MONOTONE CORRELATION

AND

MONOTONE DISJUNCT PIECES

by

Devendra Chhetry [1]

Jan de Leeuw [2]

Allan R. Sampson [1,3]

SERIES IN RELIABILITY AND STATISTICS

Department of Mathematics and Statistics
University of Pittsburgh
Pittsburgh, PA 15260

Technical Report No. 86-04

June 1986

1 Research sponsored by the Air Force Office of Scientific Research under Contract AFOSR-84-0113. The United States Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation herein.

2 The address for this author is Department of Data Theory, University of Leiden, Middelstregracht 4, 2312 TW Leiden, The Netherlands.

3 The research of this author was done in part at the Department of Statistics, Carnegie-Mellon University.

MONOTONE CORRELATION

AND

MONOTONE DISJUNCT PIECES

by

Devendra Chhetry [1]

Jan de Leeuw [2]

Allan R. Sampson [1,3]

Abstract

Suppose X,Y are random variables taking values on the lattice $\{x_1 <\ldots< x_m\} \times \{y_1 <\ldots< y_n\}$ with $Q = \{\text{Prob}(X=x_i, Y=y_j)\}$. Let $\rho_{CMC}(Q)$ and $\rho_{DMC}(Q)$ be the concordant and discordant monotone correlations defined, respectively, by the maximum and minimum of $\{\rho(f(X),g(Y))\}$ over all f,g increasing with nonzero variances. A number of results concerning $\rho_{CMC}(Q)$ and $\rho_{DMC}(Q)$ and their evaluations are obtained. One result shows that $\rho_{CMC}(Q) = 1$, if and only if Q consists of at least two increasing disjunct pieces, i.e., $Q = \text{Diag}(Q_1,Q_2)$. Necessary and sufficient conditions are also given for $\rho_{CMC}(Q) = \rho_{DMC}(Q)$.

AMS Subject Classification Primary: 15A51 Secondary: 62H20

Keywords and Phrases: Maximal correlation, concordant monotone correlation, disjunct pieces, monotone disjunct pieces.

1 Research sponsored by the Air Force Office of Scientific Research under Contract AFOSR-84-0113. The United States Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation herein.

2 The address for this author is Department of Data Theory, University of Leiden, Middelstegracht 4, 2312 TW Leiden, The Netherlands.

3 The research of this author was done in part at the Department of Statistics, Carnegie-Mellon University.

MONOTONE CORRELATION

AND

MONOTONE DISJUNCT PIECES

by

Devendra Chhetry

Jan de Leeuw

Allan R. Sampson

Department of Mathematics and Statistics
University of Pittsburgh
Pittsburgh, PA 15260

Department of Data Theory
University of Leiden
The Netherlands

Department of Mathematics and Statistics
University of Pittsburgh
Pittsburgh, PA 15260

1. Introduction.

Let X and Y be two discrete random variables taking values in the lattice $S \times T \equiv \{x_1 < ... < x_m\} \times \{y_1 < ... < y_n\}$ with $Q \equiv \{q_{ij}\} = \{\text{Prob}\ (X = x_i, Y = y_j)\}$, where we assume $r_i \equiv \sum_j q_{ij} > 0$ for all i and $c_j \equiv \sum_i q_{ij} > 0$ for all j. The corresponding cdf matrix is defined by $F \equiv \{F_{ij}\} = \{\text{Prob}(X \leq x_i, Y \leq y_j)\}$. There is a substantial literature dealing with measuring the association between X and Y (see Goodman and Kruskal (1979), Haberman (1982) or Raveh (1986)). One such measure of association introduced by Hirschfeld (1935) receiving much attention is the maximal correlation coefficient $\rho'(X,Y)$ (or $\rho'(Q)$) defined to be the max $\{\rho\ (f(X),g(Y))\}$, where the correlation function ρ is maximized over all f and g with nonzero variances. Clearly, $0 \leq \rho'\ (X,Y) \leq 1$. The properties of $\rho'\ (X,Y)$ have been extensively studied, e.g., Richter (1949), Renyi (1959), Lancaster

