217) and many other authors). This table is also order dependent of order 1 and since it is close to order dependence of order 2, the "horseshoe" phenomenon is found in its graphical representation.

In this section two important classes of models for (underlying) order dependent probability tables are given. A third model are rank 2 probability tables (cf. Proposition 3.2.4).

The first class arises from discretizations from well-known bivariate distributions. The $n \times m$ probability table P is said to be a *discretization* of the distribution F when there exist intervals $A_1 < A_2 < \ldots < A_n$ with $\cup_{i=1}^{n} A_i = \mathbb{R}$ and intervals $B_1 < \ldots < B_m$, $\cup_{j=1}^{m} B_j = \mathbb{R}$ such that

$$p_{ij} = \int_{A_i} \int_{B_j} dF \qquad \text{for } i = 1,\ldots,n;\ j = 1,\ldots,m, \tag{3.4.1}$$

and

$$r_i = \int_{A_i} dF_r > 0 \quad \text{and} \quad c_j = \int_{B_j} dF_c > 0 \qquad \text{for } i = 1,\ldots,n;\ j = 1,\ldots,m. \tag{3.4.2}$$

PROPOSITION 3.4.1. *Let the distribution* F *be* $(S)TP_s$, $(S)DR_s$, $(S)CR_s$ *or* $(S)RR_s$*-dependent, then any discretization of* F *into a* $n \times m$ *probability table is* $(S)TP_t$, $(S)DR_t$, $(S)CR_t$ *or* $(S)RR_t$*-dependent, where* $t = \min(s,n,m)$.

PROOF. First consider the case of $(S)TP_s$-dependence. Let X and Y be subsets of $\mathbb{R}$ such that F has density p with respect to the product measure $\sigma_r \times \sigma_c$ on $X \times Y$ with p $(S)TP_s$ on $X \times Y$. Consider an interval A such that $\int_A dF_r > 0$ and hence $\sigma_r(A) > 0$. Let $\tilde{X}$ be obtained from X by replacing the set $X \cap A$ by one element $a \in X \cap A$; i.e. $\tilde{X} = \{X \setminus A\} \cup \{a\}$. Define $\tilde{p}$ on $\tilde{X} \times Y$ by

$$\tilde{p}(x,y) = \begin{cases} p(x,y) & \text{if } x \in X \setminus A,\ y \in Y \\ \int_A p(x,y)\,d\sigma_r(x) & \text{if } x = a,\ y \in Y \end{cases},$$

and let s' equal s when $\tilde{X}$ contains an interval and equal the minimum of s and the number of points in $\tilde{X}$ otherwise. Then for every $t = 1,\ldots,s'$ and all $x_1 < \ldots < x_i < a < x_{i+1} < \ldots < x_{t-1}$ $(x_i \in X \setminus A)$ and all

$y_1 < \ldots < y_t$ $(y_j \in Y)$

$$\tilde{p}\begin{pmatrix} x_1,\ldots,x_i,a,x_{i+1},\ldots,x_{t-1} \\ y_1,\ldots \qquad\qquad \ldots,y_t \end{pmatrix} =$$

$$= \int_A p\begin{pmatrix} x_1,\ldots,x_i,x,x_{i+1},\ldots,x_{t-1} \\ y_1,\ldots \qquad\qquad\qquad \ldots,y_t \end{pmatrix} d\sigma_r(x) .$$

Hence $\tilde{p}$ is $(S)TP_s$, on $\tilde{X} \times Y$. Repeated application yields the desired result for $(S)TP_s$-dependence.

Next consider the case that F shows $(S)CR_s$ or $(S)RR_s$-dependence on $X \times Y$. Define in the setting above

$$\tilde{F}(x,y) = \begin{cases} F(x,y) & \text{if } x \notin A,\ y \in \mathbb{R} \\ F(b,y) & \text{if } x \in A,\ y \in \mathbb{R} \end{cases}$$

where $b = \sup\{A\}$. Hence

$$\tilde{F}_{r|c}(x|y) = \begin{cases} F_{r|c}(x|y) & \text{if } x \notin A,\ y \in Y \\ F_{r|c}(b|y) & \text{if } x \in A,\ y \in Y \end{cases}$$

and

$$\tilde{F}_{c|r}(y|x) = \begin{cases} F_{c|r}(y|x) & \text{if } x \in X \setminus A,\ y \in \mathbb{R} \\ \int_A F_{c|r}(y|x)dF_r(x) & \text{if } x = a,\ y \in \mathbb{R} \end{cases} .$$

The determinant inequalities (3.2.2) are easily verified for $\tilde{F}_{r|c}$ and $\tilde{F}_{c|r}$ and hence $\tilde{F}$ shows $(S)CR_s$, or $(S)RR_s$,-dependence on $\tilde{X} \times Y$. □

In most examples below it is easier to verify that bivariate distributions are $(S)TP_{s+1}$-dependent rather than the weaker $(S)DR_s$-dependence. To identify various types of bivariate distributions, references are given in which the distributions are derived.

Some bivariate distributions can be generated by the method of *trivariate reduction*. When Z_1, Z_2 and Z_3 are independent random variables (usually with distributions from a common family which is closed under convolutions), then the joint distribution of $(Z_1 + Z_3, Z_2 + Z_3)$ is said to be generated by trivariate reduction. It follows from Corollary 3.1.3 and Example 3.1.9 that such distributions are $(S)TP_s$-dependent when the densities of Z_1, Z_2 and Z_3 are $(S)PF_s$. It is seen from Examples 3.1.2, 3.1.7, 3.1.6 and 3.1.5 respectively that densities of the univariate normal, binomial, Poisson and gamma distributions are actually $(S)PF_\infty$. Hence the *bivariate normal* (with correlation parameter $\rho > 0$), the

bivariate binomial, the *bivariate Poisson* (HOLGATE, 1964) and the *bivariate gamma* (CHERIAN, 1941) are (S)TP_∞-dependent. In fact, EAGLESON (1964) shows that the canonical functions of these four bivariate distributions are orthogonal Hermite, Kratchouck, Poisson-Charlier and Laguerre polynomials respectively. So it alternatively follows by the canonical decomposition (2.1.9), Corollary 3.1.3 and Example 3.1.8 that the latter three distributions are STP_∞-dependent. (Note that Hermite polynomials do not satisfy the requirements of Example 3.1.8.)

STP_∞-dependence of the *negative trinomial*, the *bivariate* F (GHOSH, 1955), *the bivariate Pareto* (MARDIA, 1962), the *bivariate logistic* (GUMBEL, 1961) and the *bivariate exponential* distribution (CLAYTON and CUZICK, 1985) can be proved using the examples in Section 3.1 and Proposition 3.1.1. The *trinomial*, the *bivariate hypergeometric* and the *standard Dirichlet* or *bivariate beta* (JOHNSON, 1960) show negative dependence, but it is easily verified that the reversed distributions (reversed in one variable) are actually TP_∞-dependent. Of course not all bivariate distributions are TP_s-dependent; for instance the bivariate Cauchy density $p(x,y) = (2\pi)^{-1}(1 + x^2 + y^2)^{-\frac{3}{2}}$ is not TP_2 on $\mathbb{R}^2$.

Another, more specific model for the probability table P is the *linear by linear interaction model*

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

where $\Sigma_i\ \alpha_i = \Sigma_j\ \beta_j = \Sigma_i\ \gamma_i = \Sigma_j\ \delta_j = 0$. The probability table P is STP_n when the rows and columns are indexed such that γ_i and δ_j are both strictly increasing in their indices. Thus the row and column variables are always (strictly) order dependent of order n. GOODMAN (1981) compares maximum likelihood estimates of γ and δ in this model with the first pair of canonical functions φ_1 and ψ_1. Furthermore, he discusses the ordering of rows and columns which is present in this model by means of TP_2 and DR_1-dependence; however, he does not prove that this ordering is reflected in the components of φ_1 and ψ_1.

In Section 2.1 it was noted that CA could also be applied to incidence matrices. Incidence matrices are also often TP. A good example is the Münsingen-Rain incidence matrix (cf. KENDALL, 1971). This matrix only slightly departs from a band diagonal matrix, which is TP_n. Application of CA produces "horseshoes" (see HILL, 1974) as one could expect in view of

Theorems 3.3.1 and 3.3.2.

3.5. TESTS SENSITIVE TO ORDERED ALTERNATIVES

This section briefly discusses some tests of independence which are sensitive to ordered alternatives. The first three tests are due to Yates and the other two are based on Kendall's rank correlation tau and Spearman's rank correlation rho. The tests considered are quite easily performed and show a considerable gain of power against most regression dependent alternatives, compared to the chi-square and likelihood ratio test.

Consider the setting of Sections 2.2 and 2.3 where P_0 and P_L denote the null-hypothesis and local alternative conditional Poisson distributions for the empirical probability table $\hat{P}^{(N)}$, drawn from underlying tables (2.2.11) and (2.2.12) respectively and where the conditioning matrix A satisfies (2.2.10). Recall that P' denotes the limiting alternative table.

Before turning to the tests under consideration, the following remark should be made. In case the underlying probability table shows DR_1-dependence, one might hope that the first pair of empirical canonical functions $\hat{\varphi}_1^{(N)}$ and $\hat{\psi}_1^{(N)}$ are often monotone of order 1. Hence tests based on the statistics (2.3.2) and (2.3.3), with a and b strictly monotone too, might performe well against such alternatives. However, Proposition 2.3.4 shows that such tests are inconsistent against most DR_1-dependent alternatives. So an attempt to base tests sensitive to regression dependence on the first pair of empirical canonical variables only, fails. The strength of the dependence must be taken into account too.

YATES (1948) introduced a test of independence which is based on $N^{\frac{1}{2}}$ times the sample (product moment) correlation between row and column variables scored with preassigned values $a = (a_1,\ldots,a_n)^T$ and $b = (b_1,\ldots,b_m)^T$, which are not proportional to the vector e, i.e. based on

$$(3.5.1)\qquad T_Y^{(N)} = N^{\frac{1}{2}}(a^T\hat{P}^{(N)}b - a^T\hat{f}^{(N)}b^T\hat{c}^{(N)})/((a^T\hat{R}^{(N)}a-(a^T\hat{f}^{(N)})^2)(b^T\hat{C}^{(N)}b-(b^T\hat{c}^{(N)})^2))^{\frac{1}{2}}.$$

Against alternatives with (limiting) canonical functions φ_1 and ψ_1, Yates' test performs asymptotically best when $a = \varphi_1$ and $b = \psi_1$. Thus when DR_1-dependent alternatives are expected, one should take a and b monotone of order 1; e.g. $a = (1,\ldots,n)^T$ and $b = (1,\ldots,m)^T$ if no prior information about the canonical functions is available.

THEOREM 3.5.1. *Under the hypothesis* P_0,

$$T_Y^{(N)} \to_D N_1(0,1) \qquad \text{as } N \to \infty \tag{3.5.2}$$

and under local alternatives P_L,

$$T_Y^{(N)} \to_D N_1(\theta\tau_Y, 1) \qquad \text{as } N \to \infty \tag{3.5.3}$$

where

$$\tau_Y = (a^T P' b - a^T r b^T c)/((a^T R a - (a^T r)^2)(b^T C b - (b^T c)^2))^{\frac{1}{2}}.$$

Furthermore, the test which rejects for large (positive) values of $T_Y^{(N)}$ *is consistent against all fixed alternatives of interest.*

PROOF. Note that

$$T_Y^{(N)} = N^{\frac{1}{2}} a^T \hat{R}^{(N)\frac{1}{2}} \hat{W}^{(N)} \hat{C}^{(N)\frac{1}{2}} b / (a^T (\hat{R}^{(N)} - \hat{r}^{(N)} \hat{r}^{(N)T}) a b^T (\hat{C}^{(N)} - \hat{c}^{(N)} \hat{c}^{(N)T}) b)^{\frac{1}{2}}.$$

Since $\hat{R}^{(N)}$, $\hat{r}^{(N)}$, $\hat{C}^{(N)}$ and $\hat{c}^{(N)}$ converge under P_0 and P_L in probability to R, r, C and c respectively, the results (3.5.2) and (3.5.3) follow from (2.2.14) and (2.2.15). Furthermore, under fixed alternatives P_A, $\hat{W}^{(N)}$ converges in probability to W' and hence $N^{-\frac{1}{2}} T_Y^{(N)}$ converges in probability to τ_Y as $N \to \infty$. Therefore, the test which rejects for large positive values of $T_Y^{(N)}$ is consistent against fixed alternatives with $\tau_Y > 0$. □

In case alternatives are expected in which only φ_1 is monotone of order 1 or in case the ordering of column categories induced by regression dependence is unknown, one can use the test statistic

$$T_{YR}^{(N)} = N(a^T \hat{P}^{(N)} \hat{C}^{(N)-1} \hat{P}^{(N)T} a - (a^T \hat{r}^{(N)})^2) / (a^T \hat{R}^{(N)} a - (a^T \hat{r}^{(N)})^2)$$

where a is monotone of order 1 (not proportional to e). This statistic can be interpreted as N times the square of the maximal achievable sample correlation between scored row and column variables, when the scores of the row variable are the preassigned values $a = (a_1, \ldots, a_n)^T$. Similarly define

$$T_{YC}^{(N)} = N(b^T \hat{P}^{(N)T} \hat{R}^{(N)-1} \hat{P}^{(N)} b - (b^T \hat{c}^{(N)})^2) / (b^T \hat{C}^{(N)} b - (b^T \hat{c}^{(N)})^2).$$

THEOREM 3.5.2. *The asymptotic distributions of* $T_{YR}^{(N)}$ *and* $T_{YC}^{(N)}$ *under* P_0 *are chi-square with* m-1 *and* n-1 *degrees of freedom respectively. Their asymptotic distributions under* P_L *are non-central chi-square with the same*

degrees of freedom and non-centrality parameters

$$\theta^2\tau_{YR}^2 = \theta^2(a^TP'^TC^{-1}P'^Ta - (a^Tr)^2)/(a^TRa - (a^Tr)^2)$$

and

$$\theta^2\tau_{YC}^2 = \theta^2(b^TP'^TR^{-1}P'b - (b^Tc)^2)/(b^TCb - (b^Tc)^2)$$

respectively. Furthermore, tests based on these statistics are consistent against all fixed alternatives.

PROOF. Analogous to the proof of Theorem 3.5.1. □

The next two tests are based on probability table adaptations of Kendall's rank correlation tau and Spearman's rank correlation rho. Adaptation can be done in several ways; the present choice is motivated in Section 4.4. The probability table version of Kendall's tau for a probability table P is defined by

$$\tilde{\phi}_K(P) = 2\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}\left(\sum_{k=1}^{i-1}\sum_{\ell=1}^{j-1} p_{k\ell} - \sum_{k=1}^{i-1}\sum_{\ell=j+1}^{m} p_{k\ell}\right)$$

$$= 2\{\text{sum of all } 2\times 2 \text{ determinants}\}.$$

For an empirical probability table $\hat{P}^{(N)}$ this equals

$$\tilde{\phi}_K(\hat{P}^{(N)}) = (\#\{\text{concordant pairs of obs.}\} - \#\{\text{disconc. pairs of obs.}\})/(\tfrac{1}{2}N^2).$$

The probability table version of Spearman's rho for P is defined by

$$\tilde{\phi}_S(P) = 3\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}\left(\sum_{k=1}^{i-1} r_k - \sum_{k=i+1}^{n} r_k\right)\left(\sum_{\ell=1}^{j-1} c_\ell - \sum_{\ell=j+1}^{m} c_\ell\right).$$

Note that $\sum_{k=1}^{i-1} r_k - \sum_{k=i+1}^{n} r_k$ for $i = 1,\ldots,n$ are centered mid-ranks times $2N^{-1}$ for the categories of the row variable. Thus $\tilde{\phi}_S(P)$ is the covariance between row and column variables, scored with these centered mid-ranks. The test statistics based on Kendall's tau and Spearman's rho are

$$T_K^{(N)} = \tfrac{1}{2}N^{\frac{1}{2}}\tilde{\phi}_K(\hat{P}^{(N)})/\hat{v}^{(N)}$$

$$T_S^{(N)} = \tfrac{1}{6}N^{\frac{1}{2}}\tilde{\phi}_S(\hat{P}^{(N)})/\hat{v}^{(N)}$$

where $\hat{v}^{(N)}$ is the product of the sample standard deviations of row and column variables scored with (sample) mid-ranks, i.e.

$$\hat{v}^{(N)} = \left(\sum_{i=1}^{n}\hat{r}_i^{(N)}\left(\sum_{k=1}^{i-1}\hat{r}_k^{(N)} - \sum_{k=i+1}^{n}\hat{r}_k^{(N)}\right)^2\sum_{j=1}^{m}\hat{c}_j^{(N)}\left(\sum_{\ell=1}^{j-1}\hat{c}_\ell^{(N)} - \sum_{\ell=j+1}^{m}\hat{c}_\ell^{(N)}\right)^2\right)^{\frac{1}{2}}.$$

THEOREM 3.5.3. *Under the hypothesis* P_0,

$$T_K^{(N)} \to_D N_1(0,1), \qquad T_S^{(N)} \to_D N_1(0,1) \qquad \text{as } N \to \infty$$

and under local alternatives P_L,

$$T_K^{(N)} \to_D N_1(\theta\tau_S,1), \quad T_S^{(N)} \to_D N_1(\theta\tau_S,1) \qquad \text{as } N \to \infty$$

where

$$\tau_S = \tfrac{1}{3}\tilde{\phi}_S(P')/\Big(\sum_{i=1}^{n} r_i\Big(\sum_{k=1}^{i-1} r_k - \sum_{k=i+1}^{n} r_k\Big)^2 \sum_{j=1}^{m} c_j\Big(\sum_{\ell=1}^{j-1} c_\ell - \sum_{\ell=j+1}^{m} c_\ell\Big)^2\Big)^{\frac{1}{2}}.$$

Furthermore, tests based on both statistics are consistent against all fixed alternatives of interest.

PROOF. The proof is based on an application of the ∂-method (cf. RAO, 1973, p. 388). Denote by ρ the vector with components

$$\rho_i = \sum_{k=i+1}^{n} r_k - \sum_{k=1}^{i-1} r_k \qquad \text{for } i = 1,\ldots,n \tag{3.5.4}$$

and by γ the vector with components

$$\gamma_j = \sum_{\ell=j+1}^{m} c_\ell - \sum_{\ell=1}^{j-1} c_\ell \qquad \text{for } j = 1,\ldots,m. \tag{3.5.5}$$

Straightforward computations show that the derivatives of $\tilde{\phi}_K$ at $P = rc^\top$ are given by

$$\begin{aligned}
&\frac{\partial\tilde{\phi}_K(P)}{\partial p_{ij}}\Bigg|_{P=rc^\top} \\
&= 2\Big(\sum_{k=1}^{i-1}\sum_{\ell=1}^{j-1} p_{k\ell} - \sum_{k=1}^{i-1}\sum_{\ell=j+1}^{m} p_{k\ell} - \sum_{k=i+1}^{n}\sum_{\ell=1}^{j-1} p_{k\ell} - \sum_{k=i+1}^{n}\sum_{\ell=j+1}^{m} p_{k\ell}\Big)\Big|_{P=rc^\top} \\
&= 2\rho_i\gamma_j.
\end{aligned}$$

Similarly the derivatives of $\tilde{\phi}_S$ at $P = rc^\top$ equal

$$\begin{aligned}
&\frac{\partial\tilde{\phi}_S(P)}{\partial p_{ij}}\Bigg|_{P=rc^\top} \\
&= 3\Big(\sum_{k=1}^{i-1}\sum_{\ell=1}^{j-1} p_{k\ell} - \sum_{k=1}^{i-1}\sum_{\ell=j+1}^{m} p_{k\ell} - \sum_{k=i+1}^{n}\sum_{\ell=1}^{j-1} p_{k\ell} + \sum_{k=i+1}^{n}\sum_{\ell=j+1}^{m} p_{k\ell} + \rho_i\gamma_j\Big)\Big|_{P=rc^\top} \\
&= 6\rho_i\gamma_j.
\end{aligned}$$

Furthermore, by Lemma 2.2.1 it follows that under P_0

$$N^{\frac{1}{2}} vec\,[\hat{P}^{(N)} - r^{(N)}c^{(N)^T}] \to_D N_{nm}(0,\ddagger) \quad \text{as } N \to \infty$$

and under P_L

$$N^{\frac{1}{2}} vec\,[\hat{P}^{(N)} - r^{(N)}c^{(N)^T}] \to_D N_{nm}(\theta\, vec\,[R^{\frac{1}{2}}W'C^{\frac{1}{2}}],\ddagger) \quad \text{as } N \to \infty$$

where

$$\ddagger = (R \otimes C) - (R \otimes C)A^T(A(R \otimes C)A^T)^{-}A(R \otimes C).$$

Therefore application of the ϑ-method yields under P_0

$$N^{\frac{1}{2}}\tilde{\phi}_K(\hat{P}^{(N)}) \to_D N_1(0,4(\rho \otimes \gamma)^T\ddagger(\rho \otimes \gamma))$$

$$N^{\frac{1}{2}}\tilde{\phi}_S(\hat{P}^{(N)}) \to_D N_1(0,36(\rho \otimes \gamma)^T\ddagger(\rho \otimes \gamma))$$

and under P_L

$$N^{\frac{1}{2}}\tilde{\phi}_K(\hat{P}^{(N)}) \to_D N_1(2\theta\rho^TR^{\frac{1}{2}}W'C^{\frac{1}{2}}\gamma,4(\rho \otimes \gamma)^T\ddagger(\rho \otimes \gamma))$$

$$N^{\frac{1}{2}}\tilde{\phi}_S(\hat{P}^{(N)}) \to_D N_1(6\theta\rho^TR^{\frac{1}{2}}W'C^{\frac{1}{2}}\gamma,36(\rho \otimes \gamma)^T\ddagger(\rho \otimes \gamma))$$

as $N \to \infty$. It is easily verified that $\rho^Tr = 0$ and $\gamma^Tc = 0$. Moreover, since the conditioning matrix A is of the form (2.2.10) it follows that $(\rho \otimes \gamma)^T\ddagger(\rho \otimes \gamma) = (\rho \otimes \gamma)^TR \otimes C(\rho \otimes \gamma) = \rho^TR\rho\gamma^TC\gamma$. Furthermore, one can verify that $\rho^TR^{\frac{1}{2}}W'C^{\frac{1}{2}}\gamma = \frac{1}{3}\tilde{\phi}_S(P')$. Since $\hat{v}^{(N)}$ converges under P_0 and under P_L in probability to $(\rho^TR\rho\gamma^TC\gamma)^{\frac{1}{2}}$, the limiting distributions of the statistics under null-hypothesis and local alternatives follow. Under fixed alternatives P_A, $\hat{P}^{(N)}$ converges in probability to P', implying consistency of the tests which reject for large values of $T_K^{(N)}$ and $T_S^{(N)}$ against fixed alternatives with $\tilde{\phi}_K(P') > 0$ and $\tilde{\phi}_S(P') > 0$. □

It is shown in Section 4.4 that these tests are unbiased in finite samples against CR_1 and RR_1-dependent alternatives.