(1969), Sarmanov (1958a), (1958b), and Hall (1969). One of the interesting and important results is that $\rho'(X,Y) = 0$ is equivalent to X and Y being independent random variables, and $\rho'(X,Y) = 1$ is equivalent to Q consisting of at least two disjunct pieces, where this concept is defined as follows:

Definition 1.1 (Richter (1949). The probability matrix Q is said to consist of k *disjunct pieces* if there exists partitions $S_1, \ldots, S_k$ of S and $T_1, \ldots, T_k$ of T such that

$$\text{Prob}\,((X,Y) \varepsilon S_i \times T_i) > 0,\ i = 1,\ldots,k\ , \tag{1.1}$$

and

$$\text{Prob}\,((X,Y) \varepsilon S_i \times T_j) = 0 \quad \text{for all } i \neq j\ . \tag{1.2}$$

Additionally, the probability matrix Q is said to consist of *exactly* k disjunct pieces, if (1.1) and (1.2) hold, and Q cannot further consist of k+1 disjunct pieces. (Note that this is equivalent to Q having the first k-1 canonical correlations being one, and the kth being less than one, e.g., Richter (1949).)

Fundamental to the concept of disjunctness and the related concept of canonical correlations (see Lancaster (1969)) is the following result. If $m = n$ and Q consists of m disjunct pieces, then X and Y are called mutually completely dependent (Lancaster (1963)), and there exists a one-to-one function h such that $Y = h(X)$ w.p.1..

For the purposes of this paper we require a further refinement of the concept of disjunct pieces. To define this refinement, we employ the notation that if U,V are sets of real numbers, $U < V$ means $u < v$ for all $u \varepsilon U$ and all $v \varepsilon V$.

Definition 1.2. The probability matrix Q is said to consist of k *increasing (decreasing) disjunct pieces* if there exists partitions $S_1 < S_2 < \cdots < S_k$ of S and $T_1 < (>)\ T_2 <(>) \cdots <(>)\ T_k$ of T such that (1.1) and (1.2) hold.

We say Q consists of k *monotone disjunct pieces* if Q consists of either k increasing or decreasing disjunct pieces.

Q consisting of k increasing disjunct pieces, is equivalent to $Q = \text{Diag}(Q_1, \ldots, Q_k)$, where Q_i is an $m_i \times n_i$ matrix and $\Sigma m_i = m$, $\Sigma n_i = n$. This also can be viewed as Q being the direct sum $Q_1 \oplus \ldots \oplus Q_k$, where direct sum in this context is analogous to the direct sum of square matrices (see MacDuffe (1943, p. 114)). Moreover, if m=n and Q consists of m increasing (decreasing) disjunct pieces then X and Y are increasing (decreasing) dependent and X,Y have the upper (lower) Fréchet distribution (see Kimeldorf and Sampson (1978)).

In order to measure positive association between arbitrary random variables X and Y and also to circumvent some of the difficulties pointed out by Kimeldorf and Sampson (1978), Kimeldorf, May and Sampson (KMS) (1982) introduced the concordant monotone correlation ρ_{CMC} (or alternatively $\rho_{CMC}(Q)$), defined by

$$\rho_{CMC} = \sup \{\rho(f(X), g(Y))\} \tag{1.3}$$

where the supremum is taken over all increasing f and g with nonzero variances. Also introduced by KMS is the discordant monotone correlation, $\rho_{DMC}(Q)$ defined by (1.3) where "sup" is replaced by "inf". KMS shows that $-1 \le \rho_{DMC} \le \rho_{CMC} \le 1$, and $\rho_{DMC} = \rho_{CMC} = 0$ is equivalent to X and Y being independent random variables. Also they provide an example where $\rho_{DMC} < \rho_{CMC} = 0$ and yet X and Y are dependent random variables. It is also direct to show that $\rho_{DMC} \ge 0$ ($\rho_{CMC} \le 0$) if and only if X and Y are positively (negatively) quadrant dependent (Lehmann (1966)), i.e., Prob $(X \le x, Y \le y) \ge (\le)$ Prob $(X \le x)$ Prob $(Y \le y)$ for all x,y.

The purpose of this paper is to obtain some important additional results in the bivariate discrete setting concerning ρ_{CMC} and ρ_{DMC}, and their evaluation.

2. Some results for ρ_{CMC}.

For a given probability matrix Q the notation $\rho_Q(\underline{x}, \underline{y})$ is used for the correlation $(\underline{x}'(D_r - \underline{r}\,\underline{r}')\underline{x})^{-1/2} (\underline{y}'(D_c - \underline{c}\,\underline{c}')\underline{y})^{-1/2} (\underline{x}'(Q - \underline{r}\,\underline{c}')\underline{y})$, where $D_r = \text{Diag}(r_1, \ldots, r_m)$, D_c

$= \mathrm{Diag}\,(c_1, \ldots, c_n)$, $\underline{r} = (r_1, \ldots, r_m)'$, $\underline{c} = (c_1, \ldots, c_n)'$ and the denominator is nonzero. Furthermore, the vector $(w_1, \ldots, w_p)'$ is said to be nondecreasing if $w_1 \le \cdots \le w_p$; and $\underline{e}_k$ denotes the kth co-ordinate unit vector of the appropriate dimension.

Theorem 2.1. A necessary and sufficient condition for $\rho_{CMC}(Q) = 1$ ($\rho_{DMC}(Q) = -1$) is that Q consists of at least two increasing (decreasing) disjunct pieces.

Proof. The sufficiency follows immediately (see Kimeldorf, May and Sampson (1982, p. 120)).

Suppose $\rho_{CMC}(Q) = 1$. Then, there exist two nondecreasing vectors $\underline{x}_0$ and $\underline{y}_0$, such that $\rho_Q(\underline{x}_0, \underline{y}_0) = 1$ and thus, Q consists of at least two disjunct pieces. Assume that Q consists of exactly t disjunct pieces, where $t \ge 2$. Hence, there exist permutation matrices P_1 and P_2 such that $Q^* = P_1\, Q\, P_2'$ consists of exactly t increasing disjunct pieces, i.e., $Q^* = \mathrm{Diag}\,(Q_1^*, \ldots, Q_t^*)$, where Q_k^* is an $m_k x n_k$ matrix, such that $\Sigma m_k = m$ and $\Sigma n_k = n$. It then follows (see Richter (1949) or Bastin, Benzercri, Bourgarit and Cazes (1980)) that $\rho_{Q^*}(\underline{x}_0^*, \underline{y}_0^*) = 1$ if and only if $\underline{x}_0^* = \sum_{s=1}^{t} \lambda_s\, \underline{u}_s$, where

$\underline{u}_s = \underline{e}_{m_1 + \cdots + m_{s-1} + 1} + \cdots + \underline{e}_{m_1 + \cdots + m_s}$, and $\underline{y}_0^* = \sum_{s=1}^{t} (\alpha\lambda_s + \beta)\, \underline{v}_s$, where

$\underline{v}_s = \underline{e}_{n_1 + \cdots + n_{s-1} + 1} + \cdots + \underline{e}_{n_1 + \cdots + n_s}$, and where there exists $i < j$ such that $\lambda_i \ne \lambda_j$ and $\alpha > 0$. It is direct to show that $\rho_Q(\underline{x}_0, \underline{y}_0) = 1$ if and only if $\underline{x}_0 = P_1'\, \underline{x}_0^*$ and $\underline{y}_0 = P_2'\, \underline{y}_0^*$ for any $\underline{x}_0^*, \underline{y}_0^*$ which satisfies $\rho_{Q^*}(\underline{x}_0^*, \underline{y}_0^*) = 1$. For each vector $\underline{x}_0^*, \underline{y}_0^*$ of the preceding form, let $i^* \ge 2$ be the first value such that $\lambda_{i^*} \ne \lambda_1$; the existence of i^* follows from $\lambda_i \ne \lambda_j$ for some $i < j$. Because $\underline{x}_0$ is nondecreasing and $\underline{x}_0 = P_1'\, \underline{x}_0^*$, it follows that $P_1 = \mathrm{Diag}\,(P_1^{(1)}, P_1^{(2)})$, where $P_1^{(1)}$ is an $m^* \times m^*$ permutation matrix and $P_1^{(2)}$ is an $(m - m^*) \times (m - m^*)$ permutation matrix, where $m^* = \sum_{k=1}^{i^*-1} m_k$. Similarly, P_2 is in block diagonal form and, hence, Q consists of at least two increasing disjunct pieces.