The limiting distributions of another adaptation of Kendall's tau is studied in GOODMAN and KRUSKAL (1963, 1972).

It is easily shown that the sequences of statistics $\{T_Y^{(N)}\}_{N=1}^{\infty}$, $\{T_{YR}^{(N)^{\frac{1}{2}}}\}_{N=1}^{\infty}$, $\{T_{YC}^{(N)^{\frac{1}{2}}}\}_{N=1}^{\infty}$, $\{T_K^{(N)}\}_{N=1}^{\infty}$ and $\{T_S^{(N)}\}_{N=1}^{\infty}$ are Bahadur standard sequences with approximate slopes

$$s_Y(P') = \tau_Y^2,$$

$$s_{YR}(P') = \tau_{YR}^2,$$

$$s_{YC}(P') = \tau_{YC}^2,$$

$$s_K(P') = \tilde{\phi}_K^2(P')/(4v^2)$$

$$s_S(P') = \tilde{\phi}_S^2(P')/(36v^2)$$

where $v^2 = \rho^T R \rho \gamma^T C \gamma$, provided attention is restricted to alternatives with $\tau_Y > 0$, $\tau_{YR} \neq 0$, $\tau_{YC} \neq 0$, $\tilde{\phi}_K(P') > 0$ and $\tilde{\phi}_S(P') > 0$ respectively. This in combination with the approximate Bahadur slopes given in Section 2.3 suggest that the chi-square test, the likelihood ratio test and even the canonical correlation test always dominate Yates' test. Numerical computations show that this is certainly not true; in the present case approximate Bahadur efficiencies very badly describe the relative performances of tests. Note that the Pitman efficiency of $T_K^{(N)}$ with respect to $T_S^{(N)}$ equals unity. No (further) studies have yet been made of the relative performance of the tests in this section. GROSS (1981) compares tests based on measures of Yates, Spearman and HABERMAN (1974 b) by means of particular asymptotic relative efficiencies. NGUYEN (1982) describes tests which are superior to tests based on Spearman's rho and Kendall's tau for a restricted class of alternatives.

The vectors of preassigned scores a and b in the statistics $T_Y^{(N)}$, $T_{YR}^{(N)}$ and $T_{YC}^{(N)}$ can also be taken random. When such random vectors converge in probability to constants a and b, then the same limiting distributions apply. So $T_S^{(N)}$ can be considered as a special case of $T_Y^{(N)}$. Other choices of a and b leading to tests which are (in some sense) asymptotically most stringent against quadrant dependent alternatives are considered by SNIJDERS (1979, p. 216).

CHAPTER 4

ORDERINGS FOR POSITIVE DEPENDENCE

4.1. DEFINITIONS AND BASIC PROPERTIES

This section introduces some partial orderings for bivariate distributions which order distributions according to positive dependence. Consider two arbitrary (empirical or underlying) distribution functions F and F'. Let (X,Y) and (X',Y') be pairs of random variables with distributions F and F' respectively. Throughout this chapter, the *support* X of the marginal distribution F_r of X is defined by $X = \{x \in (-\infty,+\infty] : F_r(x') < F_r(x)$ for all $x' < x\}$. This definition of support is unusual, but it is essential that F_r is strictly increasing on X. The supports Y, X' and Y' of the marginal distributions of Y, X' and Y' are similarly defined.

Although several intuitive formulations of "more dependent" can be given, not much attention has been paid in the literature to make them precise. (An ordering according to a single measure of dependence can be misleading: bivariate dependencies are generally too complex to be represented by one single number.) One suggestion for an ordering "more dependent" for probability tables has actually already been given in the lines preceding Theorem 2.1.5. This ordering can easily be formulated for arbitrary bivariate distributions, but it is intuitively clear that it orders distributions according to dependence discarding its sign or form (e.g. negative dependence is "more dependent" than independence). WHITT (1982) discusses orderings for multivariate distributions which are related to the so called HPK inequality (see also EATON, 1982, formula (3.6) or KARLIN and RINOTT, 1980, formula (1.18)). These orderings order distributions according to positive dependence, but they seem to be quite strong because they do not order the standard bivariate normal distributions. The ordering of KOWALCZYK and PLESZCZYŃSKA (1977) is too weak for our purposes.

An ordering much weaker than that induced by the HPK inequality but stronger than the one of Kowalczyk and Pleszczyńska is introduced in

TCHEN (1976) (see also MARSHALL and OLKIN, 1979, p. 381) and independently of Tchen also in CAMBANIS, SIMONS and STOUT (1976). This ordering is of more interest to us and is defined as follows.

DEFINITION 4.1.1. The distribution F' is said to be *more concordant* than F, denoted by $F' \overset{c}{\geq} F$ or by $(X',Y') \overset{c}{\geq} (X,Y)$, when $F'_r(x) = F_r(x)$, $F'_c(y) = F_c(y)$ and $F'(x,y) \geq F(x,y)$ for all $x,y \in \mathbb{R}$.

This ordering for positive dependence, in particular its extension given in Definition 4.1.2 below, has some desirable properties. For example, it is a partial ordering, it is invariant under strictly increasing transformations of the marginals, it arises naturally in many examples and most well-known measures of dependence preserve the ordering. But still this ordering is unsatisfactory because the order preserving property does not carry over to samples in any useful sense. Therefore, a somewhat stronger ordering called *more associated* is introduced (Definition 4.1.4) which actually is stochastically preserved in samples by most measures of positive dependence whenever it is present between underlying distributions. The name "more associated" is explained by Proposition 4.1.2 which shows that each pair of variables which is more associated than an independent pair is *associated* in the sense of ESARY, PROSCHAN and WALKUP (1967). Furthermore, two interesting special cases of the ordering "more associated" are discussed. Definitions and properties below are frequently formulated in terms of random variables, but they only concern their distributions.

A disadvantage of the ordering $\overset{c}{\geq}$ is that it is only defined for distributions with the same marginals. The following extension imposes less restrictive assumptions on the marginals.

DEFINITION 4.1.2. The pair (X',Y') is said to be *more quadrant dependent* than (X,Y), denoted by $(X',Y') \overset{q}{\geq} (X,Y)$ or by $F' \overset{q}{\geq} F$, when $(F'_r(X'),F'_c(Y')) \overset{c}{\geq} (F_r(X),F_c(Y))$.

This ordering gives rise to the following equivalence relation.

DEFINITION 4.1.3. The pairs (X',Y') and (X,Y) are said to be *equally dependent*, denoted $(X',Y') \overset{d}{=} (X,Y)$ or $F' \overset{d}{=} F$, when $(F'_r(X'),F'_c(Y')) \sim (F_r(X),F_c(Y))$.

It is easily verified that the relation $\overset{d}{=}$ is reflexive, symmetric and

transitive and therefore actually is an equivalence relation. Furthermore, it can be shown that the ordering $\overset{q}{\geq}$ is a partial ordering (i.e. it is reflexive, transitive and anti-symmetric) of the equivalence classes defined by $\overset{d}{=}$ on the class of all bivariate distributions.

Obviously, two distributions can only be compared by means of $\overset{q}{\geq}$ or $\overset{d}{=}$ when they are elements of the same class F of bivariate distributions with marginals satisfying

$$(4.1.1) \qquad (X',Y') \sim F' \in F, \quad (X,Y) \sim F \in F \Rightarrow F'_r(X') \sim F_r(X) \text{ and } F'_c(Y') \sim F_c(Y).$$

Examples of such classes F are F = {bivariate distributions with continuous marginal distribution functions}, F = {empirical bivariate distribution functions based on N observations without ties} and F = {bivariate distribution functions with given marginals}.

PROPOSITION 4.1.1. $(X,Y) \overset{d}{=} (X',Y')$ *iff there exist strictly increasing functions* $\nu_1 : X \to X'$ *and* $\nu_2 : Y \to Y'$ *such that* $(X',Y') \sim (\nu_1(X),\nu_2(Y))$.

PROOF. First consider the if-part. Let $\nu_1^- : X' \to X$ and $\nu_2 : Y' \to Y$ be the inverses of ν_1 and ν_2 respectively. Since ν_1^- is strictly increasing on X' it follows from $X' \sim \nu_1(X)$ that $F'_r(x) = F_r(\nu_1^-(x))$ for all $x \in X'$. Similarly, $F'_c(y) = F_c(\nu_2^-(y))$ for all $y \in Y'$. Furthermore, $(X,Y) \sim (\nu_1^-(X'),\nu_2^-(Y'))$. Hence $(F_r(X),F_c(Y)) \sim (F_r(\nu_1^-(X')),F_c(\nu_2^-(Y'))) \sim (F'_r(X'),F'_c(Y'))$. Conversely, assume $(F'_r(X'),F'_c(Y')) \sim (F_r(X),F_c(Y))$. Let U denote the support of the distribution of $F'_r(X')$ and $F_r(X)$. Define $F'^-_r : U \to X'$ by $F'^-_r(u) = \inf\{x : F'_r(x) \geq u\}$. Obviously, the range of F'^-_r is X'. Since X' is the support of the distribution F'_r, it follows that $F'^-_r(F'_r(X')) \sim X'$. Furthermore, since X is the support of the distribution F_r, the function F_r is strictly increasing on X. Similarly, F'^-_r is strictly increasing on U. Hence the function $\nu_1 : X \to X'$ defined by $\nu_1(x) = F'^-_r(F_r(x))$ is strictly increasing on X. By the same arguments it follows that a similar defined function $\nu_2 = F'^-_c(F_c)$ is strictly increasing on Y. Moreover, $(X',Y') \sim (F'^-_r(F_r(X)),F'^-_c(F_c(Y))) \sim (\nu_1(X),\nu_2(Y))$. □

This proposition shows that the equivalence classes defined by $\overset{d}{=}$ consist of distributions which can be transformed into each other by strictly increasing transformations of the marginals. Moreover, it follows from this proposition that the ordering $\overset{q}{\geq}$ is invariant under strictly increasing transformations of the marginals.

The results above show that the ordering "more quadrant dependent" has some desirable elementary properties. The next section considers measures of positive dependence which preserve this ordering. But let's first turn to a somewhat stronger ordering for positive dependence on which this chapter is mainly focussed.

DEFINITION 4.1.4. The pair (X',Y') is said to be *more associated* than (X,Y), denoted $(X',Y') \overset{a}{\geq} (X,Y)$ or $F' \overset{a}{\geq} F$, when there exist functions $\kappa_1 : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}'$ and $\kappa_2 : \mathcal{X} \times \mathcal{Y} \to \mathcal{Y}'$ such that for all $x_1,x_2 \in \mathcal{X}$ and $y_1,y_2 \in \mathcal{Y}$

$$(4.1.2) \qquad x_1 \leq x_2,\ y_1 \leq y_2 \Rightarrow \kappa_1(x_1,y_1) \leq \kappa_1(x_2,y_2),\ \kappa_2(x_1,y_1) \leq \kappa_2(x_2,y_2)$$

$$(4.1.3) \qquad \kappa_1(x_1,y_1) < \kappa_1(x_2,y_2),\ \kappa_2(x_1,y_1) > \kappa_2(x_2,y_2) \Rightarrow x_1 < x_2,\ y_1 > y_2$$

$$(4.1.4) \qquad (X',Y') \sim (\kappa_1(X,Y),\kappa_2(X,Y)).$$

Occasionally, the briefer notation $(X',Y') \sim \kappa(X,Y)$ is used instead of (4.1.4), where κ is the vector valued function $\kappa = (\kappa_1,\kappa_2)$. Of course (4.1.2) means that κ_1 and κ_2 are increasing in both arguments and (4.1.3) excludes reflection about the diagonal. Reflection is excluded in order to make the ordering compatible with $\overset{q}{\geq}$ and $\overset{c}{\geq}$. Note that $(X',Y') \overset{q}{\geq} (X,Y)$ does not imply $(X',Y') \overset{q}{\geq} (Y,X)$.

Two special cases of the ordering $\overset{a}{\geq}$ are also of interest. First, in case where κ_1 satisfies $\kappa_1(x,y) = x$ for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$, the pair (X',Y') is said to be *more row regression dependent* than (X,Y), denoted $(X',Y') \overset{rr}{\geq} (X,Y)$. The second special case arises when $\kappa_2(x,y) = y$ for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and is called *more column regression dependent* (denoted $\overset{cr}{\geq}$). Clearly, in both cases (4.1.2) implies (4.1.3).

The names of these orderings, and also that of $\overset{q}{\geq}$, are borrowed from the form of dependence of the class of distributions which are $\overset{a}{\geq}$, $\overset{rr}{\geq}$, $\overset{cr}{\geq}$ and $\overset{q}{\geq}$ than independence; see Proposition 4.1.2 below. Let's first review the definitions of these forms of dependence. The pair (X,Y) is called *associated* by ESARY, PROSCHAN and WALKUP (1967) if $Cov\,(\kappa_1(X,Y),\kappa_2(X,Y)) \geq 0$ for all pairs of functions κ_1 and κ_2 which are increasing in both arguments and for which the covariance exists. Furthermore, the pair (X,Y) is called *quadrant dependent* by LEHMANN (1966) when their joint distribution function satisfies $F(x,y) \geq F_r(x)F_c(y)$ for all $x,y \in \mathbb{R}$. The following well-known

relation between these forms of dependence and the forms considered in Section 3.2 is proved in ESARY et al. (1967) and TONG (1980, pp. 80,86).

$$TP_2\text{-dep.} \Rightarrow DR_1\text{-dep.} \nearrow \begin{matrix} RR_1\text{-dep.} \\ CR_1\text{-dep.} \end{matrix} \searrow \text{association} \Rightarrow \text{quadrant dep.}$$

PROPOSITION 4.1.2.

(4.1.5) $(X',Y') \overset{q}{\geq} (X,Y)$, (X,Y) *is quadrant dep.* $\Rightarrow$ (X',Y') *is quadrant dep.*

(4.1.6) $(X',Y') \overset{a}{\geq} (X,Y)$, (X,Y) *is associated* $\Rightarrow$ (X',Y') *is associated*

(4.1.7) $(X',Y') \overset{rr}{\geq} (X,Y)$, (X,Y) *is* RR_1*-dep.* $\Rightarrow$ (X',Y') *is* RR_1*-dep.*

(4.1.8) $(X',Y') \overset{cr}{\geq} (X,Y)$, (X,Y) *is* CR_1*-dep.* $\Rightarrow$ (X',Y') *is* CR_1*-dep.*

Moreover, let X *and* Y *be independent then*

(4.1.9) $(X',Y') \overset{q}{\geq} (X,Y) \Leftrightarrow (X',Y')$ *is quadrant-dep.,*

and if in addition X *and* Y *have continuous distribution functions, then also*

(4.1.10) $(X',Y') \overset{rr}{\geq} (X,Y) \Leftrightarrow (X',Y')$ *is* RR_1*-dep.*

(4.1.11) $(X',Y') \overset{cr}{\geq} (X,Y) \Leftrightarrow (X',Y')$ *is* CR_1*-dep.*

PROOF. The proof of (4.1.5) is trivial and (4.1.6) follows directly from property P_4 in ESARY et al. (1967). The result (4.1.7) can be verified as follows. Let $\kappa_2 : X \times Y \to Y'$ be an increasing function such that $(X',Y') \sim (X,\kappa_2(X,Y))$. Then for $x_1 < x_2$ $(x_1,x_2 \in X)$,

$$P\{Y' \leq y | X' = x_1\} = P\{\kappa_2(X,Y) \leq y | X = x_1\} = P\{\kappa_2(x_1,Y) \leq y | X = x_1\}$$

$$\geq P\{\kappa_2(x_1,Y) \leq y | X = x_2\} \geq P\{\kappa_2(x_2,Y) \leq y | X = x_2\} = P\{Y' \leq y | X' = x_2\}.$$

Hence the family of conditional distributions of $Y'|X' = x$ is stochastically increasing in x. The proof of (4.1.8) is similar.

Since independence is a special case of quadrant, RR_1 and CR_1-dependence the implications from left to right in (4.1.9), (4.1.10) and (4.1.11) follow from the results above. The implication from right to left in (4.1.9) is trivial. The similar implication in (4.1.10) is proved as follows. Let X and Y be independent having continuous distribution functions and let (X',Y') be RR_1-dependent. Thus the conditional distribution function $F'_{c|r}(\cdot|x)$ of $Y'|X' = x$ is decreasing in x. Define in the function

$\kappa_2 : X \times Y \to Y'$ by

$$\kappa_2(x,y) = \inf\{\eta : F'_{c|r}(\eta|x) \geq F_c(y)\}.$$

Obviously, the range of κ_2 is Y' and, moreover, κ_2 is increasing in both arguments. Furthermore,

$$\kappa_2(x,y) \leq \theta \Leftrightarrow \inf\{\eta : F'_{c|r}(\eta|x) \geq F_c(y)\} \leq \theta \Leftrightarrow F'_{c|r}(\theta|x) \geq F_c(y).$$

Therefore,

$$P\{\kappa_2(X,Y) \leq y | X = x\} = P\{\kappa_2(x,Y) \leq y\} = P\{F_c(Y) \leq F'_{c|r}(y|x)\} = F'_{c|r}(y|x)$$

where the last equality holds because continuity of F_c implies that $F_c(Y)$ is uniformly distributed on the unit interval. Thus the distribution function of $(X,\kappa_2(X,Y))$ is F' and hence $(X',Y') \overset{rr}{\geq} (X,Y)$. Note that (4.1.3) follows since $\kappa_1(x,y) \equiv x$ and κ_2 is increasing. The proof of (4.1.11) is similar. □

Note that (X',Y') is more concordant than independence iff (X',Y') is quadrant dependent. Thus by the same arguments, $\overset{c}{\geq}$ could in fact also be called "more quadrant dependent". By Theorem 2.1 in ESARY et al. (1967), an independent pair is associated and hence by (4.1.6), (X',Y') is more associated than independence implies that (X',Y') is associated. However, counterexamples can be constructed showing that even in the continuous case a converse of this result does not hold. But (4.1.10) and (4.1.11) suggest that the class of associated distributions is not much larger than the class of distributions which is $\overset{a}{\geq}$ than independence.

The ordering $\overset{a}{\geq}$ has, just as $\overset{q}{\geq}$, the desirable property that it is invariant under strictly increasing transformations of the marginals.

PROPOSITION 4.1.3. *Let* $(X',Y') \overset{d}{=} (U',V')$ *and* $(X,Y) \overset{d}{=} (U,V)$, *then* $(X',Y') \overset{a}{\geq} (X,Y)$ *iff* $(U',V') \overset{a}{\geq} (U,V)$.

PROOF. Using Proposition 4.1.1 and the fact that the composition of increasing functions is again increasing, the proof is straightforward. □

So the ordering $\overset{a}{\geq}$ also gives rise to the equivalence classes defined by $\overset{d}{=}$. It can be verified that $\overset{a}{\geq}$ is reflexive and transitive. Moreover, since Theorem 4.1.4 below shows that (under some conditions on the marginals) the ordering $\overset{a}{\geq}$ is stronger than $\overset{q}{\geq}$, anti-symmetry of $\overset{a}{\geq}$ follows from anti-symmetry of $\overset{q}{\geq}$. Hence, $\overset{a}{\geq}$ *is a partial ordering on any subclass of*

bivariate distributions F *satisfying* (4.1.1) *of all equivalence classes defined by* $\overset{d}{=}$. (On the class of all bivariate distributions, anti-symmetry does not hold and $\overset{a}{\geq}$ is only a preordering.)

THEOREM 4.1.4. *Let* F *be a class of bivariate distribution functions with marginals satisfying* (4.1.1). *Then for* $F', F \in F$

$$F' \overset{a}{\geq} F \Rightarrow F' \overset{q}{\geq} F.$$

PROOF. By Proposition 4.1.3, $(X',Y') \overset{a}{\geq} (X,Y)$ iff $(F'_r(X'),F'_c(Y')) \overset{a}{\geq} (F_r(X),F_c(Y))$ and hence it is sufficient to consider distributions with the same marginals and prove $F' \overset{a}{\geq} F \Rightarrow F' \overset{c}{\geq} F$. So let $X' \sim X$, $Y' \sim Y$ and $\kappa_1 : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ and $\kappa_2 : \mathcal{X} \times \mathcal{Y} \to \mathcal{Y}$ be such that (4.1.2), (4.1.3) and (4.1.4) hold. Choose $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. By (4.1.2) and (4.1.4) it follows

$$(4.1.12) \qquad P\{X \leq x, Y \leq y\} \leq P\{X' \leq \kappa_1(x,y), Y' \leq \kappa_2(x,y)\},$$

$$(4.1.13) \qquad P\{X \geq x, Y \geq y\} \leq P\{X' \geq \kappa_1(x,y), Y' \geq \kappa_2(x,y)\},$$

and by (4.1.3) and (4.1.4)

$$(4.1.14) \qquad P\{X < x, Y > y\} \geq P\{X' < \kappa_1(x,y), Y' > \kappa_2(x,y)\},$$

$$(4.1.15) \qquad P\{X > x, Y < y\} \geq P\{X' > \kappa_1(x,y), Y' < \kappa_2(x,y)\}.$$

Next consider the following four cases.