Now suppose $\rho_{DMC}(Q) = -1$. Use the preceding argument and the fact that

$\rho_{DMC}(Q) = -\rho_{CMC}(Q^*)$ where $Q^* = Q(\underline{e}_n, \ldots, \underline{e}_1)$ to get the result. ∎

KMS show that monotone correlation $\rho^*(Q)$, introduced by Kimeldorf and Sampson (1978), is also given by $\rho^*(Q) = \max\{\rho_{CMC}(Q), -\rho_{DMC}(Q)\}$. From Theorem 2.1, it immediately follows that $\rho^*(Q) = 1$ if and only if, Q consists of at least two monotone disjunct pieces.

While Theorem 2.1 deals with the case $\rho'(Q) = \rho_{CMC}(Q) = 1$, more generally we have $\rho'(Q) \geq \rho_{CMC}(Q)$. However, in some cases Schriever (1983) shows that $\rho'(Q) = \rho_{CMC}(Q)$ without their necessarily being unity. We observe that $\rho'(Q) = \rho_{CMC}(Q)$ means that there exist at least one pair of nondecreasing function f_0 and g_0 such that $\rho(f_0(X), g_0(Y)) = \rho'(Q)$. For a further discussion of Schriever's results we need the following definition due to Lehmann (1966).

Definition (Lehmann (1966)). A random variable X is said to be *positively regression dependent (PRD)* on Y if $\text{Prob}(X > x \mid Y = y)$ is nondecreasing in y for all x.

Theorem 2.2. (Schriever (1983)). If X is PRD on Y and Y is PRD on X, then $\rho'(Q) = \rho_{CMC}(Q)$.

Chhetry and Sampson (1986) show that the conditions of Theorem 2.2 for $\rho'(Q) = \rho_{CMC}(Q)$ are not necessary. This result is not surprising because Q corresponding to Y is PRD on X (X is PRD on Y) implies that every $\tilde{Q}$ has the same property, where $\tilde{Q}$ is obtained from Q by collapsing any sets of adjacent rows or adjacent columns. As a consequence of this result and of Theorem 2.2, it follows that Q corresponding to Y is PRD on X and X is PRD on Y implies that $\rho'(\tilde{Q}) = \rho_{CMC}(\tilde{Q})$ for every collapsed $\tilde{Q}$.

In the study of bivariate dependence concepts, it oftentimes is of interest to consider $P(\underline{r},\underline{c})$, the class of all $m \times n$ probability matrices with fixed row and column marginals, $\underline{r}$ and $\underline{c}$, respectively. It is well known that (see Schriever (1985, Example 4.2.3)) $\rho_{CMC}(Q^+) \geq \rho_{CMC}(Q)$ for all $Q \,\varepsilon\, P(\underline{r}, \underline{c})$, where Q^+ is the probability matrix corresponding to the upper Fréchet cdf matrix $F^+ = \{(\min(F_i, G_j))\}$, where $F_i = \sum_{k=1}^{i} r_k$ and $G_j = \sum_{k=1}^{j} c_k$. If both marginals were continuous, the

CMC for the upper Frechet bound is 1; however, in the discrete situation it is not always the case that $\rho_{CMC}(Q^+)$ is one. In the following theorem we provide a necessary and sufficient condition for $\rho_{CMC}(Q^+) = 1$ in terms of the marginals.