(i) Suppose $\kappa_1(x,y) \leq x$ and $\kappa_2(x,y) \leq y$, then by (4.1.12)

$$(4.1.16) \qquad P\{X \leq x, Y \leq y\} \leq P\{X' \leq x, Y' \leq y\}.$$

(ii) Suppose $\kappa_1(x,y) > x$ and $\kappa_2(x,y) > y$, then by (4.1.13)

$$(4.1.17) \qquad P\{X > x, Y > y\} \leq P\{X' > x, Y' > y\}.$$

(iii) Suppose $\kappa_1(x,y) > x$ and $\kappa_2(x,y) \leq y$, then by (4.1.14)

$$(4.1.18) \qquad P\{X \leq x, Y > y\} \geq P\{X' \leq x, Y' > y\}.$$

(iv) Suppose $\kappa_1(x,y) \leq x$ and $\kappa_2(x,y) > y$, then by (4.1.15)

$$(4.1.19) \qquad P\{X > x, Y \leq y\} \geq P\{X' > x, Y' \leq y\}.$$

Since $X' \sim X$ and $Y' \sim Y$, it follows that each of the formula's (4.1.17), (4.1.18) and (4.1.19) implies (4.1.16). Hence for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$,

(4.1.16) holds, implying $F' \overset{c}{\geq} F$. □

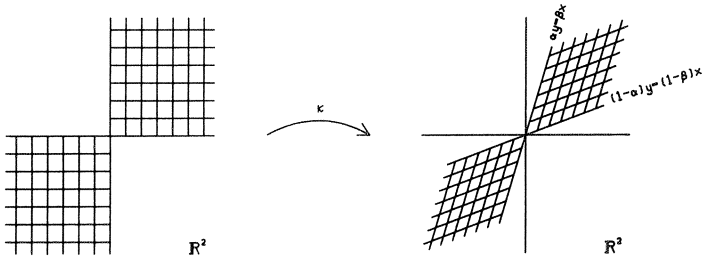
It is obvious that the converse of this theorem does not hold.

This section closes with some examples of the ordering $\overset{a}{\geq}$.

EXAMPLE 4.1.1. This first example gives an intuitive idea of the ordering $\overset{a}{\geq}$ in terms of transformations of mass. Consider the case where (X',Y') is a linear transform, with positive coefficients, of (X,Y). Since the ordering $\overset{a}{\geq}$ is invariant under location and scale transformations it is no loss of generality to assume that

(4.1.20) $\quad (X',Y') \sim ((1-\alpha)X + \alpha Y, (1-\beta)X + \beta Y)$

for $0 \leq \alpha \leq \beta \leq 1$. The figure below illustrates that the mass of the distribution F in the first and third quadrant of $\mathbb{R}^2$ is mapped by the function $\kappa(x,y) = ((1-\alpha)x + \alpha y, (1-\beta)x + \beta y)$ onto the region between the lines $(1-\alpha)y = (1-\beta)x$ and $\alpha y = \beta x$.



Clearly, the mass of F' is more concentrated around the line y = x than that of F.

Special cases of the transform (4.1.20) give rise to one-parameter families of bivariate distributions $\{F^{(\alpha)}\}_{\alpha \in A}$, e.g.

$$(X^{(\alpha)},Y^{(\alpha)}) \sim ((1-\alpha)X + \alpha Y, Y) \qquad \text{for } \alpha \in A = [0,1].$$

For this family it is easily shown that $0 \leq \alpha \leq \alpha' \leq 1$ implies $F^{(\alpha')} \overset{cr}{\geq} F^{(\alpha)}$. Another example is the family

$$(X^{(\alpha)},Y^{(\alpha)}) \sim ((1-\alpha)X + \alpha Y, \alpha X + (1-\alpha)Y) \text{ for } \alpha \in A = [0,\tfrac{1}{2}],$$

which is a special case of a family considered by KONIJN (1956); see also RUYMGAART (1973). Elementary calculations yield for this family that $0 \leq \alpha \leq \alpha' \leq \frac{1}{2}$ iff $F^{(\alpha')} \overset{a}{\geq} F^{(\alpha)}$.

Linear transforms (4.1.20) form perhaps the most important examples of the ordering $\overset{a}{\geq}$ for continuous distributions. Unfortunately, families $\{F^{(\alpha)}\}_{\alpha>0}$ generated by trivariate reduction

$$(X^{(\alpha)},Y^{(\alpha)}) \sim (Z_1+\alpha Z_3, Z_2+\alpha Z_3) \qquad \text{for } \alpha > 0,$$

where Z_1, Z_2 and Z_3 are independent, generally do not give an example of the ordering $\overset{a}{\geq}$.

EXAMPLE 4.1.2. Let (X,Y) and (X',Y') have bivariate normal distributions with correlation parameters ρ and ρ' respectively. Then

$$(4.1.21) \qquad (X',Y') \overset{a}{\geq} (X,Y) \qquad \text{iff} \quad -1 < \rho \leq \rho' \leq 1.$$

In fact, the stronger orderings $\overset{rr}{\geq}$ and $\overset{cr}{\geq}$ hold for the standardized distributions. Moreover, these orderings hold in any elliptical family. Recall that *elliptical distributions* have densities of the form

$$\det(\Sigma)^{-\frac{1}{2}}\, h\left(\begin{pmatrix}x-\xi\\ y-\eta\end{pmatrix}^{\top}\Sigma^{-1}\begin{pmatrix}x-\xi\\ y-\eta\end{pmatrix}\right),$$

where Σ is a symmetric strictly positive definite 2×2 matrix and the function h determines the family (cf. MUIRHEAD, 1982, p. 34). Denote this distribution by $E(h;(\xi,\eta),\Sigma)$. The correlation for this elliptical distribution is $\rho = \sigma_{12}/(\sigma_{11}\,\sigma_{22})^{\frac{1}{2}}$, where σ_{ij} is the (i,j)-th element of Σ. Independence is equivalent to $\rho = 0$ only in normal distributions. It is easily verified that elliptical distributions are closed under linear transformations; i.e. for any 2×2 matrix K the linear transform $(X,Y)K \sim E(h;(\xi,\eta)K,K\Sigma K^{\top})$ whenever $(X,Y) \sim E(h;(\xi,\eta),\Sigma)$. Hence for $(X',Y') \sim E(h;(0,0),\Sigma')$ and $(X,Y) \sim E(h;(0,0),\Sigma)$, where

$$\Sigma' = \begin{pmatrix}1 & \rho'\\ \rho' & 1\end{pmatrix} \qquad \text{and} \qquad \Sigma = \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix},$$

it follows by putting $\kappa_2(x,y) \equiv y$ and

$$\kappa_1(x,y) \equiv ((1-\rho'^2)/(1-\rho^2))^{\frac{1}{2}}x + (\rho' - \rho((1-\rho'^2)/(1-\rho^2))^{\frac{1}{2}})y$$

that $(X',Y') \overset{cr}{\geq} (X,Y)$ iff $1 < \rho \leq \rho'$. It is easily seen that (4.1.21) holds for elliptical distributions with arbitrary expectations and covariance matrices.

The following proposition is merely a characterization of the

orderings $\overset{a}{\geq}$ and $\overset{q}{\geq}$ for empirical distributions. Recall that two pairs of observations (x_1,y_1) and (x_2,y_2) are called *concordant* when $(x_2 - x_1)(y_2 - y_1) > 0$ and *disconcordant* when $(x_2 - x_1)(y_2 - y_1) < 0$. Furthermore, they *tie in the row variable* when $x_1 = x_2$ and *tie in the column variable* when $y_1 = y_2$.

PROPOSITION 4.1.5. *Let* $\hat{F}^{(N)}$ *and* $\hat{F}'^{(N)}$ *be two empirical distribution functions based on two sets of* N *observations* $(x_1,y_1),\ldots,(x_N,y_N)$ *and* $(x'_1,y'_1),\ldots,(x'_N,y'_N)$ *respectively. Suppose there are no ties in both sets of observations. Then*

(i) $\hat{F}'^{(N)} \overset{a}{\geq} \hat{F}^{(N)}$ *iff for some permutation* $(\pi_1,\ldots,\pi_N)$ *of* $(1,\ldots,N)$

$$x_i < x_j, y_i < y_j \Rightarrow x'_{\pi_i} < x'_{\pi_j},\ y'_{\pi_i} < y'_{\pi_j} \tag{4.1.22}$$

and

$$x'_{\pi_i} < x'_{\pi_j},\ y'_{\pi_i} > y'_{\pi_j} \Rightarrow x_i < x_j,\ y_i > y_j. \tag{4.1.23}$$

(ii) $\hat{F}'^{(N)} \overset{q}{\geq} \hat{F}^{(N)}$ *iff there exists a permutation* $(\pi_1,\ldots,\pi_N)$ *of* $(1,\ldots,N)$ *such that for* $i = 1,\ldots,N$ *the number of observations in the first set which is concordant with* (x_i,y_i) *is less than or equal to the number of observations in the second set concordant with* (x'_{π_i},y'_{π_i}).

PROOF. Let $X = \{x_1,\ldots,x_N\}$, $Y = \{y_1,\ldots,y_N\}$, $X' = \{x'_1,\ldots,x'_N\}$ and $Y' = \{y'_1,\ldots,y'_N\}$.

(i) First suppose $\hat{F}'^{(N)} \overset{a}{\geq} \hat{F}^{(N)}$. Let $\kappa_1 : X \times Y \to X'$ and $\kappa_2 : X \times Y \to Y'$ be functions satisfying (4.1.2), (4.1.3) and (4.1.4). Clearly (4.1.4) implies that there exists a permutation $(\pi_1,\ldots,\pi_N)$ of $(1,\ldots,N)$ such that $x'_{\pi_i} = \kappa_1(x_i,y_i)$ and $y'_{\pi_i} = \kappa_2(x_i,y_i)$ for $i = 1,\ldots,N$. Since there are no tied observations (4.1.2) and (4.1.3) imply (4.1.22) and (4.1.23).

Conversely, suppose there exists a permutation $(\pi_1,\ldots,\pi_N)$ of $(1,\ldots,N)$ such that (4.1.22) and (4.1.23) hold. Define functions κ_1 and κ_2 in the N points of the first set of observations by $\kappa_1(x_i,y_i) = x'_{\pi_i}$ and $\kappa_2(x_i,y_i) = y'_{\pi_i}$ for $i = 1,\ldots,N$. Furthermore, let $x_0 = \inf\{X\}$ and $y_0 = \inf\{Y\}$ and define $\kappa_1(x_0,y_0) = \inf\{X'\}$ and $\kappa_2(x_0,y_0) = \inf\{Y'\}$. Extend the functions κ_1 and κ_2 in other points $(\xi,\eta) \in X \times Y$ by

$$\kappa_1(\xi,\eta) = \sup\{\kappa_1(x_i,y_i) : x_i \leq \xi,\ y_i \leq \eta,\ 0 \leq i \leq N\},$$

$$\kappa_2(\xi,\eta) = \sup\{\kappa_2(x_i,y_i) : x_i \leq \xi,\ y_i \leq \eta,\ 0 \leq i \leq N\}.$$

In order to verify (4.1.2) let $\xi_1 \leq \xi_2$ and $\eta_1 \leq \eta_2$, then

$\{(x_i,y_i): x_i \le \xi_1, y_i \le \eta_1,\ 0 \le i \le N\} \subset \{(x_i,y_i): x_i \le \xi_2, y_i \le \eta_2,\ 0 \le i \le N\}$ and hence $\kappa_1(\xi_1,\eta_1) \le \kappa_1(\xi_2,\eta_2)$ and $\kappa_2(\xi_1,\eta_1) \le \kappa_2(\xi_2,\eta_2)$. Furthermore, verify (4.1.3) as follows. Let $\xi_1,\xi_2 \in X$; $\eta_1,\eta_2 \in Y$ such that $\kappa_1(\xi_1,\eta_1) < \kappa_1(\xi_2,\eta_2)$ and $\kappa_2(\xi_1,\eta_1) > \kappa_2(\xi_2,\eta_2)$. Then for the observation (x_i,y_i) such that $x_i \le \xi_1$, $y_i \le \eta_1$ and $\kappa_2(x_i,y_i) = \kappa_2(\xi_1,\eta_1)$ and for the observation (x_j,y_j) such that $x_j \le \xi_2$, $y_j \le \eta_2$ and $\kappa_1(x_j,y_j) = k_1(\xi_2,\eta_2)$ it follows that $\kappa_1(x_i,y_i) < \kappa_1(x_j,y_j)$ and $\kappa_2(x_i,y_i) > \kappa_2(x_j,y_j)$. By (4.1.23) this implies $\xi_1 < \xi_2$ and $\eta_1 > \eta_2$.

(ii) The proof of the second part of the proposition follows directly from the definition of $\overset{q}{\geq}$. □

4.2. MEASURES OF DEPENDENCE PRESERVING THE ORDERINGS

It is interesting to know which measures of positive dependence preserve the orderings $\overset{a}{\geq}$, $\overset{q}{\geq}$ and $\overset{c}{\geq}$. Measures of interest can be written as real valued functionals of the (empirical) distribution function. Such a measure ϕ is said to preserve $\overset{a}{\geq}$ when

$$F' \overset{a}{\geq} F \Rightarrow \phi(F') \geq \phi(F).$$

TCHEN (1976) and CAMBANIS et al. (1976) considered this problem for the ordering $\overset{c}{\geq}$; their result is formulated in Theorem 4.2.1 below.

DEFINITION 4.2.1. A real valued function ψ defined on $\mathbb{R}^2$ is called *lattice-superadditive* (L-superadditive) when

$$x_1 < x_2,\ y_1 < y_2 \Rightarrow \psi(x_1,y_1) + \psi(x_2,y_2) \geq \psi(x_1,y_2) + \psi(x_2,y_1).$$

Note that ψ is L-superadditive iff $\exp(\psi)$ is TP_2. So examples of L-superadditive functions are easily derived from examples in Section 3.1.

THEOREM 4.2.1. *Let* $F' \overset{c}{\geq} F$ *and* ψ *be right continuous and L-superadditive on* $\mathbb{R}^2$, *then*

$$\iint_{\mathbb{R}^2} \psi(x,y)\,dF'(x,y) \geq \iint_{\mathbb{R}^2} \psi(x,y)\,dF(x,y) \tag{4.2.1}$$

provided the integrals exist and either of the following conditions is satisfied:

(i) ψ *is symmetric and the expectations* $\int \psi(x,x)\,dF_r(x)$ *and* $\int \psi(y,y)\,dF_c(y)$ *are finite,*

(ii) *the expectations* $\int \psi(x,y_0)dF_r(x)$ *and* $\int \psi(x_0,y)dF_c(y)$ *are finite for some* y_0 *and* x_0.

Conversely, if the marginal distributions of F' *and* F *are the same and* (4.2.1) *holds for all L-superadditive functions* ψ, *then* $F' \overset{c}{\geq} F$.

PROOF. The first part of the theorem is proved in CAMBANIS et al. (1976). Its converse follows by noting that the indicator function of the set $(-\infty,x] \times (-\infty,y]$ is L-superadditive. □

The conditions (i) and (ii) of this theorem can be weakened and perhaps they are not necessary at all. These conditions or their weaker versions are satisfied in all cases of interest; e.g. ψ bounded, $\psi(x,y) \equiv xy$ or any example in which the support of F and F' is bounded (cf. CAMBANIS et al., 1976).

PROPOSITION 4.2.2. *Let* $\mathcal{F}$ *be a class of bivariate distribution functions satisfying* (4.1.1) *and let* ϕ *be a functional which preserves* $\overset{c}{\geq}$ *and is invariant under strictly increasing transformations of the marginals. Then* ϕ *preserves* $\overset{a}{\geq}$ *and* $\overset{q}{\geq}$ *on* $\mathcal{F}$, *i.e.*

$$F' \overset{q}{\geq} F; \quad F',F \in \mathcal{F} \Rightarrow \phi(F') \geq \phi(F)$$

and a similar implication a fortiori holds for $\overset{a}{\geq}$.

PROOF. Follows directly from Propositions 4.1.1, 4.1.3 and Theorem 4.1.4. □

By Proposition 4.1.3 the condition that ϕ is invariant under strictly increasing transformations of the marginals is necessary also. For example, Pearson's product moment correlation preserves $\overset{c}{\geq}$, but it does not preserve $\overset{a}{\geq}$ nor $\overset{q}{\geq}$ because it is not invariant under strictly increasing transformations of the marginals. It is obvious that rank statistics and measures based on canonical correlations have this invariance property.

EXAMPLE 4.2.1. *Linear rank statistics.*

Let J be a real valued function defined on $(0,1] \times (0,1]$ and consider measures of the form

$$\phi(F) = \iint_{\mathbb{R}^2} J(F_r(x),F_c(y))dF(x,y). \tag{4.2.2}$$

For an empirical distribution function $\hat{F}^{(N)}$ based on N observations $(x_1,y_1),\ldots,(x_N,y_N)$ without ties, (4.2.2) equals

$$(4.2.3) \qquad \phi(\hat{F}^{(N)}) = \sum_{i=1}^{N} N^{-1} J(N^{-1}R_i, N^{-1}Q_i)$$

where R_i is the rank of x_i among $x_1,\ldots,x_N$ and Q_i is the rank of y_i among $y_1,\ldots,y_N$. Statistics of the form (4.2.3) are called linear rank statistics. For underlying distributions F the measure (4.2.2) may be interpreted as a population value. The function J is called the (generating) score function. In practice the score function usually depends on the sample size N and is derived from the generating score function in either of the following two ways. The frequently used approximate score function is defined by

$$(4.2.4) \qquad J_N(u,v) = J(i/(N+1), j/(N+1))$$

for (u,v) in $((i-1)/N, i/N] \times ((j-1)/N, j/N]$, $i = 1,\ldots,N$ and $j = 1,\ldots,N$. The exact score function is obtained by using

$$J_N(u,v) = EJ(U_{i:N}, V_{j:N})$$

instead of (4.2.4), where $U_{i:N}$ denotes the i-th order statistic in a random sample of size N from the uniform distribution on (0,1) and $V_{i:N}$ is an independent copy of $U_{i:N}$.

The (generating, approximate or exact) score functions of most well-known rank statistics are of product type $J(u,v) = f(u)g(v)$, where f and g are increasing, right continuous and integrable functions. Such score functions are L-superadditive and satisfy the additional condition (ii) of Theorem 4.2.1 (because $\psi(x,y) \equiv xy$ satisfies this condition). Moreover, linear rank statistics are invariant under strictly increasing transformations of the marginals. Therefore, the following specific examples of linear rank statistics preserve $\overset{c}{\geq}$ and preserve $\overset{a}{\geq}$ and $\overset{q}{\geq}$ on classes of distributions F satisfying (4.1.1).

Spearman's rank correlation rho has the generating score function

$$(4.2.5) \qquad J(u,v) = 3(2u - 1)(2v - 1).$$

Using the approximate score function in the empirical case (without ties), this statistic equals the sample product moment correlation of the ranks.

Fisher-Yates' normal score statistic has the generating score function

$$J(u,v) = \Phi^{-}(u)\Phi^{-}(v),$$

where Φ^{-} is the inverse of the standard normal distribution function. In

the empirical case the Fisher-Yates normal score statistic is computed with the exact score function, whereas the approximate score function produces the *vander Waerden statistic*.

The generating score function of the *quadrant statistic* is given by

$$J(u,v) = \text{sign}\,(u - \tfrac{1}{2})\ \text{sign}\,(v - \tfrac{1}{2})$$

where sign (u) = -1,0,1 as $u < 0$, $= 0$ or > 0.

EXAMPLE 4.2.2. *Non-linear rank statistics.*

Consider measures of the form

$$\phi(F) = \iint_{\mathbb{R}^2} K(F(x,y))dF(x,y). \tag{4.2.6}$$

Since any bivariate distribution function F is L-superadditive it follows from MARSHALL and OLKIN (1979, p. 151) that K(F) is L-superadditive when K is increasing and convex. So if in addition K is also right continuous and either one of the conditions (i) or (ii) of Theorem 4.2.1 is satisfied, then $F' \overset{c}{\geq} F$ implies $\iint K(F')dF' \geq \iint K(F')dF \geq \iint K(F)dF$. The last inequality holds because K is increasing. Hence under these conditions on K, the measure (4.2.6) preserves $\overset{c}{\geq}$ and preserves $\overset{a}{\geq}$ and $\overset{q}{\geq}$ on classes of distributions F satisfying (4.1.1).

Kendall's rank correlation tau is a rank statistic of the form (4.2.6) with

$$K(u) = 4u - 1. \tag{4.2.7}$$

For empirical distributions based on N observations (without ties) one usually takes

$$K_N(u) = (N-1)^{-1}(4Nu - N - 3). \tag{4.2.8}$$

Clearly, these functions K have the desired properties. Inserting (4.2.8) in (4.2.6) yields the well-known formula

$$\phi_K(\hat{F}^{(N)}) = (\#\{\text{concordant pairs}\} - \#\{\text{disconc. pairs}\})/\#\{\text{pairs}\},$$

provided there are no ties. Therefore it alternatively follows directly from Proposition 4.1.5 that Kendall's tau preserves $\overset{a}{\geq}$ and $\overset{q}{\geq}$ for empirical distribution functions without ties.