Theorem 2.3. A necessary and sufficient condition for $\rho_{CMC}(Q^+) = 1$ is that there exist $s < m$ and $t < n$ such that $F_s = G_t$.

Proof: In view of Theorem 2.1, we need to show that $Q^+ = \text{Diag}\ (Q_1^+, Q_2^+)$ if and only if $F_s = G_t$, where Q_1^+ is $s \times t$ and Q_2^+ is $(m-s) \times (n-t)$. Obviously, $Q^+ = \text{Diag}\ (Q_1^+, Q_2^+)$ implies that $F_s = G_t$. To prove the converse assume that $F_s = G_t$. Let F_{ij}^+ be the (i,j) th element of F^+; then it can be easily checked that

$$F_{ij}^+ = \begin{cases} F_i & \text{if } i = 1,2,\ldots,s \text{ and } j \geq t \\ G_j & \text{if } i = s, \text{ and } j < t \\ G_j & \text{if } i > s, j \leq t\,. \end{cases}$$

This implies that Q^+ is of the required form. ∎

To motivate the next theorem, consider first the simple case when Q is a 2×2 probability matrix. Then it is trivial to show that $\rho_{CMC}(Q) = \rho_{DMC}(Q)$; additionally, $\rho_{CMC}(Q) = -1$ ($\rho_{DMC}(Q) = 1$) if and only if $q_{11} = q_{22} = 0$ ($q_{12} = q_{21} = 0$). The analogous results do not continue to hold when $m > 2$ or $n > 2$, as we now show.

Theorem 2.4. If $m > 2$ or $n > 2$, then $\rho_{CMC}(Q) = \rho_{DMC}(Q)$ if and only if X and Y are independent.

Proof: Suppose $\rho_{CMC}(Q) = \rho_{DMC}(Q) = \eta \neq 0$ (if $\eta = 0$, independence follows). Without loss of generality assume $m > 2$, so that we can choose three nondecreasing functions a_1, a_2, and b such that (i) $\rho(a_1(X), a_2(X)) < 1$ and (ii) $\text{Var}\ [a_1(X)] = \text{Var}\ [a_2(X)] = \text{Var}\ [b(Y)] = 1$. Then, by assumption,

$$\eta = \rho(a_1(X) + a_2(X), b(Y)) = 2\eta(2 + 2\ \rho(a_1(X), a_2(X)))^{-1/2}\ ,$$

which implies that $\rho(a_1(X), a_2(X)) = 1$, a contradiction. ∎

Corollary 2.5: If $m > 2$ or $n > 2$, then $\rho_{CMC}(Q) > -1$ and $\rho_{DMC}(Q) < 1$.

The proof of Corollary 2.5 is obvious.

3. Some results concerning evaluation.

While the quantities $\rho'(Q)$ and $\rho_{CMC}(Q)$ are of interest in their own right as measures of association, the vectors at which these maxima occur play an important role in scaling. The vectors which maximize $\rho'(Q)$ can be derived from certain correspondence analysis (Benzecri (1973)) and Hill (1974)). The increasing vectors which yield $\rho_{CMC}(Q)$ can be interpreted as either providing dual scales for ordinal contingency tables or a form of ordinal correspondence analysis. However, their evaluation is substantially more complicated than the non-ordinal case, e.g., see KMS, or Breiman and Friedman (1985), and the comments of Buja and Kass (1985). Chhetry and Sampson (CS) (1986) provide an approach which simplifies somewhat the calculation of $\rho_{CMC}(Q)$ and the maximizing vectors. We briefly discuss that approach and then detail how to employ it effectively when the ordinal table is collapsed. The latter issue is important in building heirarchial models for ordinal tables in which collapsing is used for model simplification.