EXAMPLE 4.2.3. *Signed canonical correlation.*

The signed canonical correlation of a pair (X,Y) with distribution function

F is defined by

$$(4.2.9) \qquad \phi_{SC}(F) = \sup Corr\,(\varphi(X),\psi(Y))$$

where the supremum is taken over all increasing functions φ and ψ for which the correlation exists. It is easily verified that the signed canonical correlation preserves $\overset{c}{\geq}$ and preserves $\overset{q}{\geq}$ and $\overset{a}{\geq}$ on classes $\mathcal{F}$ satisfying (4.1.1). To this end let $F' \overset{c}{\geq} F$. Choose $\varepsilon > 0$ and let φ_1 and ψ_1 be increasing functions such that $\phi_{SC}(F) \leq Corr\,(\varphi_1(X),\psi_1(Y)) + \varepsilon$. Then by Theorem 4.2.1, $\iint \varphi_1\psi_1 dF' \geq \iint \varphi_1\psi_1 dF$. Since $\varphi_1(X') \sim \varphi_1(X)$ and $\psi_1(Y') \sim \psi_1(Y)$ it follows that $\phi_{SC}(F') \geq Corr\,(\varphi_1(X'),\psi_1(Y')) \geq \phi_{SC}(F) - \varepsilon$. Since ε is arbitrary, $\phi_{SC}(F') \geq \phi_{SC}(F)$. Application of Proposition 4.2.2 yields the desired result for $\overset{a}{\geq}$ and $\overset{q}{\geq}$.

These examples show that all familiar measures of positive dependence preserve $\overset{a}{\geq}$ and $\overset{q}{\geq}$. It follows directly from Example 4.1.2 that omnibus measures such as Pearson's χ^2 (2.1.8), the canonical correlation, the correlation ratio, the rank statistic $(\iint (F(x,y) - F_r(x)F_c(y))^2 dF(x,y))^{\frac{1}{2}}$ of HOEFFDING (1948), etc. do not preserve these orderings. But the canonical correlation actually preserves both orderings on the class of DR_1-dependent probability tables (distributions) satisfying (4.1.1), because by Theorem 3.3.2 the canonical correlation and the signed canonical correlation are equal for such distributions.

4.3. SAMPLING PROPERTIES OF THE ORDERING $\overset{a}{\geq}$

The main results formulated in this section show that the order preserving properties of the ordering $\overset{a}{\geq}$ given in the previous section carry over from population distributions to finite samples from these distributions. The ordering $\overset{q}{\geq}$ does not share this property with $\overset{a}{\geq}$. This is the principal reason for introducing $\overset{a}{\geq}$.

Let F and F' be arbitrary bivariate distribution functions and let $\hat{F}^{(N)}$ and $\hat{F}'^{(N)}$ be empirical distribution functions based on samples of N i.i.d. observations from F and F' respectively.

THEOREM 4.3.1. *Let ϕ be a measure which preserves $\overset{a}{\geq}$. Then*

$$(4.3.1) \qquad F' \overset{a}{\geq} F \Rightarrow P\{\phi(\hat{F}'^{(N)}) \geq t\} \geq P\{\phi(\hat{F}^{(N)}) \geq t\}$$

for all t *and* N; *i.e.* $\phi(\hat{F}'^{(N)})$ *is stochastically larger than* $\phi(\hat{F}^{(N)})$ *for all sample sizes* N.

PROOF. Let $(X_1,Y_1),\ldots,(X_N,Y_N)$ be a sample of N i.i.d. observations from F and let $\hat{F}$ be the corresponding empirical distribution function (the superscript (N) is suppressed). Since $F' \overset{a}{\geq} F$ there exist functions κ_1 and κ_2 such that $(\kappa_1(X_i,Y_i),\kappa_2(X_i,Y_i))$ for $i = 1,\ldots,N$ is a sample of N i.i.d observations from F'. Denote its empirical distribution function by $\hat{H}'$. Clearly, the empirical distribution function $\hat{F}'$ of any other sample of N i.i.d. observations from F' has the same distribution as $\hat{H}'$. Furthermore, it is obvious that $P\{\hat{H}' \overset{a}{\geq} \hat{F}\} = 1$, which implies $P\{\phi(\hat{H}') \geq \phi(\hat{F})\} = 1$. Hence for all t, $P\{\phi(\hat{F}) > t \geq \phi(\hat{H}')\} = 0$ and (4.3.1) follows. □

It is clear that similar results also holds for $\overset{rr}{\geq}$ and $\overset{cr}{\geq}$ and measures preserving these orderings. The measures discussed in the previous section preserve $\overset{a}{\geq}$ on classes of distributions with marginals satisfying (4.1.1) and hence the theorem is not directly applicable to these statistics. But if the underlying distributions have continuous marginal distribution functions, the class of all empirical distribution functions drawn from such underlying distributions satisfies (4.1.1) and therefore the following corollary holds.

COROLLARY 4.3.2. *Let* F *and* F' *be bivariate distributions with continuous marginal distribution functions and let* ϕ *be a measure which preserves* $\overset{a}{\geq}$ *on classes of distribution functions satisfying* (4.1.1), *then* (4.3.1) *holds.*

It follows from the examples in the previous section that when F and F' have continuous marginals, then (4.3.1) holds for Spearman's rank correlation rho, Fisher-Yates' normal score statistic, vander Waerden's statistic, the quadrant statistic, Kendall's rank correlation tau and the signed canonical correlation. This result implies that tests based on these statistics have a higher power against F' than against F. In particular, it follows from Proposition 4.1.2. that these tests are unbiased against all RR_1-dependent and CR_1-dependent alternatives with continuous marginals. This unbiasedness result is, for some of these statistics, also proved in LEHMANN (1966).

Note that (4.3.1) also holds for these statistics when the functions κ_1 and κ_2 of Definition 4.1.4 are such that for any sample $(X_1,Y_1),\ldots,(X_N,Y_N)$ from F the empirical distribution function of this sample and the empirical distribution function of $(\kappa_1(X_1,Y_1),\kappa_2(X_1,Y_1)),\ldots,(\kappa_1(X_N,Y_N),\kappa_2(X_N,Y_N))$ satisfy (4.1.1).

Another aspect of the order preserving property of the ordering $\overset{a}{\geq}$ is the following. Suppose F' and F are associated and $F' \overset{a}{\geq} F$. Then samples from F' are more frequently associated than samples from F.

PROPOSITION 4.3.3. *Let* F *be associated and* $F' \overset{a}{\geq} F$, *then for all sample sizes* N

$$P\{\hat{F}'^{(N)} \text{ is associated}\} \geq P\{\hat{F}^{(N)} \text{ is associated}\}.$$

Similar results hold for RR_1*-dependence and* $\overset{rr}{\geq}$ *and for* CR_1*-dependence and* $\overset{cr}{\geq}$.

PROOF. By (4.1.6), $F' \overset{a}{\geq} F$ and F is associated imply F' is associated. Using the same set-up as in the proof of Theorem 4.3.1, it follows that $\hat{F}$ is associated implies $\hat{H}'$ is associated. Hence
$P\{\hat{F} \text{ is associated}\} \leq P\{\hat{H}' \text{ is associated}\}$. □

4.4. THE DISCRETE CASE

If the distributions F and F' contain atoms, the main result given in the previous section is generally not applicable to the statistics considered in Section 4.2. Moreover, these rank statistics are not well-defined for samples containing ties. Furthermore, Definition 4.1.4 requires that the whole mass of each atom of F is moved to an atom of F'. In particular, when F and F' correspond to probability tables P and P' of the same size, this means that cell's of P are pooled in order to obtain P' and empty cell's in P' arise. Thus the ordering $\overset{a}{\geq}$ is only present between probability tables in very few and simple cases and although Theorem 4.3.1 is valid, statistics to which this theorem can be applied are still unknown.

This section discusses a modified definition of the ordering "more associated" which is more appropriate for probability tables. It is shown that the results of the previous sections continue to hold for the modified ordering in the probability table case, provided the measures of dependence are appropriately defined. The idea used here can also be applied to discrete distributions in general, but the notation is then more complicated.

This section is only concerned with the probability table case and it is assumed that (X,Y) and (X',Y') are distributed on $\mathcal{X} \times \mathcal{Y}$ and $\mathcal{X}' \times \mathcal{Y}'$ respectively, where $\mathcal{X} = \{1,\ldots,n\}$, $\mathcal{Y} = \{1,\ldots,m\}$, $\mathcal{X}' = \{1,\ldots,n'\}$ and $\mathcal{Y}' = \{1,\ldots,m'\}$. Their distributions F and F' correspond to probability

tables P and P' with entries $p_{ij} = P\{X = i, Y = j\}$ and $p'_{ij} = P\{X' = i, Y' = j\}$ for appropriate indices i and j.

DEFINITION 4.4.1. In the probability table case (X',Y') is said to be *more associated* than (X,Y), denoted $(X',Y') \overset{\sim}{\geq}^{a} (X,Y)$, when

$$(X' + U', Y' + V') \overset{a}{\geq} (X + U, Y + V)$$

where U', U, V' and V are independently and uniformly distributed on [0,1) and also independent of (X',Y') and (X,Y). Equivalently to $(X',Y') \overset{\sim}{\geq}^{a} (X,Y)$ the notation $F' \overset{\sim}{\geq}^{a} F$ or $P' \overset{a}{\geq} P$ is used. Orderings $\overset{\sim}{\geq}^{rr}$, $\overset{\sim}{\geq}^{cr}$, $\overset{\sim}{\geq}^{q}$ and the equivalence relation $\overset{\sim}{=}^{d}$ are similarly defined.

The assumption of uniformity of U, U', V and V' is no restriction. Their independence is not essential but has convenient consequences (e.g. Proposition 4.4.1).

Clearly, (X' + U', Y' + V') and (X + U, Y + V) are distributed on $[1,n'+1) \times [1,m'+1)$ and $[1,n+1) \times [1,m+1)$ respectively and have continuous marginal distribution functions. These *continuous versions* of (X',Y') and (X,Y) are denoted by $\tilde{F}'$ and $\tilde{F}$. Thus $(X',Y') \overset{\sim}{\geq}^{a} (X,Y)$ iff $\tilde{F}' \overset{a}{\geq} \tilde{F}$.

It follows from the results in Section 4.1 that the orderings $\overset{\sim}{\geq}^{a}$ and $\overset{\sim}{\geq}^{q}$ are partial orderings of the equivalence classes defined by $\overset{\sim}{=}^{d}$ over all probability tables.

Furthermore, it is easily verified that the pair (X,Y) is independent, quadrant dependent, associated, RR_1-dependent or CR_1-dependent iff its continuous version (X + U, Y + V) has the corresponding form of dependence. Therefore, the following proposition follows directly from Proposition 4.1.2.

PROPOSITION 4.4.1.

$(X',Y') \overset{\sim}{\geq}^{q} (X,Y)$, (X,Y) *is quadrant dep.* ⇒ (X',Y') *is quadrant dep.*,

$(X',Y') \overset{\sim}{\geq}^{a} (X,Y)$, (X,Y) *is associated* ⇒ (X',Y') *is associated*,

$(X',Y') \overset{\sim}{\geq}^{rr} (X,Y)$, (X,Y) *is* RR_1*-dep.* ⇒ (X',Y') *is* RR_1*-dep.*,

$(X',Y') \overset{\sim}{\geq}^{cr} (X,Y)$, (X,Y) *is* CR_1*-dep.* ⇒ (X',Y') *is* CR_1*-dep.*.

Moreover, when X *and* Y *are independent, then*

$(X',Y') \overset{\sim}{\geq}^{q} (X,Y) \Leftrightarrow$ (X',Y') *is quadrant dep.*,

$(X',Y') \overset{\sim}{\geq}^{rr} (X,Y) \Leftrightarrow$ (X',Y') *is* RR_1*-dep.*,

$(X',Y') \overset{\sim}{\geq}^{cr} (X,Y) \Leftrightarrow$ (X',Y') *is* CR_1*-dep.*.

PROPOSITION 4.4.2. $(X',Y') \overset{\sim}{\geq}^{a} (X,Y) \Rightarrow (X',Y') \overset{\sim}{\geq}^{q} (X,Y)$.

Next consider a measure of dependence ϕ which is (properly) defined for bivariate distributions with continuous marginal distribution functions. Define the *probability table version* $\tilde{\phi}$ of ϕ for a probability table P by

$$\tilde{\phi}(P) = \phi(\tilde{F}) \tag{4.4.1}$$

where $\tilde{F}$ is the continuous version of the distribution function F corresponding to P. Thus even for an empirical probability table the probability table version $\tilde{\phi}$ is defined as ϕ computed at its continuous version. The probability table version of rank statistics show similarity to the randomization procedure of removing tied observations: the randomization procedure is not performed but the randomization distribution is considered instead. In the case that the score function J in the measure (4.2.2) is bilinear and in the case that the function K in the measure (4.2.6) is linear, the probability table versions $\tilde{\phi}$ of these measures ϕ are equivalent to the statistic ϕ using mid-ranks for ties (see Examples 4.4.1 and 4.4.2). In general $\tilde{\phi}$ need not be equivalent to ϕ with well-known treatment of ties (such as: expectation of the statistic over randomization of ties, expectation of score function over randomization, mid-ranks, least favourable value of the statistic, most favourable value, etc; cf. HÁJEK and ŠIDÁK, 1967, pp. 118-124). Nor need $\tilde{\phi}$ correspond to other well-known adaptations of rank statistics to contingency (probability) tables. But the version (4.4.1) is a natural choice for adapting statistics for which results, similar to those given in the previous sections, can easily be derived.

In the following two examples the probability table versions of Kendall's tau and Spearman's rho are calculated. First note that for an $n \times m$ (empirical) probability table P the distribution function of its continuous version equals

$$\tilde{F}(x,y) = \sum_{k=1}^{i-1} \sum_{\ell=1}^{j-1} p_{k\ell} + (x-i) \sum_{\ell=1}^{j-1} p_{i\ell} + (y-j) \sum_{k=1}^{i-1} p_{kj} + (x-i)(y-j)p_{ij} \tag{4.4.2}$$

for $i \leq x \leq i+1$ and $j \leq y \leq j+1$ and $i = 1,\ldots,n$; $j = 1,\ldots,m$. A summation $\sum_{k=1}^{0}$ is assumed to be zero.

EXAMPLE 4.4.1. *Kendall's rank correlation tau.*

By (4.4.1), (4.4.2), (4.2.6) and (4.2.7) the probability table version of Kendall's tau for an $n \times m$ (empirical) probability table P equals

$$\tilde{\phi}_K(P) = \iint 4\tilde{F}d\tilde{F} - 1 = 4 \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \int_i^{i+1} \int_j^{j+1} \tilde{F}(x,y)dydx - 1$$

$$= 4 \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \left(\sum_{k=1}^{i-1} \sum_{\ell=1}^{j-1} p_{k\ell} + \tfrac{1}{2} \sum_{k=1}^{i-1} p_{kj} + \tfrac{1}{2} \sum_{\ell=1}^{j-1} p_{i\ell} + \tfrac{1}{4} p_{ij} \right) - 1 .$$

Elementary calculations yield

$$\tilde{\phi}_K(P) = 2 \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \left(\sum_{k=1}^{i-1} \sum_{\ell=1}^{j-1} p_{k\ell} - \sum_{k=1}^{i-1} \sum_{\ell=j+1}^{m} p_{k\ell} \right) \tag{4.4.3}$$

$$= 2\{\text{sum of all } 2 \times 2 \text{ determinants}\}.$$

To show this we employ a more transparent notation. Let (X_1,Y_1) and (X_2,Y_2) be independent, both with distribution specified by P and let U_1, V_1, U_2 and V_2 be independent and uniformly distributed on $[0,1)$. Then

$$\tilde{\phi}_K(P) = 4P\{X_1 + U_1 \le X_2 + U_2, Y_1 + V_1 \le Y_2 + V_2\} - 1$$

$$= 4(P\{X_1 < X_2, Y_1 < Y_2\} + \tfrac{1}{2}P\{X_1 < X_2, Y_1 = Y_2\} \tag{4.4.4}$$

$$+\tfrac{1}{2}P\{X_1 = X_2, Y_1 < Y_2\} + \tfrac{1}{4}P\{X_1 = X_2, Y_1 = Y_2\}) - 1.$$

Furthermore

$$1 = 2(P\{X_1 < X_2, Y_1 < Y_2\} + P\{X_1 < X_2, Y_1 > Y_2\} + P\{X_1 = X_2, Y_1 < Y_2\} \tag{4.4.5}$$

$$+ P\{X_1 < X_2, Y_1 = Y_2\} + \tfrac{1}{2}P\{X_1 = X_2, Y_1 = Y_2\}).$$

Inserting (4.4.5) in (4.4.4) yields

$$\tilde{\phi}_K(P) = 2(P\{X_1 < X_2, Y_1 < Y_2\} - P\{X_1 < X_2, Y_1 > Y_2\})$$

$$= 2(P\{(X_1,Y_1) \text{ is concordant with } (X_2,Y_2)\}$$

$$-P\{(X_1,Y_1) \text{ is disconc. with } (X_2,Y_2)\}),$$

establishing (4.4.3). In case of an empirical probability table $\hat{P}^{(N)}$ based on N observations, (4.4.3) becomes

$$\tilde{\phi}_K(\hat{P}^{(N)}) = (\#\{\text{concordant pairs}\} - \#\{\text{disconc. pairs}\})/(\tfrac{1}{2}N^2). \tag{4.4.6}$$

This is equivalent to two other adaptations of Kendall's tau to contingency tables. First, KENDALL (1975, p. 34) suggests using the well-known definition

$$(\#\{\text{conc. pairs}\} - \#\{\text{disconc. pairs}\})/\#\{\text{pairs}\}$$

without removing ties. Clearly, this yields $N\tilde{\phi}_K(\hat{P}^{(N)})/(N-1)$. Secondly, it is equivalent to using mid-ranks for tied observations. Denoting the number of observations in the (i,j)-th cell by $N_{ij} = N\hat{p}_{ij}^{(N)}$, the mid-rank of the observations in this cell is

$$Q_{ij} = \sum_{k=1}^{i-1}\sum_{\ell=1}^{j-1} N_{k\ell} + \tfrac{1}{2}\sum_{k=1}^{i-1} N_{kj} + \tfrac{1}{2}\sum_{\ell=1}^{j-1} N_{i\ell} + \tfrac{1}{4}(N_{ij} - 1) + 1$$

(i.e. the expected value of the randomized ranks of the observations in the cell, where the expectation is taken over the randomization; see HÁJEK and ŠIDÁK, 1967, p. 121). Therefore, Kendall's tau (4.2.8) with mid-ranks equals

$$4N\left(\sum_{i=1}^{n}\sum_{j=1}^{m} N^{-2}N_{ij}Q_{ij} - N - 3\right)/(N-1) = N\tilde{\phi}_K(\hat{P}^{(N)})/(N-1).$$

Other adaptations of Kendall's tau, such as the measure proposed by GOODMAN and KRUSKAL (1954)

$$(\#\{\text{conc. pairs}\} - \#\{\text{disconc. pairs}\})/(\#\{\text{conc. pairs}\} + \#\{\text{disconc. pairs}\})$$

or the second adaptation proposed by KENDALL (1975, p. 34)

$$(\#\{\text{conc. pairs}\} - \#\{\text{disconc. pairs}\})/$$
$$(\#\{\text{pairs with no tie in row var.}\}\#\{\text{pairs with no tie in col. var.}\})^{\frac{1}{2}}$$

are not equivalent but quite similar to (4.4.6). It is not clear whether Proposition 4.4.3 and Theorem 4.4.4 hold for these two measures.

EXAMPLE 4.4.2. *Spearman's rank correlation rho.*
The probability table version of Spearman's rho for an $n \times m$ probability table P equals

$$(4.4.7)\qquad \tilde{\phi}_S(P) = 3\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}\left(2\sum_{k=1}^{i-1} r_k + r_i - 1\right)\left(2\sum_{\ell=1}^{j-1} c_\ell + c_j - 1\right)$$

where $r_i = \Sigma_{j=1}^{m} p_{ij}$ for $i = 1,\ldots,n$ and $c_j = \Sigma_{i=1}^{n} p_{ij}$ for $j = 1,\ldots,m$. For an empirical probability table $\hat{P}^{(N)}$, (4.4.7) is equivalent to Spearman's rho (formula (11) on p. 113 of HÁJEK and ŠIDÁK, 1967) with mid-ranks $R_i = \Sigma_{k=1}^{i-1} N\hat{r}_k^{(N)} + \frac{1}{2}(N\hat{r}_i^{(N)} + 1)$ for the observations in the i-th row and with mid-ranks $Q_j = \Sigma_{\ell=1}^{j-1} N\hat{c}_\ell^{(N)} + \frac{1}{2}(N\hat{c}_j^{(N)} + 1)$ for the observations in the j-th column. The statistic obtained in this way equals $N^2\tilde{\phi}_S(\hat{P}^{(N)})/(N^2 - 1)$.

The following results easily follow from their counterparts in Sections 4.2 and 4.3.

PROPOSITION 4.4.3. *Let ϕ be a measure of dependence which preserves $\overset{a}{\geq}$ ($\overset{rr}{\geq}$, $\overset{cr}{\geq}$ or $\overset{q}{\geq}$) for distributions with continuous marginal distribution functions. Then its probability table version $\tilde{\phi}$ preserves $\overset{\tilde{a}}{\geq}$ ($\overset{\tilde{rr}}{\geq}$, $\overset{\tilde{cr}}{\geq}$ or $\overset{\tilde{q}}{\geq}$).*

THEOREM 4.4.4. *Let $\hat{P}^{(N)}$ and $\hat{P}'^{(N)}$ be empirical probability tables based on multinomial samples of size N drawn from P and P' respectively. Then for any measure $\tilde{\phi}$ which preserves $\overset{\tilde{a}}{\geq}$,*

$$P' \overset{\tilde{a}}{\geq} P \Rightarrow P\{\tilde{\phi}(\hat{P}'^{(N)}) \geq t\} \geq P\{\tilde{\phi}(\hat{P}^{(N)}) \geq t\}$$

for all t and N.

Thus when $P' \overset{\tilde{a}}{\geq} P$, the probability table versions of Spearman's rho, Kendall's tau, Fisher-Yates' normal score statistic, the quadrant statistic and the signed canonical correlation based on multinomial samples of size N drawn from P' are stochastically larger than those based on multinomial samples from P. In particular, tests based on these statistics are unbiased against RR_1-dependent and CR_1-dependent alternatives.

In case $P' \overset{\tilde{rr}}{\geq} P$, the result of Theorem 4.4.4 also holds when $\hat{P}'^{(N)}$ and $\hat{P}^{(N)}$ are based on samples of size N from P' and P which are conditioned on row sums. A similar property holds of course for $\overset{\tilde{cr}}{\geq}$ and conditioning on column sums. It is not clear whether such results extend to other conditional Poisson samples; this problem cannot be solved by the method of proof used in Theorem 4.3.1.

Proposition 4.3.3 immediately leads to the following result.

PROPOSITION 4.4.5. *Let P and P' be associated and $P' \overset{\tilde{a}}{\geq} P$, then for multinomial samples of size N,*

$$P\{\hat{P}'^{(N)} \text{ is associated}\} \geq P\{\hat{P}^{(N)} \text{ is associated}\}.$$

Similar results hold for RR_1-dependence and $\overset{\tilde{rr}}{\geq}$ and CR_1-dependence and $\overset{\tilde{cr}}{\geq}$.

When P and P' are DR_1-dependent and $P' \overset{\tilde{rr}}{\geq} P$ and $P' \overset{\tilde{cr}}{\geq} P$ then $P\{\hat{P}'^{(N)}$ is DR_1-dep.$\} \geq P\{\hat{P}^{(N)}$ is DR_1-dep.$\}$. This indicates that the ordering property of CA more frequently arises in multinomial samples from P' than in multinomial samples from P.

It should be emphasized that Theorems 4.3.1 and 4.4.4 generally do not hold for statistics which preserve $\overset{a}{\geq}$, or $\overset{\tilde{a}}{\geq}$ only on special subclasses of bivariate distributions. For example, the canonical correlation preserves $\overset{\tilde{a}}{\geq}$

on the class of all DR_1-dependent probability tables (cf. Section 4.2). But since samples from such probability tables need not be DR_1-dependent, Theorem 4.4.4 does not apply for the canonical correlation. The following example shows that the reversed stochastic inequality may even hold.

EXAMPLE 4.4.3. Let the distributions of (X',Y') and (X,Y) be specified by the 2×2 probability tables

$$P' = \begin{pmatrix} \frac{1}{4} & 0 \\ \frac{1}{2} & \frac{1}{4} \end{pmatrix} \quad \text{and} \quad P = \begin{pmatrix} 3/16 & 1/16 \\ 9/16 & 3/16 \end{pmatrix}$$

respectively. Note that X and Y are independent. Furthermore, $P' \overset{\widetilde{rr}}{\geq} P$. Let $\widetilde{\phi}_{CC}$ denote the canonical correlation functional. Then for multinomial samples of size $N = 3$ the following exact distribution functions can easily be computed

$$P\{\widetilde{\phi}_{CC}(\hat{P}'^{(3)}) \leq 0\} = .71875, \qquad P\{\widetilde{\phi}_{CC}(\hat{P}^{(3)}) \leq 0\} = .68359375,$$
$$P\{\widetilde{\phi}_{CC}(\hat{P}'^{(3)}) \leq \tfrac{1}{2}\} = .90625, \qquad P\{\widetilde{\phi}_{CC}(\hat{P}^{(3)}) \leq \tfrac{1}{2}\} = .89453125,$$
$$P\{\widetilde{\phi}_{CC}(\hat{P}'^{(3)}) \leq 1\} = 1.0, \qquad P\{\widetilde{\phi}_{CC}(\hat{P}^{(3)}) \leq 1\} = 1.0.$$

Thus $\widetilde{\phi}_{CC}(\hat{P}^{(3)})$ is stochastically larger than $\widetilde{\phi}_{CC}(\hat{P}'^{(3)})$. Since for 2×2 probability tables the canonical correlation test is equivalent to the chi-square test, this example shows that the canonical correlation test and the chi-square test are biased. In fact one can prove that for all N, $P\{\widetilde{\phi}_{CC}(\hat{P}'^{(N)}) = 1\} < P\{\widetilde{\phi}_{CC}(\hat{P}^{(N)}) = 1\}$ and hence for all N there exists an α such that the power of the level α test against P' is strictly less than α.

CHAPTER 5

THE MULTIVARIATE CASE

5.1. MULTIPLE CORRESPONDENCE ANALYSIS

CA can be generalized to the case when more variables are involved. This generalization is called multiple correspondence analysis (abbreviated MCA); other names in the literature are homogeneity analysis (GIFI, 1981) and first order correspondence analysis (HILL, 1974). This section summarizes some basic properties of MCA.

In order to attack the multivariate case, notation must be altered at some places. Let $X_1, X_2, \ldots, X_k$ be nominal variables, where X_ℓ has n_ℓ integer labeled categories ($\ell = 1, \ldots, k$). The bivariate marginal distribution of X_j and X_ℓ is given in the $n_j \times n_\ell$ probability table $P_{j\ell}$, for $j, \ell = 1, \ldots, k$. It is clear that $P_{j\ell}^\top = P_{\ell j}$ for $j, \ell = 1, \ldots, k$. The diagonal matrix containing the univariate marginal distribution of X_ℓ is denoted by $R_\ell = P_{\ell\ell}$, for $\ell = 1, \ldots, k$, assumed to be non-singular.

MCA seeks k functions $\varphi_{11}, \varphi_{21}, \ldots, \varphi_{k1}$, defined on the categories of $X_1, X_2, \ldots, X_k$ respectively, such that the first principal component of the correlation matrix of $\varphi_{11}(X_1), \varphi_{21}(X_2), \ldots, \varphi_{k1}(X_k)$ has maximal variance. This principal component is called the first MCA component. Subsequently, MCA seeks a second component which has maximal variance but which is uncorrelated with the first, etc.

DEFINITION 5.1.1. The t-*th MCA component* is the linear combination of transformed variables

$$Y_t = \sum_{\ell=1}^{k} \alpha_{\ell t} \varphi_{\ell t}(X_\ell)$$

for which $\mu_t = Var\,(Y_t)$ is maximal subject to

$$
\begin{aligned}
& E\varphi_{\ell t}(X_\ell) = 0, \qquad Var\,(\varphi_{\ell t}(X_\ell)) = 1 \text{ for } \ell = 1,\ldots,k \\
& \sum_{\ell=1}^{k} \alpha_{\ell t}^2 = 1, \qquad (5.1.1) \\
& Corr\,(Y_t, Y_s) = 0 \quad \text{for } s = 1,\ldots,t-1.
\end{aligned}
$$

The *MCA solution* consists of all (k+1)-tuples $(\mu_t, \alpha_{1t}\varphi_{1t}(X_1),\ldots,\alpha_{kt}\varphi_{kt}(X_k))$ for $t = 1,2,\ldots$.

The present definition of MCA may depart from other definitions given in the literature with respect to the normalization of the variable weights $\alpha_{1t},\ldots,\alpha_{kt}$. It is clear from

$$Var\,(Y_t) = \sum_{j=1}^{k} \sum_{\ell=1}^{k} \alpha_{jt}\alpha_{\ell t}\, Corr\,(\varphi_{jt}(X_j), \varphi_{\ell t}(X_\ell))$$

that MCA considers only the bivariate marginals of the k-dimensional distribution, which is a drawback of the technique.

MCA may be interpreted as a generalization of principal component analysis for categorical variables. Moreover, in the case that $X_1, X_2, \ldots, X_k$ are all dichotomous (i.e. $n_\ell = 2$ for $\ell = 1,2,\ldots,k$), the functions $\varphi_{\ell t}$ are determined, up to a change of sign, by their normalization. Hence maximalization of the variance of Y_t is only with respect to the variable weights $\alpha_{\ell t}$ for $\ell = 1,\ldots,k$. Therefore, MCA is in the dichotomous case equivalent to principal component analysis of the correlation matrix of the nominal variables $X_1, X_2, \ldots, X_k$. Note that for dichotomous variables the correlation matrix is defined up to changes of sign of rows and columns. This property of MCA was also noted by DE LEEUW (1973), GUTTMAN (1950) and NAKHLÉ (1976).

Another special case arises when there are only two variables involved. The following proposition states that in case k = 2, MCA is equivalent to CA (see also GIFI, 1981, p. 139; GREENACRE, 1984, p. 135 and LEBART et al., 1977, p. 139).

PROPOSITION 5.1.1. *Consider the case* k = 2 *and assume, without loss of generality, that* $n_1 \leq n_2$.

(i) *Suppose* $(\lambda_t, \varphi_{1t}(X_1), \varphi_{2t}(X_2))$ *for* $t = 1,\ldots,n_1-1$ *is a* CA *solution. Let* $\varphi_{2t}(X_2)$ *for* $t = n_1,\ldots,n_2-1$ *be normalized variables such that* $\varphi_{2s}(X_2)$ *is uncorrelated with* $\varphi_{2t}(X_2)$ *for* $s < t$ *and* $t = 1,\ldots,n_2-1$. *Then a MCA solution is given by*

$$
\begin{aligned}
& (1+\lambda_t,\ 2^{-\frac{1}{2}}\varphi_{1t}(X_1),\ 2^{-\frac{1}{2}}\varphi_{2t}(X_2)) \qquad \text{for } t = 1,\ldots,n_1-1, \\
& (1,\ 0,\ \varphi_{2t}(X_2)) \qquad \text{for } t = n_1,\ldots,n_2-1,
\end{aligned}
$$

$(1-\lambda_s,\ 2^{-\frac{1}{2}}\varphi_{1s}(X_1),\ -2^{-\frac{1}{2}}\varphi_{2s}(X_2))$ *where* $s = n_1+n_2-t-1$ *for* $t = n_2,\ldots,n_1+n_2-2$.

(ii) *Conversely, for* $t = 1,\ldots,n_1-1$ *the variables* $\varphi_{1t}(X_1)$ *and* $\varphi_{2t}(X_2)$ *of the MCA solution are a pair of* t*-th canonical row and column variables with corresponding correlation* $\lambda_t = \mu_t - 1$, *provided* $\lambda_t > 0$.

PROOF. Follows directly from (2.1.4), (2.1.5) and Theorem 5.1.3 below. □

For general k, the MCA solution satisfies the following generalization of Proposition 2.1.1 (simultaneous linear regression).

PROPOSITION 5.1.2. *The tuple* $(\mu_t,\alpha_{1t}\varphi_{1t}(X_1),\ldots,\alpha_{kt}\varphi_{kt}(X_k))$ *is a* t*-th tuple of the MCA solution iff*

$$(5.1.2)\qquad (\mu_t - 1)\alpha_{\ell t}\varphi_{\ell t}(X_\ell) = \sum_{j\neq\ell} \alpha_{jt}E(\varphi_{jt}(X_j)|X_\ell) \quad a.s.\ for\ \ell = 1,\ldots,k$$

where μ_t *is maximal subject to* (5.1.1).

PROOF. The proof is quite similar to the proof of Proposition 2.1.1. First suppose that the tuple satisfies (5.1.2) with μ_t maximal subject to (5.1.1). Multiplying (5.1.2) on left and right hand side by $\alpha_{\ell t}\varphi_{\ell t}(X_\ell)$ and taking expectation over X_ℓ yields

$$(\mu_t - 1)\alpha_{\ell t}^2 Var\,(\varphi_{\ell t}(X_\ell)) =$$

$$= \sum_{j\neq\ell} \alpha_{\ell t}\alpha_{jt} Cov\,(\varphi_{\ell t}(X_\ell),\varphi_{jt}(X_j)) \text{ for } \ell = 1,\ldots,k.$$

Adding these k equotions produces $\mu_t = Var\,(Y_t)$. Hence the tuple maximizes $Var(Y_t)$ subject to (5.1.1).

Conversely, suppose that the tuple maximizes $\mu_t = Var\,(Y_t)$ subject to (5.1.1). Let θ, $\beta_{\ell t}$ and $\psi_{\ell t}(X_\ell)$ for $\ell = 1,\ldots,k$ be such that $\Sigma_{\ell=1}^k \beta_{\ell t}^2 = 1$, $Var(\psi_{\ell t}(X_\ell)) = 1$ for $\ell = 1,\ldots,k$ and

$$(5.1.3)\qquad (\theta-1)\beta_{\ell t}\psi_{\ell t}(X_\ell) = \sum_{j\neq\ell} \alpha_{jt}E(\varphi_{jt}(X_j)|X_\ell) \text{ a.s. for } \ell = 1,\ldots,k.$$

Define $Y_t' = \Sigma_{\ell=1}^k \beta_{\ell t}\psi_{\ell t}(X_\ell)$. Notice that

$$(\mu_t\, Var\,(Y_t'))^{\frac{1}{2}} \geq Cov\,(Y_t,Y_t') = \theta \sum_{\ell=1}^k Var\,(\beta_{\ell t}\psi_{\ell t}(X_\ell)) = \theta.$$

Furthermore, by Cauchy-Schwarz,

$$\mu_t = Var\,(Y_t) = \theta \sum_{\ell=1}^{k} Cov\,(\alpha_{\ell t}\varphi_{\ell t}(X_\ell), \beta_{\ell t}\psi_{\ell t}(X_\ell))$$

$$\leq \theta \left(\sum_{\ell=1}^{k} Var\,(\alpha_{\ell t}\varphi_{\ell t}(X_\ell)) \right)^{\frac{1}{2}} \left(\sum_{\ell=1}^{k} Var\,(\beta_{\ell t}\psi_{\ell t}(X_\ell)) \right)^{\frac{1}{2}} = \theta$$

with equality only if

(5.1.4) $$\alpha_{\ell t}\varphi_{\ell t}(X_\ell) = \beta_{\ell t}\psi_{\ell t}(X_\ell) \quad \text{a.s. for } \ell = 1,\ldots,k.$$

Thus $Var\,(Y_t') \geq Var\,(Y_t)$ with equality only when (5.1.4) holds. Moreover, one can show by induction over t that Y_t' is uncorrelated with Y_s for $s = 1,\ldots,t-1$ whenever Y_t is. To this end suppose that (5.1.2) holds, with t replaced by s, for $s = 1,\ldots,t-1$. Then for $s < t$,

$$\mu_s\, Cov\,(Y_t, Y_s) = \theta\mu_s \sum_{\ell=1}^{k} Cov\,(\alpha_{\ell s}\varphi_{\ell s}(X_\ell), \beta_{\ell t}\psi_{\ell t}(X_\ell)) = \theta\, Cov\,(Y_t', Y_s).$$

This proves the proposition. □

This proposition already indicates that the MCA solution can be obtained by solving a matrix eigenvalue problem. In order to formulate the eigenvalue problem introduce the following notation. Let $\nu_{\ell t}$ denote the n_ℓ-dimensional vector of category scores

$$\nu_{\ell t} = (\alpha_{\ell t}\varphi_{\ell t}(1),\ldots,\alpha_{\ell t}\varphi_{\ell t}(n_\ell))^T$$

of the ℓ-th variable on the t-th MCA component for $\ell = 1,\ldots,k$ and $t = 1,2,\ldots$. All category scores on the t-th MCA component are summarized in the vector of length $n = \Sigma_{\ell=1}^{k} n_\ell$,

$$\nu_t = (\alpha_{1t}\varphi_{1t}(1),\ldots,\alpha_{1t}\varphi_{1t}(n_1),\ldots \quad \ldots,\alpha_{kt}\varphi_{kt}(1),\ldots,\alpha_{kt}\varphi_{kt}(n_k))^T.$$

Let R be the $n \times n$ diagonal matrix with the diagonal elements of $R_1,\ldots,R_k$ on the diagonal and let B be the $n \times n$ block matrix containing the bivariate marginals

$$B = \begin{pmatrix} R_1 & P_{12} & \cdots & P_{1k} \\ P_{21} & R_2 & \cdots & P_{2k} \\ \vdots & \vdots & & \vdots \\ P_{k1} & P_{k2} & \cdots & R_k \end{pmatrix}.$$

Recall that e denotes the vector having all its components equal to unity. Define the n-dimensional vectors $u_1,\ldots,u_{k-1}$ as follows. Let the first n_1

components of u_1 be equal to unity, the next n_2 components equal to -1 and all other components zero; let u_2 be the vector having the first n_1 components equal to unity, the next n_2 components zero, then n_3 components equal to -1 and all further components zero; etc., until u_{k-1} which has its first n_1 components equal to unity, the last n_k components equal to -1 and all other components zero.

THEOREM 5.1.3. *The vectors* ν_t *containing all category scores on the* t*-th MCA component exist for* $t = 1,\ldots,n-k$. *Moreover, the vectors* $e,\nu_1,\nu_2,\ldots,\nu_{n-k},u_1,\ldots,u_{k-1}$ *can be any complete set of suitably normalized eigenvectors of the matrix* $R^{-1}B$ *corresponding to the eigenvalues* $k \geq \mu_1 \geq \mu_2 \geq \ldots \geq \mu_{n-k} \geq 0 = \ldots = 0$ (*where* μ_t *is the variance of the* t*-th MCA component*).

PROOF. First note that a complete set of eigenvectors of $R^{-1}B$ exists and that $e,u_1,\ldots,u_{k-1}$ are eigenvectors corresponding to eigenvalues $k,0,\ldots,0$. By Proposition 5.1.2 the MCA solution satisfies

$$\mu_t \nu_t = R^{-1}B\nu_t \tag{5.1.5}$$

where μ_t is maximal subject to

$$e^T R_\ell \nu_{\ell t} = 0, \quad \text{for } \ell = 1,\ldots,k, \quad \nu_t^T R \nu_t = 1,$$
$$\nu_s^T B \nu_t = 0, \quad \text{for } s = 1,\ldots,t-1.$$

Since (5.1.5) holds,

$$\mu_t e^T R_\ell \nu_{\ell t} = e^T R_\ell (R_\ell^{-1}P_{\ell 1},\ldots,R_\ell^{-1}P_{\ell k})\nu_t = e^T R \nu_t,$$
$$\mu_t \nu_s^T R \nu_t = \nu_s^T B \nu_t.$$

Therefore the MCA solution satisfies (5.1.5) with μ_t maximal subject to

$$e^T R \nu_t = 0, \quad \nu_t^T R \nu_t = 1,$$
$$\nu_s^T R \nu_t = 0 \quad \text{for } s = 1,\ldots,t-1.$$

This in turn is equivalent to the eigenvalue problem of the matrix $R^{-1}B$ and hence ν_t for $t = 1,\ldots,n-k$ can be taken as the remaining eigenvectors of $R^{-1}B$. □

GIFI (1981), GREENACRE (1984), HILL (1974) and LEBART et al. (1977) describe

MCA from other points of view, which also lead to the present eigenvalue problem. GIFI (1981, pp. 103-104) and HILL (1974) show that the first MCA component, according to their definitions, determines a transformation of the variables which maximizes the largest eigenvalue of their correlation matrix.

Similar to CA, the MCA category scores are usually displayed in a q-dimensional graphical representation. The i-th category of the ℓ-th variable is represented by the point $\xi_{i\ell} = (\mu_1\alpha_{\ell 1}\varphi_{\ell 1}(i),\ldots,\mu_q\alpha_{\ell q}\varphi_{\ell q}(i))$ for $i = 1,\ldots,n_\ell$ and $\ell = 1,\ldots,k$. The interpretation of this graphical display differs form the interpretation of its CA analogon and is less transparent (cf. GIFI, 1981, p. 102 and KESTER and SCHRIEVER, 1982). Moreover, in empirical situations it is also quite common to display the scores of the individuals on the first q MCA components. For further properties, results and applications of MCA see DE LEEUW (1984), GIFI (1981), GREENACRE (1984), LEBART et al. (1977) and NISHISATO (1980).

5.2. ORDERING PROPERTIES IN MULTIPLE CORRESPONDENCE ANALYSIS

In Section 3.3 it was demonstrated that (S)DR_s-dependence implies that the t-th canonical row and column functions are (strictly) monotone of order t, for $t = 1,\ldots,s$. In the present section it is shown that a similar property holds for MCA with respect to a (strict) multiple double regression dependence of order 1. However, it need not hold with respect to multiple regression dependence of higher order.

DEFINITION 5.2.1. The variables $X_1,X_2,\ldots,X_k$ are called (*strictly*) *multiple double regression dependent of order* s (abbreviated (S)MDR_s-dependent) if all pairs (X_j,X_ℓ) for $j,\ell = 1,\ldots,k$; $j \neq \ell$ are (S)DR_s-dependent. Furthermore, k categorical variables are called (*strictly*) *multiple order dependent of order* s when there exist integer labels for their categories such that the labeled integer valued variables are (S)MDR_s-dependent.