For every $m \times n$ probability matrix Q, CS define the $(m + n - 2) \times (m + n - 2)$ matrix $\Sigma(Q)$ (denoted where there is no ambiguity as Σ) by

$$\Sigma(Q) = \begin{bmatrix} \overline{A}' & 0 \\ 0 & \overline{B}' \end{bmatrix} \begin{bmatrix} D_r & Q \\ Q' & D_c \end{bmatrix} \begin{bmatrix} \overline{A} & 0 \\ 0 & \overline{B} \end{bmatrix}, \tag{3.1}$$

where $\overline{A} = (I_m - \underline{1}_m \underline{1}_m' D_r)\Psi_m$, $\overline{B} = (I_n - \underline{1}_n \underline{1}_n' D_c)\Psi_n$, and Ψ_p is the $p \times (p-1)$ matrix whose (i,j)th element is 0, if $i \leq j$, and 1, otherwise. Let $\Sigma_{11} = \overline{A}' D_r \overline{A}$, $\Sigma_{12} = \overline{A}' Q \overline{B}$, $\Sigma_{22} = \overline{B}' D_c \overline{B}$ and $\Sigma_{21} = \Sigma_{12}'$. CS also show that Σ_{11} and Σ_{22} are positive definite and Σ is a nonnegative definite matrix. For any Q, let Σ be given by (3.1) and define for $\underline{\alpha} \in R^{m-1}$, $\underline{\beta} \in R^{n-1}$

$$r_Q(\underline{\alpha}, \underline{\beta}) = (\underline{\alpha}' \Sigma_{11} \underline{\alpha})^{-1/2} (\underline{\alpha}' \Sigma_{12} \underline{\beta}) (\underline{\beta}' \Sigma_{22} \underline{\beta})^{-1/2} \tag{3.2}$$

where $\underline{\alpha} \neq 0$ and $\underline{\beta} \neq 0$. Then CS show that

$$\rho'(Q) = \max_{\underline{\alpha}, \underline{\beta}} r_Q(\underline{\alpha}, \underline{\beta}) , \tag{3.3a}$$

$$\rho_{CMC}(Q) = \max_{\underline{\alpha} \geq 0, \underline{\beta} \geq 0} r_Q(\underline{\alpha}, \underline{\beta}) , \tag{3.3b}$$

and

$$\rho_{DMC}(Q) = \min_{\underline{\alpha} \geq 0, \underline{\beta} \geq 0} r_Q(\underline{\alpha}, \underline{\beta}) . \tag{3.3c}$$

The relationships of (3.3) can be viewed as simplifying computation by reducing dimensionality. Also note that if $\underline{\alpha}_0$ and $\underline{\beta}_0$ optimize any of (3.3a), (3.3b) or (3.3c), then the corresponding maximizing vectors $\underline{x}_0$ and $\underline{y}_0$ defining the left hand sides are

related by $\underline{x}_0 = \overline{A}\, \underline{\alpha}_0$ and $\underline{y}_0 = \overline{B}\, \underline{\beta}_0$. For example, if $r_Q(\underline{\alpha}, \underline{\beta})$ is maximized at $\underline{\alpha}_0, \underline{\beta}_0$, then $\rho_Q(\underline{x}, \underline{y})$ is maximized at $\underline{x}_0 = \overline{A}\, \underline{\alpha}_0$ and $\underline{y}_0 = \overline{B}\, \underline{\beta}_0$.

An additonal advantage of the problem formulation given by (3.2) and (3.3) is that these optimization problems can be reformulated analogously to the problem of finding the canonical correlation for the multivariate normal. A description of this relationship is given in the following lemma whose proof follows from Lemma 4.1 and Theorem 4.2 of CS.

Lemma 3.1. The positive square root of the largest eigenvalue ρ_1^2 of $\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ (or $\Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$) is $\rho'(Q)$. If $\underline{\alpha}^{(1)} \neq 0$ and $\underline{\beta}^{(1)} \neq 0$ satisfy the equations:

$$\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \underline{\alpha}^{(1)} = \rho_1^2 \underline{\alpha}^{(1)} \tag{3.4a}$$

and

$$\underline{\beta}^{(1)} = \Sigma_{22}^{-1} \Sigma_{21} \underline{\alpha}^{(1)} , \tag{3.4b}$$

then $\rho_Q(\underline{\alpha}^{(1)}, \underline{\beta}^{(1)}) = \rho'(Q)$. Moreover, $\rho'(Q) = \rho_{CMC}(Q)$ if and only if there exist nonnegative vectors $\underline{\alpha}^{(1)}$ and $\underline{\beta}^{(1)}$ satisfying (3.4).