MDR_1-dependence preserves monotonicity of order 1 in the following sense. Consider the linear regression of X_ℓ on the transformed variables $f_j(X_j)$ for $j \neq \ell$,

$$(5.2.1) \qquad g(x) = \sum_{j\neq\ell} E(f_j(X_j)|X_\ell = x).$$

It is easily seen that g is increasing (monotone of order 1) whenever all functions f_j are increasing (monotone of order 1 in the same direction) and

all pairs (X_j,X_ℓ) for $j = 1,\ldots,k$ are DR_1-dependent. However, MDR_2-dependence need not preserve monotonicity of order 2 in a similar sense. In case $X_1,X_2,\ldots,X_k$ are MDR_2-dependent and all functions f_j are monotone of order 2 and oscillate in the same direction, then in spite of the fact that by Theorem 3.2.3 all functions $E(f_j(X_j)|X_\ell = x)$ are monotone of order ≤ 2, their sum need not be monotone of order ≤ 2. Moreover, examples are easily constructed showing that such a property does not always hold for the multivariate normal distribution. Thus even if stronger forms of ordinal dependence are present in the k-dimensional distribution, for example if the joint density of the k variables is TP_2 in every pair when the remaining variables are kept fixed, then still monotonicity of order 2 need not be preserved in the regression (5.2.1). (Note that the density of a multivariate normal distribution is TP_∞ in pairs iff the off-diagonal elements of the inverse covariance matrix are negative.) Monotonicity of order 2 is preserved when all functions $E(f_j(X_j)|X_\ell = x)$ attain their maximum (minimum) at the same place. These considerations indicate that some unnatural condition in addition to MDR_2-dependence is required. Moreover, in view of (5.1.2), these arguments make it not surprising that MCA does not necessarily reflect the higher order monotonicity properties.

Nevertheless, MCA reflects the ordering of categories induced by (strict) multiple order dependence of order 1 in the category scores on the first MCA component.

THEOREM 5.2.1. *Suppose that, for given category labels,* $X_1,X_2,\ldots,X_k$ *are* $SMDR_1$*-dependent. Then the variance of the first MCA component is strictly larger than unity and each vector of category scores* $\nu_{\ell 1}$ *of the* ℓ*-th variable on the first MCA component is strictly increasing for* $\ell = 1,\ldots,k$.

PROOF. Let T denote the $n \times n$ block matrix with diagonal blocks $S_{n_1},S_{n_2},\ldots,S_{n_k}$ and all off-diagonal blocks zero. Furthermore, let $\bar{Q}$ denote the $n \times n$ matrix $\bar{Q} = T^{-1}R^{-1}BT$ and let the matrix Q of size $(n-k) \times (n-k)$ be obtained by deleting the rows and columns corresponding to the k indices $n_1,n_1+n_2,\ldots,n_1+n_2+\ldots n_k$. Similarly to the proof of Theorem 3.3.1 it follows from Theorem 5.1.3 that the vectors $\nu_{11},\nu_{21},\ldots,\nu_{k1}$ are strictly monotone of order 1 in the same direction iff the eigenvector of Q corresponding to its largest eigenvalue μ_1 has strictly positive or strictly negative components. Since all bivariate marginal probability tables show SDR_1-dependence, the elements of Q are positive and even strictly positive exept on

diagonal blocks. This implies that Q^2 is STP_1. Application of Lemma 3.1.5 yields that $\nu_{11},\nu_{21},\ldots,\nu_{k1}$ are strictly monotone of order 1 in the same direction and that $\mu_1 > \mu_2$. Since the diagonal elements of Q are unities and since each row has some strictly positive off-diagonal elements, the largest eigenvalue must be strictly larger than unity. □

THEOREM 5.2.2. *Suppose that, for given category labels,* $X_1,X_2,\ldots,X_k$ *are* MDR_1*-dependent, then the vectors with category scores* $\nu_{11},\nu_{21},\ldots,\nu_{k1}$ *of the* k *variables on the first MCA component can be taken monotone of order* 1 *in the same direction.*

PROOF. Follows from Lemma 3.1.6, Theorem 5.2.1 and continuity considerations. □

These theorems show that when k categorical variables are multiple order dependent of order 1, then the (meaningful) ordering of categories induced by the dependence is reflected in the order of the category scores on the first MCA component. This supports the use of $\alpha_{\ell 1}\varphi_{\ell 1}(X_\ell)$ as numerical versions of the nominal variables X_ℓ, for $\ell = 1,\ldots,k$.

5.3. SCALING OF ORDERED LATENT STRUCTURE MODELS

This section discusses an application of MCA in ordered latent structure models. Such models are developed for the following situation which frequently arises in e.g. psychology and medicine. In a population individuals must be ordered according to their value on some *unobservable* characteristic (e.g. intelligence, knowledge of a subject, attitude in a given context, a specific disease). For this purpose responses on a set of variables related to the characteristic are collected for each individual (e.g. an intelligence test). In this section attention is restricted to the simple case in which there is only one such characteristic, called the *latent variable*, and in which the collected responses are dichotomous. The variables involved are called *items* and the set of all items is called the *test*. Since the characteristic of interest is often hard to separate from other characteristics, the assumption that responses on the items systematically depend on only one latent variable is for most applications more restrictive than the dichotomy assumption.

Introducing TP_2 in this situation leads to models with realistic orderings between items. These ordered models are special cases of the

model introduced by MOKKEN (1971). Theorem 5.3.2 shows that the orderings between items in such models are reflected in the MCA scores. Furthermore, it is shown that the TP_2 property is present in well-known examples of latent structure models. The ordering property of MCA, as described in Theorem 5.3.2, has already been noted for these examples by GIFI (1981, Chapter 9), however proofs have not been given.

The latent structure model considered in this section supposes that the responses of the individuals on the k dichotomous items $X_1,X_2,\ldots,X_k$ can be accounted for, to a substantial degree, by an univariate latent variable Z. It is assumed that conditionally on Z the items $X_1,X_2,\ldots,X_k$ are stochastically independent. This assumption of *local independence* is essential in latent structure models. It means that each individual responds independently on the items. Local independence implies that *the (global) dependence structure between the items is caused and hence can completely be explained by variation in the latent variable.*

Let the distribution function of the latent variable Z be denoted by H. Results in this section are not based on any assumption on H and thus the ordering properties of MCA hold for any (sampled) population. Furthermore, the two response categories of each item shall be labeled by 1 ("correct") and 0 ("wrong"). The responses on an item X_ℓ are for each latent value z described by the so called *trace line*

$$\pi_\ell(z) = P\{X_\ell = 1 \,|\, Z = z\} \qquad \text{for } \ell = 1,\ldots,k.$$

This probability of a correct response to item X_ℓ for an individual with latent value z may be interpreted as the (local) difficulty of item X_ℓ for this individual. The unconditional probability of a correct response to item X_ℓ,

$$\pi_\ell = \int_{\mathbb{R}} \pi_\ell(z)dH(z) = P\{X_\ell = 1\} \qquad \text{for } \ell = 1,\ldots,k,$$

is the (global) difficulty of item X_ℓ for the population. By local independence the joint probability of correct responses to both item X_ℓ and item X_j, $\ell \neq j$, for an individual with latent value z equals $\pi_\ell(z)\pi_j(z)$. The unconditional joint probability of correct responses to both items is denoted by

$$\pi_{\ell j} = \int_{\mathbb{R}} \pi_\ell(z)\pi_j(z)dH(z) \qquad \text{for } \ell,j = 1,\ldots,k;\ j \neq \ell;$$

however,

$$\pi_{\ell\ell} = \pi_\ell = \int_{\mathbb{R}} \pi_\ell(z)dH(z) \qquad \text{for } \ell = 1,\ldots,k.$$

Note that the correlation between items X_ℓ and X_j equals

$$\sigma_{\ell j} = (\pi_{\ell j} - \pi_\ell\pi_j)/(\pi_\ell(1-\pi_\ell)\pi_j(1-\pi_j))^{\frac{1}{2}} \qquad \text{for } \ell,j = 1,\ldots,k.$$

Denote by $\sharp = (\sigma_{\ell j})$ the correlation matrix of the items.

MOKKEN (1971) imposes two natural conditions on the trace lines of the items in the test. First, he assumes that for each item the probability of a correct response increases as the individual scores higher on the latent variable, i.e.

(5.3.1) $\quad z_1 < z_2 \Rightarrow \pi_\ell(z_1) \leq \pi_\ell(z_2)$ and not dH-a.e. equality, for $\ell = 1,\ldots,k$.

Secondly, Mokken assumes that if for one individual an item is more difficult than another item, then it must be more difficult for all individuals, i.e.

(5.3.2) the items in the test can be indexed such that
$1 \leq \ell < j \leq k \Rightarrow \pi_\ell(z) \geq \pi_j(z)$ for all z and not dH-a.e. equality.

Indexing is then from easy to difficult. Tests satisfying (5.3.1) and (5.3.2) are called *doubly monotone*. More about interpretation and examples of doubly monotone tests can be found in MOKKEN (1971). In many examples, see for instance at the end of this section, double monotonicity typically occurs in combination with TP_2 of trace lines

(5.3.3) $\quad z_1 < z_2,\ 1 \leq \ell < j \leq k \Rightarrow \pi_\ell(z_1)\pi_j(z_2) \geq \pi_\ell(z_2)\pi_j(z_1),$

and also in combination with a similar TP_2 property with respect to wrong responses,

(5.3.4) $\quad z_1 < z_2,\ 1 \leq \ell < j \leq k \Rightarrow (1-\pi_\ell(z_1))(1-\pi_j(z_2)) \geq (1-\pi_\ell(z_2))(1-\pi_j(z_1)).$

The increasing property (5.3.1) implies that all items are positively dependent, because

(5.3.5) $\quad \pi_{\ell j} - \pi_\ell\pi_j = \frac{1}{2}\iint_{\mathbb{R}^2}(\pi_\ell(z_1)-\pi_\ell(z_2))(\pi_j(z_1)-\pi_j(z_2))dH(z_1)dH(z_2) > 0$

for $\ell,j = 1,\ldots,k$. Thus correlations are even strictly positive. Moreover, it trivially follows from (5.3.2) that

(5.3.6) $\quad 1 \leq \ell < j \leq k \Rightarrow \pi_\ell > \pi_j$ and $\pi_{\ell i} \geq \pi_{ji}$ for $i \neq j$.

Large departures from double monotonicity violate (5.3.5) and (5.3.6) and might be detected by inspection of these properties. Note that (5.3.5) and (5.3.6) only concern properties of the observable items.

Analysis of a latent structure model satisfying the assumptions above with MCA may be motivated by the interpretation of this technique and by Theorem 5.3.2 below. Recall that in the dichotomous case the first MCA component

$$Y_1 = \sum_{\ell=1}^{k} \alpha_{\ell 1}\varphi_{\ell 1}(X_\ell)$$

equals the first principal component of the correlation matrix $\sharp$ of the items $X_1, X_2, \ldots, X_k$. Therefore, Y_1 "best explains" the dependence structure between the items among all linear combinations of items. Since the latent variable completely explains this dependence, Y_1 can be interpreted as the linear combination of the items which best fits the latent variable Z in this sense. The category scores on the first MCA component will be used to analyse the model. Let $\gamma_1 = (\gamma_{11}, \ldots, \gamma_{k1})^\top$ be the vector with these scores for the correct categories of the items, thus $\gamma_{\ell 1} = \alpha_{\ell 1}\varphi_{\ell 1}(1)$ for $\ell = 1, \ldots, k$. Similarly, denote by $\omega_1 = (\omega_{11}, \ldots, \omega_{k1})^\top$, where $\omega_{\ell 1} = \alpha_{\ell 1}\varphi_{\ell 1}(0)$ for $\ell = 1, \ldots, k$, the vector with wrong category scores. Both vectors can be computed by solving an eigenvalue problem.

PROPOSITION 5.3.1. *The components of the vectors* γ_1 *and* ω_1 *satisfy*

$$\gamma_{\ell 1} = ((1-\pi_\ell)/\pi_\ell)^{\frac{1}{2}}\alpha_{\ell 1} \quad \textit{and} \quad \omega_{\ell 1} = -(\pi_\ell/(1-\pi_\ell))^{\frac{1}{2}}\alpha_{\ell 1}$$

for $\ell = 1, \ldots, k$, *where the vector* $\alpha_1 = (\alpha_{11}, \ldots, \alpha_{k1})^\top$ *is an eigenvector, normalized such that* $\alpha_1^\top\alpha_1 = 1$ *and first non-zero component positive, corresponding to the largest eigenvalue of the correlation matrix* $\sharp$ *between the items.*

PROOF. By the normalization conditions (5.1.1), $\varphi_{\ell 1}(1) = ((1-\pi_\ell)/\pi_\ell)^{\frac{1}{2}}$ and $\varphi_{\ell 1}(0) = -(\pi_\ell/(1-\pi_\ell))^{\frac{1}{2}}$ for $\ell = 1, \ldots, k$, and thus the variance of Y_1 is only maximized with respect to the vector α_1 containing the weights $\alpha_{\ell 1}$ for $\ell = 1, \ldots, k$. This is equivalent to finding the first principal component of the covariance matrix of $\varphi_{11}(X_1), \varphi_{21}(X_2), \ldots, \varphi_{k1}(X_k)$, that is of the correlation matrix $\sharp$ of $X_1, X_2, \ldots, X_k$. □

Suppose a subset of items in the test satisfy the double monotonicity and total positivity conditions. Then these items possess strong orderings with respect to their difficulties. The next theorem shows that these orderings are reflected in the correct and wrong scores even when the remaining items do not match the orderings of the items in the subset (i.e. cannot be added to the subset without disturbing double monotonicity or total positivity). Note that by Proposition 5.3.1 the correct and wrong scores do not depend on the order in which items are presented to MCA.

THEOREM 5.3.2. *Suppose the test consists of* m *items which all satisfy* (5.3.1), *with* k *replaced by* m. *Furthermore, suppose* k *of the items, which without loss of generality can be taken as the first* k, *can be ordered such that* (5.3.2) *and* (5.3.3) *hold. Then the correct category scores of these* k *items satisfy*

$$(5.3.7) \qquad 0 < \gamma_{11} < \gamma_{21} < \ldots < \gamma_{k1}.$$

Similarly, if (5.3.1), (5.3.2) *and* (5.3.4) *hold for these first* k *items then*

$$(5.3.8) \qquad \omega_{11} < \omega_{21} < \ldots < \omega_{k1} < 0.$$

PROOF. It follows from Proposition 5.3.1 that $\gamma_1 = (\gamma_{11},\ldots,\gamma_{m1})^T$ is an eigenvector corresponding to the largest eigenvalue of the matrix C with elements

$$c_{\ell j} = (\pi_{\ell j} - \pi_\ell \pi_j)/(\pi_\ell(1-\pi_j)) \qquad \text{for } \ell,j = 1,\ldots,m.$$

Recall that S_k denotes the $k \times k$ upper triangular matrix with unit elements on and above the diagonal and all other elements zero. Denote by T the $m \times m$ block diagonal matrix with diagonal blocks S_k^T and the identity matrix I_{m-k} respectively. Then its inverse T^{-1} is a block diagonal matrix with diagonal blocks S_k^{-1T} and I_{m-k}. Since T is non-singular, γ_1 is an eigenvector corresponding to the largest eigenvalue of C iff $d = T^{-1}\gamma_1$ is an eigenvector corresponding to the largest eigenvalue of $D = T^{-1}CT$.

Under the conditions of the theorem, D turns out to be TP_1 and D^2 even turns out to be STP_1. This can be verified as follows. The elements of the matrix $T^{-1}C = B = (b_{\ell j})$ are given by

$$b_{\ell j} = (\pi_{\ell j} - \pi_\ell \pi_j)/(\pi_\ell(1-\pi_j)) \quad \text{for } \ell = 1,k+1,\ldots,m \text{ and } j = 1,\ldots,m,$$

$$b_{\ell j} = (\pi_{\ell-1}\pi_{\ell j} - \pi_{\ell}\pi_{\ell-1\,j})/(\pi_{\ell}\pi_{\ell-1}(1-\pi_j)) \quad \text{for } \ell = 2,\ldots,k \text{ and } j = 1,\ldots,m.$$

Since (5.3.1) holds for all m items, it follows as in (5.3.5) that all correlations are strictly positive and hence $b_{\ell j} > 0$ for $\ell = 1,k+1,\ldots,m$ and $j = 1,\ldots,m$. Furthermore, by (5.3.4), $\pi_{\ell}^{-1}\pi_{\ell}(z)$ is TP_2 on $\{1,\ldots,k\}\times\mathbb{R}$ and, moreover, satisfies

$$\int_{\mathbb{R}} \pi_{\ell}^{-1}\pi_{\ell}(z)dH(z) = 1 \qquad \text{for } \ell = 1,\ldots,k.$$

Therefore, since $\pi_j(z)$ is increasing in z for each $j = 1,\ldots,m$, it follows by Proposition 3.1 in KARLIN (1968, p. 22) or by Theorem 3.2.3 that for each $j = 1,\ldots,m$, $\int \pi_{\ell}^{-1}\pi_{\ell}(z)\pi_j(z)dH(z)$ is increasing in ℓ, i.e.

$$\pi_{\ell-1}\pi_{\ell j} - \pi_{\ell}\pi_{\ell-1\,j} \geq 0 \quad \text{for } \ell = 2,\ldots,k \text{ and all } j \neq \ell-1,\ell.$$

Hence $b_{\ell j} \geq 0$ for $\ell = 2,\ldots,k$ and all $j \neq \ell-1,\ell$. Obviously, $b_{\ell\ell} > 0$ and $b_{\ell\ell-1} < 0$ for $\ell = 2,\ldots,k$. Thus the matrix B has positive elements exept for $b_{\ell\ell-1}$, $\ell = 2,\ldots,k$. But by (5.3.2), $\pi_{\ell-1} > \pi_{\ell}$ for $\ell = 2,\ldots,k$ and it easily follows that $b_{\ell\ell-1} + b_{\ell\ell} > 0$ for $\ell = 2,\ldots,k$. Therefore, $D = BT$ is TP_1. Moreover, since $b_{\ell j} > 0$ for $\ell = 1,k+1,\ldots,m$ and $j = 1,\ldots,m$ and since $b_{\ell\ell-1} + b_{\ell\ell} > 0$ for $\ell = 2,\ldots,k$, it follows that the elements in the first row of D and the elements in the first column of D are strictly positive. This implies that D^2 is STP_1.

Application of the Perron-Frobenius theorem (cf. GANTMACHER, 1977, vol II, p. 53) or of Lemma 3.1.5 with $s = 1$ yields that the eigenvector $d = (d_1,\ldots,d_m)^T$ corresponding to the largest eigenvalue of the matrix D has strictly positive components, i.e. $d_{\ell} > 0$ for $\ell = 1,\ldots,m$. Since $d = T^{-1}\gamma_1$ or equivalently $d_{\ell} = \gamma_{\ell 1}$ for $\ell = 1,k+1,\ldots,m$ and $d_{\ell} = \gamma_{\ell 1} - \gamma_{\ell-1\,1}$ for $\ell = 2,\ldots,k$ the result (5.3.7) follows. The proof of (5.3.8) is similar. □

The result (5.3.7) also holds when (5.3.2) and (5.3.3) are relaxed to

(5.3.9) $$1 \leq \ell < j \leq k \Rightarrow \pi_{\ell} > \pi_j$$

and

(5.3.10) $$\int_{-\infty}^{z} \pi_{\ell-1}^{-1}\pi_{\ell-1}(\zeta)dH(\zeta) \geq \int_{-\infty}^{z} \pi_{\ell}^{-1}\pi_{\ell}(\zeta)dH(\zeta) \text{ for all } z \in \mathbb{R} \text{ and } \ell = 2,\ldots,k;$$

however, the latter assumption involves the distribution H of Z.

Theorem 5.3.2 shows that the MCA correct and wrong category scores

reflect the difficulties of the items. This in combination with (5.3.5) and (5.3.6) can be used for a first investigation of the model assumptions. Moreover, Theorem 5.3.2 suggests that ordering the individuals according to their MCA test score Y_1 is reasonable: responding a difficult item correctly yields a large contribution to this test score and responding it wrongly does not cost much, whereas for an easy item it is the other way around. The test score Y_1 is a weighted sum of items with small weights for items which are less related to the latent variable than the other items. This procedure is an alternative to the method proposed by MOKKEN (1971) in which individuals are ordered according to the (unweighted) sum of correct responses. It is unknown, however, in which cases (exept in the Rasch model; see below) which procedure actually works better. An advantage of the MCA approach is that a natural generalization of Theorem 5.3.2 to the case of items with three or more ordered categories can be given.

In practice the probabilities π_ℓ for $\ell = 1,\ldots,k$ and $\pi_{\ell j}$ for $\ell,j = 1,\ldots,k$ have to be estimated by the relative frequencies of correct responses. Although the MCA scores based on these estimates need not reflect the difficulties of the items even when the underlying model satisfies the conditions of Theorem 5.3.2, one would expect the total score Y_1 to reflect the appropriate ordering of the individuals quite well.

Latent structure models for dichotomous variables studied in the literature (e.g. ANDERSEN, 1980; FISCHER, 1974; LORD and NOVICK, 1968) are commonly of parametric form, that is, the functional form of the trace lines is specified. However, there is often no evidence that the specific functional form is actually present in the test at hand. The parametric examples below generally satisfy the double monotonicity and total positivity conditions and therefore analysis both with MCA or with Mokken's method is legitimate. These examples are also discussed in MOKKEN (1971). Moreover, the ordering properties of MCA under these parametric models have already been noted by GIFI (1981, Chapter 9), but proofs where not given.

In *Guttman's model* the responses on the items are deterministic functions of the latent variable. The trace lines are given by

$$\pi_\ell(z) = \begin{cases} 1 & \text{if } z \geq \delta_\ell \\ 0 & \text{if } z < \delta_\ell \end{cases} \quad \text{for } \ell = 1,\ldots,k,$$

where the item parameters satisfy $\delta_1 < \delta_2 < \ldots < \delta_k$. In this model an

individual cannot respond correctly to a difficult item and wrongly to an easy item. Hence a perfect analysis is possible. The conditions (5.3.1), (5.3.2), (5.3.3) and (5.3.4) are easily verified and therefore the results of Theorem 5.3.2 hold. Moreover, stronger ordering properties of the correct and wrong categories scores on higher MCA components hold. Denote by $\gamma_t = (\gamma_{1t},\ldots,\gamma_{kt})^\top$ the vector with correct category scores on the t-th MCA component, so $\gamma_{\ell t} = \alpha_{\ell t}\varphi_{\ell t}(1)$ for $\ell = 1,\ldots,k$ and $t = 1,\ldots,k$, and write $\omega_t = (\omega_{1t},\ldots,\omega_{kt})^\top$ with $\omega_{\ell t} = \alpha_{\ell t}\varphi_{\ell t}(0)$ for $\ell = 1,\ldots,k$ and $t = 1,\ldots,k$.