We now relate the computation of the maximal correlation and the monotone correlations for collapsed contingency tables to the original uncollapsed tables. A recent discussion on the general issue of collapsing non-ordinal contingency tables is given by Gilula and Krieger (1983)and Gilula (1986). The following definition is useful in our discussion.

Definition 3.2. An $m \times n$ matrix $P = \{p_{ij}\}$, $m \le n$, is said to be a C-matrix if (a) the rank of P is m, (b) each column of P has one and only one nonzero element, and the nonzero element is unity, and (c) if $p_{ij} = p_{ik} = 1$ for $k > j$ implies $p_{ie} = 1$ for all $e = j+1,\ldots,k-1$.

Obviously, in the above definition, if m=n then P is a permutation matrix; and if $m < n$ then appropriate multiplication of a probability matrix by P collapses sets of adjacent rows or columns. Suppose Q is transformed to $\tilde{Q}$ by $\tilde{Q} = P_1 Q P_2'$, where P_1 and P_2 are, respectively $s \times m$ and $t \times n$ C-matrices. Then, $\tilde{Q}$ is an $s \times t$ probability matrix with row and column marginals $\underline{\tilde{r}} = P_1 \underline{r} = (\tilde{r}_1, \ldots, \tilde{r}_s)'$ and $\underline{\tilde{c}} = P_2 \underline{c} = (\tilde{c}_1, \ldots, \tilde{c}_t)'$, respectively. Moreover, if $D_{\tilde{r}} = \text{Diag}(\tilde{r}_1, \ldots, \tilde{r}_s)$ and $D_{\tilde{c}} = \text{Diag}(\tilde{c}_1, \ldots, \tilde{c}_t)$, then $D_{\tilde{r}} = P_1 D_r P_1'$ and $D_{\tilde{c}} = P_2 D_c P_2'$.

In the following theorem, we establish the relationship between $\Sigma(Q)$ and $\Sigma(\tilde{Q})$.

Theorem 3.3. If $\tilde{Q} = P_1 Q P_2'$, where P_1 and P_2 are, respectively, $s \times m$ and $t \times n$ C-matrices, then

$$\Sigma(\tilde{Q}) = \text{Diag}(K_m', K_n')\, \Sigma(Q)\, \text{Diag}(K_m, K_n),$$

where $K_m = \Delta_m' P_1' \Psi_s$, $K_n = \Delta_n' P_2' \Psi_t$, and Δ_p is the $p \times (p-1)$ matrix $(\underline{e}_2 - \underline{e}_1, \underline{e}_3 - \underline{e}_2, \ldots, \underline{e}_p - \underline{e}_{p-1})$.

Proof: From CS (Lemma 3.2(i))

$$\Sigma_{12}(\tilde{Q}) = \Psi_s' (\tilde{Q} - \underline{\tilde{r}}\, \underline{\tilde{c}}') \Psi_t$$

$$= \Psi_s' P_1 (Q - \underline{r}\, \underline{c}') P_2' \Psi_t .$$

From the quadrant dependence decomposition (CS (Equation (3.4))), we obtain

$$\Sigma_{12}(\tilde{Q}) = \Psi_s' P_1 \Delta_m \Sigma_{12}(Q) \Delta_n' P_2' \Psi_t$$

$$= K_m' \Sigma_{12}(Q) K_n .$$

The relationships concerning $\Sigma_{11}(\tilde{Q})$ and $\Sigma_{22}(\tilde{Q})$ are established similarly. ∎

Note that the results of Theorem 3.3 also hold if P_1 and P_2 are more general in that they collapse nonadjacent rows and columns; however, such matrices would not be meaningful for ordinal tables. The usefulness of Theorem 3.3 especially when used in conjunction with Lemma 3.1 can be seen in the following example.