THEOREM 5.3.3. *Consider Guttman's model. For arbitrary real numbers* $a_r,a_{r+1},\ldots,a_s$ $(1 \le r \le s \le k,\ \Sigma_{t=r}^{s} a_t^2 > 0)$ *the numbers of changes of sign of the vectors*

$$\gamma = \sum_{t=r}^{s} a_t\gamma_t, \quad \omega = \sum_{t=r}^{s} a_t\omega_t \quad \textit{and}\ \alpha = \sum_{t=r}^{s} a_t\alpha_t$$

satisfy

$$r-1 \le S^-(\gamma) \le S^+(\gamma) \le s-1, \quad r-1 \le S^-(\omega) \le S^+(\omega) \le s-1,$$

$$r-1 \le S^-(\alpha) \le S^+(\alpha) \le s-1.$$

PROOF. First consider the assertion for the vector α. By Proposition 5.3.1, the vectors $\alpha_1,\alpha_2,\ldots,\alpha_k$ are eigenvectors of the correlation matrix $\sharp$ corresponding to eigenvalues $\mu_1 \ge \mu_2 \ge \ldots \ge \mu_k$. In Guttman's model, $\pi_{j\ell} = \min(\pi_j,\pi_\ell)$ for $j,\ell = 1,\ldots,k$ and hence the correlations between the items are given by

$$\sigma_{\ell j} = \begin{cases} \pi_j(1-\pi_\ell)/(\pi_\ell(1-\pi_j))^{\frac{1}{2}} & \text{if } \ell \le j \\ \pi_\ell(1-\pi_j)/(\pi_j(1-\pi_\ell))^{\frac{1}{2}} & \text{if } \ell > j \end{cases}.$$

Therefore, $\sharp$ is a so called Green's matrix which by a result in KARLIN (1968, p. 111) is TP_k and has some iterate which is STP_k iff $\pi_1^{-1}(1-\pi_1) < \pi_2^{-1}(1-\pi_2) < \ldots < \pi_k^{-1}(1-\pi_k)$ or equivalently iff $\pi_1 > \pi_2 > \ldots > \pi_k$. Application of Lemma 3.1.5 gives the desired result. By Proposition 5.3.1, vectors γ_t and ω_t are eigenvectors of the matrices $C = (c_{\ell j})$ and $W = (w_{\ell j})$ respectively with elements $c_{\ell j} = (\pi_{\ell j} - \pi_\ell\pi_j)/(\pi_\ell(1-\pi_j))$ and $w_{\ell j} = (\pi_{\ell j} - \pi_\ell\pi_j)/(\pi_j(1-\pi_\ell))$ for $j,\ell = 1,\ldots,k$. By the same arguments as above the result follows for the vectors γ and ω. □

The practical relevance of these stronger ordering properties is limited. A slightly weaker result, with respect to the principal components

$\alpha_1,\alpha_2,\ldots,\alpha_k$ has been proved by GUTTMAN (1950) and an interpretation is given in GUTTMAN (1954); cf. GIFI (1981).

A somewhat more realistic generalization of the previous model is the *latent distance model* of LAZARSFELD and HENRY (1968). Its trace lines satisfy

$$\pi_\ell(z) = \begin{cases} 1-\varepsilon_{1\ell} & \text{if } z \geq \delta_\ell \\ \varepsilon_{2\ell} & \text{if } z < \delta_\ell \end{cases} \qquad \text{for } \ell = 1,\ldots,k$$

where $\delta_1 < \delta_2 < \ldots < \delta_k$ and $\varepsilon_{2\ell} < 1-\varepsilon_{1\ell}$ for $\ell = 1,\ldots,k$. If $\varepsilon_{2\ell} > 0$, $\varepsilon_{1\ell} > 0$ for $\ell = 1,\ldots,k$ then the double monotonicity and total positivity conditions never hold simultaneously. But the weaker conditions (5.3.9) and (5.3.10) are satisfied when $\pi_{\ell-1} > \pi_\ell$, $\pi_\ell\varepsilon_{2\ell-1} \geq \pi_{\ell-1}\varepsilon_{2\ell}$, $\pi_\ell(1-\varepsilon_{1\ell-1}) \geq \pi_{\ell-1}(1-\varepsilon_{1\ell})$ for $\ell = 2,\ldots,k$ and hence (5.3.7) remains valid.

In the *linear model* of Lazarsfeld the trace lines are given by $\pi_\ell(z) = a_\ell z + b_\ell$ provided $0 \leq a_\ell z + b_\ell \leq 1$ for $\ell = 1,\ldots,k$. The conditions (5.3.1) and (5.3.2) are for instance satisfied when $a_{\ell-1} \leq a_\ell$ and $b_{\ell-1} \geq b_\ell$ for $\ell = 2,\ldots,k$. The total positivity conditions (5.3.3) and (5.3.4) hold when $a_{\ell-1}b_\ell \leq a_\ell b_{\ell-1}$ and $a_{\ell-1}(1+b_\ell) \geq a_\ell(1+b_{\ell-1})$ for $\ell = 2,\ldots,k$.

RASCH (1960) developped a model in which the (unweighted) sum of all correct responses is sufficient for Z. The trace lines for $\ell = 1,\ldots,k$ satisfy $\pi_\ell(z) = z/(z+\delta_\ell)$ (or = 0) when $z \geq 0$ (or $z < 0$), and where $0 < \delta_1 < \ldots < \delta_k$. This model is actually a special case of a model considered by Birnbaum. Moreover, Birnbaum proved that in this model Mokken's (unweighted) sum of correct responses "uniformly best discriminates" the individuals (see MOKKEN, 1971, p. 141). Double monotonicity and total positivity of trace lines is easily verified.

The last example consists of models based on shifts in distribution functions. For an (univariate) distribution function F the trace lines are defined by $\pi_\ell(z) = F(z-\delta_\ell)$ for $\ell = 1,\ldots,k$ and where $\delta_1 < \delta_2 < \ldots < \delta_k$. Double monotonicity is obvious. It follows from Corollary 3.1.3 and Example 3.1.1 that the TP_2 conditions (5.3.3) and (5.3.4) hold whenever the density p of F (with respect to some measure) is PF_2. Special choices of such PF_2 distribution functions F yield well-known models, e.g. degenerate distribution (Guttman's model), logistic distribution (Rasch's model), normal distribution (models of LAWLEY, 1943 and LORD, 1952). Other PF_2 distribution functions are the gamma, Poisson and binomial distribution function.

REFERENCES

[1] ANDERSEN, E.B. (1980). *Discrete Statistical Models with Social Science Applications*. Amsterdam: North-Holland Publ. Co.

[2] ANDERSON, T.W. (1958). *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.

[3] BAHADUR, R.R. (1960). Stochastic comparison of tests. *Ann. Math. Statist.* 31, pp. 276-295.

[4] BENZÉCRI, J.P. (1973). *L'Analyse des Données* II: *l'Analyse des Correspondances*. Paris: Dunod. (2nd edition, 1976).

[5] BENZÉCRI, J.P. (1982). *Histoire et Prehistoire de l'Analyse des Données*. Paris: Dunod.

[6] BURT, C. (1950). The factorial analysis of qualitative data. *British J. of Psychology, Statist. Section* 3, pp. 166-185.

[7] CAMBANIS, S., G. SIMONS & W. STOUT (1976). Inequalities for $Ek(X,Y)$ when the marginals are fixed. *Z. Wahrsch. verw. Gebiete* 36, pp. 285-294.

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

[9] CHESSON, P.L. (1976). The canonical decomposition of bivariate distributions. *J. Mult. Analysis* 6, pp. 526-537.

[10] CLAYTON, D. & J. CUZICK (1985). Multivariate generalizations of the proportional hazards model. *J. Royal Statist. Soc. A* 148, to appear.

[11] CORSTEN, L.C.A. (1976). Matrix approximation, a key to application of multivariate methods. In: *Proc. 9th Biometric Conf.* 1. Raleigh, N.C.: Biometric Society, pp. 61-77.

[12] DAS GUPTA, S., T.W. ANDERSON & G.S. MUDHOLKAR (1964). Monotonicity of the power functions of some tests of the multivariate linear hypothesis. *Ann. Math. Statist.* 35, pp. 200-205.

[13] DAUXOIS, J. & A. POUSSE (1977). Some convergence problems in factor analysis. In: J.R. Barra et al. (eds.). *Recent Developments in Statistics*. Amsterdam: North-Holland Publ. Co., pp. 387-402.

[14] DE LEEUW, J. (1973). *Canonical Analysis of Categorical Data.* Psychological Institute, University of Leiden.

[15] DE LEEUW, J. (1983). On the prehistory of correspondence analysis. *Statistica Neerlandica* 37, pp. 161-164.

[16] DE LEEUW, J. (1984). *Statistical Properties of Multiple Correspondence Analysis.* Paper presented at the conference on "New Multivariate Methods in Statistics". The 1984 Joint Summer Research Conference Series in Mathematical Sciences. June 10-16, Bowdoin College, Brunswick, Maine. (Report RR-84-06, Dept. of Data Theory, University of Leiden).

[17] DE LEEUW, J. (1984 b). The Gifi system of nonlinear multivariate analysis. In: E. Diday et al. (eds.). *Data Analysis and Informatics III.* Amsterdam: North-Holland Publ. Co., pp. 415-424.

[18] EAGLESON, G.K. (1964). Polynomial expansions of bivariate distributions. *Ann. Math. Statist.* 35, pp. 1208-1215.

[19] EAGLESON, G.K. & H.O. LANCASTER (1967). The regression system of sums with random elements in common. *Austr. J. Statist.* 9, pp. 119-125.

[20] EATON, M.L. (1982). A review of selected topics in multivariate probability inequalities. *Ann. Statist.* 10, pp. 11-43.

[21] ERDÉLYI, A. (1956). *Asymptotic Expansions.* New York: Dover Publishing Inc.

[22] ESARY, J.D., F. PROSCHAN & D.W. WALKUP (1967). Association of random variables, with applications. *Ann. Math. Statist.* 38, pp. 1466-1474.

[23] FIENBERG, S.E. (1977). *The Analysis of Cross-Classified Categorical Data.* Cambridge (Mass): MIT Press.

[24] FISCHER, G.H. (1974). *Einfürung in die Theorie der Psychologischer Tests.* Bern: Huber.

[25] FISHER, R.A. (1936). The use of multiple measurements in taxonomic problems. *Ann. Eugen.* 7, pp. 179-188.

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

[27] GABRIEL, K.R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58, pp. 453-467.

[28] GANTMACHER, F.R. (1977). *Matrix Theory* I, II. New York: Chelsea.

[29] GANTMACHER, F.R. & M.G. KREIN (1950). *Oscillation Matrices and Kernels and Small Vibrations of Mechanical Systems*. Transl. from Russian, issued (1961), AEC-tr-4481. Washington: US Atomic Energy Commision.

[30] GAUCH, H.G. (1982). *Multivariate Analysis in Community Ecology*. Cambridge: Cambridge University Press.

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

[32] GIFI, A. (1981). *Non-linear Multivariate Analysis*. Dept. of Data Theory, University of Leiden.

[33] GOODMAN, L.A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *J. Amer. Statist. Assoc.* 74, pp. 537-552.

[34] GOODMAN, L.A. (1981). Association models and canonical correlation in the analysis of cross-classifications having ordered categories. *J. Amer. Statist. Assoc.* 76, pp. 320-334.

[35] GOODMAN, L.A. (1985). The analysis of cross-classified data having ordered and/or unordered categories: association models, correlation models and asymmetry models for contingency tables with or without missing entries. *Ann. Statist.* 13, pp. 10-69.

[36] GOODMAN, L.A. & W.H. KRUSKAL (1954). Measures of association for cross-classifications. *J. Amer. Statist. Assoc.* 49, pp. 732-764.

[37] GOODMAN, L.A. & W.H. KRUSKAL (1963). Measures of association for cross-classifications III: approximate sampling theory. *J. Amer. Statist. Assoc.* 58, pp. 310-364.

[38] GOODMAN, L.A. & W.H. KRUSKAL (1972). Measures of association for cross-classifications IV: simplification of asymptotic variances. *J. Amer. Statist. Assoc.* 67, pp. 415-421.

[39] GREENACRE, M.J. (1981). Practical correspondence analysis. In: V. Barnett (ed.). *Interpreting Multivariate Data*. New York: Wiley, pp. 119-146.

[40] GREENACRE, M.J. (1984). *Theory and Applications of Correspondence Analysis*. New York: Academic Press.

[41] GROSS, S.T. (1981). On asymptotic power and efficiency of tests of independence in contingency tables with ordered classifications. *J. Amer. Statist. Assoc.* 76, pp. 935-941.

[42] GUMBEL, E.J. (1961). Bivariate logistic distributions. *J. Amer. Statist. Assoc.* 56, pp. 335-349.

[43] GUTTMAN, L. (1941). The quantification of a class of attributes: a theory and method of scale construction. In: The Committee on Social Adjustment (ed.). *The Prediction of Personal Adjustment*. New York: Social Science Research Council, pp. 319-348.

[44] GUTTMAN, L. (1950). The principal components of scale analysis. In: S.A. Stouffer et al.. *Measurement and Prediction*. Princeton: Princeton University Press, pp. 312-361.

[45] GUTTMAN, L. (1954). The principal components of scalable attitudes. In: P.F. Lazarsfeld (ed.). *Mathematical Thinking in the Social Sciences*. Glencoe: Free Press, pp. 216-257.

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

[47] HABERMAN, S.J. (1974 b). Log-linear models for frequency tables with ordered classifications. *Biometrics* 36, pp. 589-600.

[48] HABERMAN, S.J. (1981). Tests for independence in two-way contingency tables based on canonical correlation and on linear-by-linear interaction. *Ann. Statist.* 9, pp. 1178-1186.

[49] HÁJEK, J. & Z. ŠIDÁK (1967). *Theory of Rank Tests*. New York: Academic Press.

[50] HILL, M.O. (1973). Reciprocal averaging: an eigenvector method of ordination. *J. Ecol.* 61, pp. 237-251.

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

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

[53] HOEFFDING, W. (1948). A non-parametric test of independence. *Ann. Math. Statist.* 19, pp. 546-557.

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

[55] HORST, P. (1935). Measuring complex attitudes. *J. of Social Psychology* 6, pp. 369-374.

[56] JOHNSON, N.L. (1960). An approximation to the multinomial distribution: some properties and applications. *Biometrika* 47, pp. 93-102.

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

[58] KARLIN, S. & J.L. McGREGOR (1959). Coincidence properties of birth and death processes. *Pasific J. Math.* 9, pp. 1109-1140.

[59] KARLIN, S. & Y. RINOTT (1980). Classes of orderings of measures and related correlation inequalities I: Multivariate totally positive distributions. *J. Mult. Analysis* 10, pp. 467-498.

[60] KARLIN, S. & W.J. STUDDEN (1966). *Tchebycheff Systems: with Applications in Analysis and Statistics.* New York: Interscience Publishers (Wiley).

[61] KENDALL, D.G. (1971). Seriation from abundance matrices. In: Hodson, F.R., D.G. Kendall & P. Tautu (eds.). *Mathematics in the Archeological and Historical Sciences.* Edinburgh: University Press, pp. 215-252.

[62] KENDALL, M.G. (1975). *Rank Correlation Methods* (4th edition). London: Griffin.

[63] KESTER, N.K. & B.F. SCHRIEVER (1982). *Analysis of Association of Categorical Variables by Numerical Scores and Graphical Representation.* Report SW 85/82. Amsterdam: Mathematisch Centrum.

[64] KONIJN, H.S. (1956). On the power of certain tests for independence in bivariate populations. *Ann. Math. Statist.* 27, pp. 300-323.

[65] KOWALCZYK, T. & E. PLESZCZYŃSKA (1977). Monotonic dependence functions of bivariate distributions. *Ann Statist.* 6, pp. 1221-1227.

[66] LANCASTER, H.O. (1958). The structure of bivariate distributions. *Ann. Math. Statist.* 29, pp. 719-736.

[67] LANCASTER, H.O. (1963). Correlations and canonical forms of bivariate distributions. *Ann. Math. Statist.* 34, pp. 532-538.

[68] LANCASTER, H.O. (1975). Joint probability distributions in the Meixner classes. *J. Royal Statist. Soc. B* 37, pp. 434-443.

[69] LANCASTER, H.O. (1980). Orthogonal models for contingency tables. In: P.R. Krishnaiah (ed.). *Developments in Statistics* 3. New York: Academic Press, pp. 99-157.

[70] LAWLEY, D.N. (1943). On problems connected with item selection and test construction. *Proc. Royal Society of Edinburgh* 61, pp. 273-287.

[71] LAZARSFELD, P.F. & N.W. HENRY (1968). *Latent Structure Analysis.* Boston: Houghton-Mifflin.

[72] LEBART, L., A. MORINEAU & N. TABARD (1977). *Techniques de la Description Statistique.* Paris: Dunod. (English translation: L. Lebart, A. Morineau & K.M. Warwick (1984). *Multivariate Descriptive Statistical Analysis.* New York: Wiley.)

[73] LEHMANN, E.L. (1966). Some concepts of dependence. *Ann. Math. Statist.* 37, pp. 1137-1153.

[74] LEHMANN, E.L. (1975). *Nonparametrics: Statistical Methods based on Ranks.* San Francisco: Holden-Day.

[75] LORD, F.M. (1952). *A Theory of Test Scores.* Psychometric Monograph No 7.

[76] LORD, F.M. & M.R. NOVICK (1968). *Statistical Theories of Mental Test Scores.* Reading (Mass.): Addison Wesley.

[77] MARSHALL, A.W. & I. OLKIN (1979). *Inequalities: Theory of Majorization and Its Applications.* New York: Academic Press.

[78] MAUNG, K. (1941). Measurement of association in a contingency table with special reference to the pigmentation of hair and eye colours of Scottish school children. *Ann. Eugen.* 11, pp. 189-223.

[79] MARDIA, K.V. (1962). Multivariate Pareto distributions. *Ann. Math. Statist.* 33, pp. 1008-1015.

[80] McCULLAGH, P. (1978). A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika* 65, pp. 413-418.

[81] McCULLAGH, P. (1980). Regression models for ordinal data. *J. Royal Statist. Soc. B* 42, pp. 109-142.

[82] MOKKEN, R.J. (1971). *A Theory and Procedure of Scale Analysis.* The Hague: Mouton.

[83] MUIRHEAD, R.J. (1982). *Aspects of Multivariate Analysis.* New York: John Wiley and Sons.

[84] NAKHLÉ, F. (1976). Sur l'analyse d'un tableau de notes dédoublées. *Les Cahiers de l'Analyse des Données* I. *No* 3, pp. 243-257.

[85] NAOURI, J.C. (1970). Analyse factorielle des correspondances continues. *Publ. Inst. Statist. Univ. Paris* 19, pp. 1-100.

[86] NGUYEN, T.T. (1982). *Positive Dependence of Bivariate Discrete Distributions.* Ph. D. thesis, University of Pittsburgh.

[87] NISHISATO, S. (1980). *Analysis of Categorical Data: Dual Scaling and its Applications.* Toronto: University of Toronto Press.

[88] O'NEILL, M.E. (1978). Asymptotic distributions of the canonical correlation from contingency tables. *Austr. J. Statist.* 20, pp. 75-82.

[89] O'NEILL, M.E. (1978 b). Distributional expansions for canonical correlations from contingency tables. *J. Royal Statist. Soc. B* 40, pp. 303-312.

[90] PERLMAN, M.D. & I. OLKIN (1980). Unbiasedness of invariant tests for MANOVA and other multivariate problems. *Ann. Statist.* 8, pp. 1326-1341.

[91] RAO, C.R. (1973). *Linear Statistical Inference and its Applications.* New York: Wiley.

[92] RASCH, G. (1960). *Probabilistic Models for some Intelligence and Attainment Tests.* Copenhagen: Danish Institute for Educational Research. (Reprinted 1980: University of Chicago Press.)

[93] RÉNYI, A. (1959). On measures of dependence. *Acta Math. Acad. Sc. Hungar* 10, pp. 441-451.

[94] RICHARDSON, M. & G.F. KUDER (1933). Making a rating scale that measures. *Personnel Journal* 12, pp. 36-40.

[95] ROTHE, G. (1981). Some properties of the asymptotic relative Pitman efficiency. *Ann. Statist.* 9, pp. 663-669.

[96] RUYMGAART, F.H. (1973). *Asymptotic Theory of Rank Tests for Independence*. Mathematical Centre Tract 43. Amsterdam: Mathematisch Centrum.

[97] SCHRIEVER, B.F. (1983). Scaling of order dependent categorical variables with correspondence analysis. *Intern. Statist. Review* 51, pp. 225-238.

[98] SNIJDERS, T.A.B. (1979). *Asymptotic Optimality Theory for Testing Problems with Restricted Alternatives*. Mathematical Centre Tract 113. Amsterdam: Mathematisch Centrum.

[99] TCHEN, A.H-T. (1976). *Inequalities for Distributions with given Marginals*. Technical Report 19, Dept. of Statist. Stanford University.

[100] TONG, Y.L. (1980). *Probability Inequalities in Multivariate Distributions*. New York: Academic Press.

[101] VAN ZWET, W.R. (1968). On convexity preserving families of probability distributions. *Statistica Neerlandica* 22, pp. 23-32.

[102] WHITT, W. (1982). Multivariate monotone likelihood ratio and uniform conditional stochastic order. *J. Appl. Prob.* 19, pp. 695-701.

[103] WHITTAKER, R.H. (1967). Gradient analysis of vegetation. *Biological Reviews* 42, pp. 207-264.

[104] WIEAND, H.S. (1976). A condition under which the Pitman and Bahadur approaches to efficiency coincide. *Ann. Statist.* 4, pp. 1003-1011.

[105] WILKINSON, J.H. (1965). *The Algebraic Eigenvalue Problem*. Oxford: Clarendon.

[106] YATES, F. (1948). The analysis of contingency tables with groupings based on quantitative characters. *Biometrika* 35, pp. 176-181.

SYMBOL INDEX

P, P'	probability tables; 6, 8, 21, 78
$P^{(N)}$, $P'^{(N)}$	underlying probability tables; 17, 21
$\hat{P}^{(N)}$, $\hat{P}'^{(N)}$	empirical probability tables; 17, 82
$P_{j\ell}$	bivariate marginal probability table; 85
p	density of F; 6
$P\{\ldots\}$	probability on event
P_0, P_L, P_A	sequences of null-hyp., loc. alt. and fixed alt. cond. Poisson distr.; 21, 22
Q_r, Q_c	matrices; 47
q_r, q_c	kernels; 42
R, $R^{(N)}$, $\hat{R}^{(N)}$	diagonal matrices containing row sums of P, $P^{(N)}$ and $\hat{P}^{(N)}$ respectively; 8
r, $r^{(N)}$, $\hat{r}^{(N)}$	vectors containing row sums of P, $P^{(N)}$ and $\hat{P}^{(N)}$ respectively; 8
R	super diagonal matrix containing marginals of $X_1,\ldots,X_k$; 88
R_ℓ	diagonal matrix containing marginals of X_ℓ; 85
S, S_n	upper triangular matrix; 5
$s_\cdot(\cdot)$	approximate Bahadur slope; 30
$S^-(\cdot)$, $S^+(\cdot)$, $S_\sigma^-(\cdot)$	numbers of sign changes; 38
$SM(\cdot)$	strict order of monotonicity; 44
$T_\cdot^{(N)}$	test statistic; 26, 27, 33, 55, 56, 57
U, U'	uniform random variables; 13, 78
V, V'	uniform random variables; 13, 78
$Var(\cdot)$	variance; 6
$vec[\cdot]$	5
W, W', $W^{(N)}$, $\hat{W}^{(N)}$	matrices derived from P, P', $P^{(N)}$ and $\hat{P}^{(N)}$ respectively; 17, 18
$W_\cdot(\cdot,\cdot,\cdot)$	Wishart distribution; 22
X, X'	row variables; 6, 8, 40, 61
X	standard normally distributed matrix; 22
X_ℓ	categorical variable; 85
X, X'	subsets of real line, supports; 35, 61
Y, Y'	column variables; 6, 8, 40, 61
Y_t	t-th MCA component; 85
y, y'	subsets of real line, supports; 35, 61
$Z^{(N)}$	random matrix; 17
Z	latent variable; 93

$\alpha_{\ell t}$	MCA variable weight; 85, 86
$B(\cdot,\cdot)$	beta distribution; 33, 34
γ_t	vector of MCA correct category scores; 95, 99
δ_ℓ	item parameter; 98-100
η_j	CA column point; 14
θ	contamination factor; 22
$\kappa, \kappa_1, \kappa_2$	increasing functions ; 64
$\lambda_t, \lambda_t^{(N)}, \hat{\lambda}_t^{(N)}$	t-th canonical correlation of P, $P^{(N)}$ and $\hat{P}^{(N)}$ respectively; 9, 16, 21
Λ	diagonal matrix containing canonical correlations; 19, 22
μ_t	variance of t-th MCA component; 85
$\nu_\ell, \nu_{\ell t}$	vectors of MCA category scores; 88
$\xi_i, \xi_{i\ell}$	CA row point, MCA category point; 14, 90
Π_0	set of null-hypothesis probability tables; 21
$\pi_\ell(\cdot)$	trace line; 93
$\pi_\ell, \pi_{\ell j}$	probabilities; 93
$\varphi_t(\cdot), \varphi_t$	t-th canonical row function (vector) of P; 9, 10
$\varphi_t^{(N)}, \hat{\varphi}_t^{(N)}$	t-th canonical row functions of $P^{(N)}$ and $\hat{P}^{(N)}$ respectively; 16, 21
$\varphi_{\ell t}(\cdot)$	MCA function; 85, 86
$\phi_\cdot(\cdot)$	measure of dependence; 14, 57, 71, 72, 74, 75
$\tilde{\phi}_\cdot(\cdot)$	probability table version of $\phi_\cdot(\cdot)$; 79
$\psi_t(\cdot), \psi_t$	t-th canonical column function (vector) of P; 9, 10
$\psi_t^{(N)}, \hat{\psi}_t^{(N)}$	t-th canonical functions of $P^{(N)}$ and $\hat{P}^{(N)}$ respectively; 16, 21
ψ	function; 71
ω_t	vector of MCA wrong category scores; 95, 99
^{T}	transposed of matrix or vector; 5
$\circledast$	Kronecker matrix product; 5
$\sim$	"is distributed as"; 6
$\xrightarrow{P}$	convergence in probability; 6
$\xrightarrow{D}$	convergence in distribution; 6
$\dot{\geq}$	ordering for positive dependence; 62, 64
$\doteq$	equivalence relation for positive dependence; 62
$\tilde{\geq}$	ordering for positive dependence in probability table; 78
$\tilde{=}$	equivalence relation for positive dependence in probability table; 78
$\#\{\ldots\}$	number of elements in set (cardinal number)

SUBJECT INDEX

MATHEMATICAL CENTRE TRACTS

1 T. van der Walt. *Fixed and almost fixed points.* 1963.

2 A.R. Bloemena. *Sampling from a graph.* 1964.

3 G. de Leve. *Generalized Markovian decision processes, part I: model and method.* 1964.

4 G. de Leve. *Generalized Markovian decision processes, part II: probabilistic background.* 1964.

5 G. de Leve, H.C. Tijms, P.J. Weeda. *Generalized Markovian decision processes, applications.* 1970.

6 M.A. Maurice. *Compact ordered spaces.* 1964.

7 W.R. van Zwet. *Convex transformations of random variables.* 1964.

8 J.A. Zonneveld. *Automatic numerical integration.* 1964.

9 P.C. Baayen. *Universal morphisms.* 1964.

10 E.M. de Jager. *Applications of distributions in mathematical physics.* 1964.

11 A.B. Paalman-de Miranda. *Topological semigroups.* 1964.

12 J.A.Th.M. van Berckel, H. Brandt Corstius, R.J. Mokken, A. van Wijngaarden. *Formal properties of newspaper Dutch.* 1965.

13 H.A. Lauwerier. *Asymptotic expansions.* 1966, out of print; replaced by MCT 54.

14 H.A. Lauwerier. *Calculus of variations in mathematical physics.* 1966.

15 R. Doornbos. *Slippage tests.* 1966.

16 J.W. de Bakker. *Formal definition of programming languages with an application to the definition of ALGOL 60.* 1967.

17 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 1.* 1968.

18 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 2.* 1968.

19 J. van der Slot. *Some properties related to compactness.* 1968.

20 P.J. van der Houwen. *Finite difference methods for solving partial differential equations.* 1968.

21 E. Wattel. *The compactness operator in set theory and topology.* 1968.

22 T.J. Dekker. *ALGOL 60 procedures in numerical algebra, part 1.* 1968.

23 T.J. Dekker, W. Hoffmann. *ALGOL 60 procedures in numerical algebra, part 2.* 1968.

24 J.W. de Bakker. *Recursive procedures.* 1971.

25 E.R. Paërl. *Representations of the Lorentz group and projective geometry.* 1969.

26 European Meeting 1968. *Selected statistical papers, part I.* 1968.

27 European Meeting 1968. *Selected statistical papers, part II.* 1968.

28 J. Oosterhoff. *Combination of one-sided statistical tests.* 1969.

29 J. Verhoeff. *Error detecting decimal codes.* 1969.

30 H. Brandt Corstius. *Exercises in computational linguistics.* 1970.

31 W. Molenaar. *Approximations to the Poisson, binomial and hypergeometric distribution functions.* 1970.

32 L. de Haan. *On regular variation and its application to the weak convergence of sample extremes.* 1970.

33 F.W. Steutel. *Preservation of infinite divisibility under mixing and related topics.* 1970.

34 I. Juhász, A. Verbeek, N.S. Kroonenberg. *Cardinal functions in topology.* 1971.

35 M.H. van Emden. *An analysis of complexity.* 1971.

36 J. Grasman. *On the birth of boundary layers.* 1971.

37 J.W. de Bakker, G.A. Blaauw, A.J.W. Duijvestijn, E.W. Dijkstra, P.J. van der Houwen, G.A.M. Kamsteeg-Kemper, F.E.J. Kruseman Aretz, W.L. van der Poel, J.P. Schaap-Kruseman, M.V. Wilkes, G. Zoutendijk. *MC-25 Informatica Symposium.* 1971.

38 W.A. Verloren van Themaat. *Automatic analysis of Dutch compound words.* 1972.

39 H. Bavinck. *Jacobi series and approximation.* 1972.

40 H.C. Tijms. *Analysis of (s,S) inventory models.* 1972.

41 A. Verbeek. *Superextensions of topological spaces.* 1972.

42 W. Vervaat. *Success epochs in Bernoulli trials (with applications in number theory).* 1972.

43 F.H. Ruymgaart. *Asymptotic theory of rank tests for independence.* 1973.

44 H. Bart. *Meromorphic operator valued functions.* 1973.

45 A.A. Balkema. *Monotone transformations and limit laws.* 1973.

46 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 1: the language.* 1973.

47 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 2: the compiler.* 1973.

48 F.E.J. Kruseman Aretz, P.J.W. ten Hagen, H.L. Oudshoorn. *An ALGOL 60 compiler in ALGOL 60, text of the MC-compiler for the EL-X8.* 1973.

49 H. Kok. *Connected orderable spaces.* 1974.

50 A. van Wijngaarden, B.J. Mailloux, J.E.L. Peck, C.H.A. Koster, M. Sintzoff, C.H. Lindsey, L.G.L.T. Meertens, R.G. Fisker (eds.). *Revised report on the algorithmic language ALGOL 68.* 1976.

51 A. Hordijk. *Dynamic programming and Markov potential theory.* 1974.

52 P.C. Baayen (ed.). *Topological structures.* 1974.

53 M.J. Faber. *Metrizability in generalized ordered spaces.* 1974.

54 H.A. Lauwerier. *Asymptotic analysis, part 1.* 1974.

55 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 1: theory of designs, finite geometry and coding theory.* 1974.

56 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 2: graph theory, foundations, partitions and combinatorial geometry.* 1974.

57 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 3: combinatorial group theory.* 1974.

58 W. Albers. *Asymptotic expansions and the deficiency concept in statistics.* 1975.

59 J.L. Mijnheer. *Sample path properties of stable processes.* 1975.

60 F. Göbel. *Queueing models involving buffers.* 1975.

63 J.W. de Bakker (ed.). *Foundations of computer science.* 1975.

64 W.J. de Schipper. *Symmetric closed categories.* 1975.

65 J. de Vries. *Topological transformation groups, 1: a categorical approach.* 1975.

66 H.G.J. Pijls. *Logically convex algebras in spectral theory and eigenfunction expansions.* 1976.

68 P.P.N. de Groen. *Singularly perturbed differential operators of second order.* 1976.

69 J.K. Lenstra. *Sequencing by enumerative methods.* 1977.

70 W.P. de Roever, Jr. *Recursive program schemes: semantics and proof theory.* 1976.

71 J.A.E.E. van Nunen. *Contracting Markov decision processes.* 1976.

72 J.K.M. Jansen. *Simple periodic and non-periodic Lamé functions and their applications in the theory of conical waveguides.* 1977.

73 D.M.R. Leivant. *Absoluteness of intuitionistic logic.* 1979.

74 H.J.J. te Riele. *A theoretical and computational study of generalized aliquot sequences.* 1976.

75 A.E. Brouwer. *Treelike spaces and related connected topological spaces.* 1977.

76 M. Rem. *Associons and the closure statement.* 1976.

77 W.C.M. Kallenberg. *Asymptotic optimality of likelihood ratio tests in exponential families.* 1978.

78 E. de Jonge, A.C.M. van Rooij. *Introduction to Riesz spaces.* 1977.

79 M.C.A. van Zuijlen. *Emperical distributions and rank statistics.* 1977.

80 P.W. Hemker. *A numerical study of stiff two-point boundary problems.* 1977.

81 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 1.* 1976.

82 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 2.* 1976.

83 L.S. van Benthem Jutting. *Checking Landau's "Grundlagen" in the AUTOMATH system.* 1979.

84 H.L.L. Busard. *The translation of the elements of Euclid from the Arabic into Latin by Hermann of Carinthia (?), books vii-xii.* 1977.

85 J. van Mill. *Supercompactness and Wallman spaces.* 1977.

86 S.G. van der Meulen, M. Veldhorst. *Torrix I, a programming system for operations on vectors and matrices over arbitrary fields and of variable size.* 1978.

88 A. Schrijver. *Matroids and linking systems.* 1977.

89 J.W. de Roever. *Complex Fourier transformation and analytic functionals with unbounded carriers.* 1978.

90 L.P.J. Groenewegen. *Characterization of optimal strategies in dynamic games.* 1981.

91 J.M. Geysel. *Transcendence in fields of positive characteristic.* 1979.

92 P.J. Weeda. *Finite generalized Markov programming.* 1979.

93 H.C. Tijms, J. Wessels (eds.). *Markov decision theory.* 1977.

94 A. Bijlsma. *Simultaneous approximations in transcendental number theory.* 1978.

95 K.M. van Hee. *Bayesian control of Markov chains.* 1978.

96 P.M.B. Vitányi. *Lindenmayer systems: structure, languages, and growth functions.* 1980.

97 A. Federgruen. *Markovian control problems; functional equations and algorithms.* 1984.

98 R. Geel. *Singular perturbations of hyperbolic type.* 1978.

99 J.K. Lenstra, A.H.G. Rinnooy Kan, P. van Emde Boas (eds.). *Interfaces between computer science and operations research.* 1978.

100 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 1.* 1979.

101 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 2.* 1979.

102 D. van Dulst. *Reflexive and superreflexive Banach spaces.* 1978.

103 K. van Harn. *Classifying infinitely divisible distributions by functional equations.* 1978.

104 J.M. van Wouwe. *Go-spaces and generalizations of metrizability.* 1979.

105 R. Helmers. *Edgeworth expansions for linear combinations of order statistics.* 1982.

106 A. Schrijver (ed.). *Packing and covering in combinatorics.* 1979.

107 C. den Heijer. *The numerical solution of nonlinear operator equations by imbedding methods.* 1979.

108 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 1.* 1979.

109 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 2.* 1979.

110 J.C. van Vliet. *ALGOL 68 transput, part I: historical review and discussion of the implementation model.* 1979.

111 J.C. van Vliet. *ALGOL 68 transput, part II: an implementation model.* 1979.

112 H.C.P. Berbee. *Random walks with stationary increments and renewal theory.* 1979.

113 T.A.B. Snijders. *Asymptotic optimality theory for testing problems with restricted alternatives.* 1979.

114 A.J.E.M. Janssen. *Application of the Wigner distribution to harmonic analysis of generalized stochastic processes.* 1979.

115 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 1.* 1979.

116 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 2.* 1979.

117 P.J.M. Kallenberg. *Branching processes with continuous state space.* 1979.

118 P. Groeneboom. *Large deviations and asymptotic efficiencies.* 1980.

119 F.J. Peters. *Sparse matrices and substructures, with a novel implementation of finite element algorithms.* 1980.

120 W.P.M. de Ruyter. *On the asymptotic analysis of large-scale ocean circulation.* 1980.

121 W.H. Haemers. *Eigenvalue techniques in design and graph theory.* 1980.

122 J.C.P. Bus. *Numerical solution of systems of nonlinear equations.* 1980.

123 I. Yuhász. *Cardinal functions in topology - ten years later.* 1980.

124 R.D. Gill. *Censoring and stochastic integrals.* 1980.

125 R. Eising. *2-D systems, an algebraic approach.* 1980.

126 G. van der Hoek. *Reduction methods in nonlinear programming.* 1980.

127 J.W. Klop. *Combinatory reduction systems.* 1980.

128 A.J.J. Talman. *Variable dimension fixed point algorithms and triangulations.* 1980.

129 G. van der Laan. *Simplicial fixed point algorithms.* 1980.

130 P.J.W. ten Hagen, T. Hagen, P. Klint, H. Noot, H.J. Sint, A.H. Veen. *ILP: intermediate language for pictures.* 1980.

131 R.J.R. Back. *Correctness preserving program refinements: proof theory and applications.* 1980.

132 H.M. Mulder. *The interval function of a graph.* 1980.

133 C.A.J. Klaassen. *Statistical performance of location estimators.* 1981.

134 J.C. van Vliet, H. Wupper (eds.). *Proceedings international conference on ALGOL 68.* 1981.

135 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part I.* 1981.

136 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part II.* 1981.

137 J. Telgen. *Redundancy and linear programs.* 1981.

138 H.A. Lauwerier. *Mathematical models of epidemics.* 1981.

139 J. van der Wal. *Stochastic dynamic programming, successive approximations and nearly optimal strategies for Markov decision processes and Markov games.* 1981.

140 J.H. van Geldrop. *A mathematical theory of pure exchange economies without the no-critical-point hypothesis.* 1981.

141 G.E. Welters. *Abel-Jacobi isogenies for certain types of Fano threefolds.* 1981.

142 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 1.* 1981.

143 J.M. Schumacher. *Dynamic feedback in finite- and infinite-dimensional linear systems.* 1981.

144 P. Eijgenraam. *The solution of initial value problems using interval arithmetic; formulation and analysis of an algorithm.* 1981.

145 A.J. Brentjes. *Multi-dimensional continued fraction algorithms.* 1981.

146 C.V.M. van der Mee. *Semigroup and factorization methods in transport theory.* 1981.

147 H.H. Tigelaar. *Identification and informative sample size.* 1982.

148 L.C.M. Kallenberg. *Linear programming and finite Markovian control problems.* 1983.

149 C.B. Huijsmans, M.A. Kaashoek, W.A.J. Luxemburg, W.K. Vietsch (eds.). *From A to Z, proceedings of a symposium in honour of A.C. Zaanen.* 1982.

150 M. Veldhorst. *An analysis of sparse matrix storage schemes.* 1982.

151 R.J.M.M. Does. *Higher order asymptotics for simple linear rank statistics.* 1982.

152 G.F. van der Hoeven. *Projections of lawless sequences.* 1982.

153 J.P.C. Blanc. *Application of the theory of boundary value problems in the analysis of a queueing model with paired services.* 1982.

154 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part I.* 1982.

155 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part II.* 1982.

156 P.M.G. Apers. *Query processing and data allocation in distributed database systems.* 1983.

157 H.A.W.M. Kneppers. *The covariant classification of two-dimensional smooth commutative formal groups over an algebraically closed field of positive characteristic.* 1983.

158 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 1.* 1983.

159 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 2.* 1983.

160 A. Rezus. *Abstract AUTOMATH.* 1983.

161 G.F. Helminck. *Eisenstein series on the metaplectic group, an algebraic approach.* 1983.

162 J.J. Dik. *Tests for preference.* 1983.

163 H. Schippers. *Multiple grid methods for equations of the second kind with applications in fluid mechanics.* 1983.

164 F.A. van der Duyn Schouten. *Markov decision processes with continuous time parameter.* 1983.

165 P.C.T. van der Hoeven. *On point processes.* 1983.

166 H.B.M. Jonkers. *Abstraction, specification and implementation techniques, with an application to garbage collection.* 1983.

167 W.H.M. Zijm. *Nonnegative matrices in dynamic programming.* 1983.

168 J.H. Evertse. *Upper bounds for the numbers of solutions of diophantine equations.* 1983.

169 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 2.* 1983.

CWI TRACTS

1 D.H.J. Epema. *Surfaces with canonical hyperplane sections.* 1984.

2 J.J. Dijkstra. *Fake topological Hilbert spaces and characterizations of dimension in terms of negligibility.* 1984.

3 A.J. van der Schaft. *System theoretic descriptions of physical systems.* 1984.

4 J. Koene. *Minimal cost flow in processing networks, a primal approach.* 1984.

5 B. Hoogenboom. *Intertwining functions on compact Lie groups.* 1984.

6 A.P.W. Böhm. *Dataflow computation.* 1984.

7 A. Blokhuis. *Few-distance sets.* 1984.

8 M.H. van Hoorn. *Algorithms and approximations for queueing systems.* 1984.

9 C.P.J. Koymans. *Models of the lambda calculus.* 1984.

10 C.G. van der Laan, N.M. Temme. *Calculation of special functions: the gamma function, the exponential integrals and error-like functions.* 1984.

11 N.M. van Dijk. *Controlled Markov processes; time-discretization.* 1984.

12 W.H. Hundsdorfer. *The numerical solution of nonlinear stiff initial value problems: an analysis of one step methods.* 1985.

13 D. Grune. *On the design of ALEPH.* 1985.

14 J.G.F. Thiemann. *Analytic spaces and dynamic programming: a measure theoretic approach.* 1985.

15 F.J. van der Linden. *Euclidean rings with two infinite primes.* 1985.

16 R.J.P. Groothuizen. *Mixed elliptic-hyperbolic partial differential operators: a case-study in Fourier integral operators.* 1985.

17 H.M.M. ten Eikelder. *Symmetries for dynamical and Hamiltonian systems.* 1985.

18 A.D.M. Kester. *Some large deviation results in statistics.* 1985.

19 T.M.V. Janssen. *Foundations and applications of Montague grammer, part 1: Philosophy, framework, computer science.* 1985.

20 B.F. Schriever. *Order dependence.* 1986.

21 D.P. van der Vecht. *Inequalities for stopped Brownian motion.* 1986.