Example 3.4. Let P_1 and P_2 be C- matrices of orders $(m-s) \times m$ and $(n-t) \times n$, respectively, where $P_1 \equiv (\underline{e}_1, \ldots, \underline{e}_1, \underline{e}_2, \ldots, \underline{e}_{m-s})$ and $P_2 \equiv (\underline{e}_1, \ldots, \underline{e}_1, \underline{e}_2, \ldots, \underline{e}_{n-t})$. Then, the matrices K_m and K_n defined in Theorem 3.3 reduce to the form:

$$K_m' = (0_1, I_{(m-s-1)}) \text{ and } K_n' = (0_2, I_{(n-s-1)}) \tag{3.5}$$

where 0_1 and 0_2 are zero matrices of orders $(m-s-1) \times s$ and $(n-t-1) \times t$, respectively. Hence, using (3.5) in Theorem 3.3, we obtain

$$\Sigma_{12}(\tilde{Q}) = \Sigma_{12} [1,2,\ldots,s; 1,2,\ldots,t] ,$$

$$\Sigma_{11}(\tilde{Q}) = \Sigma_{11} [1,2,\ldots,s; 1,2,\ldots,s] ,$$

and

$$\Sigma_{22}(\tilde{Q}) = \Sigma_{22} [1,2,\ldots,t; 1,2,\ldots,t] ,$$

where $\Sigma_{11} [1,2,\ldots,i; 1,2,\ldots,k]$ is the submatrix obtained from $\Sigma_{11}(Q)$ by deleting the first i rows and the first k columns, etc.

REFERENCES

Bastin, Ch., Benzecri, J. P., Bourgarit, Ch., and Cazes, P. (1980). *Pratique de l'Analyse de Donnees.* Dunod, Paris.

Benzecri, J. P. (1983). *L'Analyse des Donnees II,* Dunod, Paris.

Breiman, L. and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation, *Journal of the American Statistical Association,* 80, 580-598.

Buja, A. and Kass, R. E. (1985). Some observations on ACE methodology. *Journal of the American Statistical Association,* 80, 602-607.

Chhetry, D. and Sampson, A. (1986). A projection decomposition for bivariate discrete probability distribution. Technical Report No. 86-03, Series in Reliability and Statistics, Department of Mathematics and Statistics, University of Pittsburgh.

Gilula, Z. (1986). Grouping and assoication in contingency tables: An exploratory canonical correlation approach. *Journal of the American Statistical Association 81,* To appear.

Gilula, Z. and Krieger, A. M. (1983). The decomposability and monotonicity of Pearson's chi-square for collapsed contingency tables, *Journal of the American Statistical Association 78,* 176-80.

Goodman, L. and Kruskal, W. (1979). *Measures of Association for Cross Classifications.* Springer-Verlag, New York.

Haberman, S. (1982). "Association, measures of" in *Encyclopedia of Statistical Sciences (vol. 1).* Eds. S. Kotz, N. Johnson. Wiley, New York, pp. 130-137.

Hall, W. J. (1969). On characterizing dependence in joint distributions. In *Essays in Probability and Statistics.* Eds. Bose et al. University of North Carolina Press, Chapel Hill.

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

Hirschfeld, H. O. (1935). A connection between correlation and contingency. *Proc Camb. Philos. Soc. 31,* 520-524.

Kimeldorf, G. and Sampson, A. R. (1978). Monotone dependence. *Ann. Statistist 6,* 895-903.

Kimeldorf, G., May, J., and Sampson, A. R. (1982). Concordant and discordant monotone correlations and their evaluation by nonlinear optimization. In *Optimization in Statistics*, Eds. Zanakis and Rustagi, *TIMS Studies Management Sci. 19*, 117-130.

Lancaster, H. O. (1969). *The Chi-squared Distribution*. John Wiley, New York.

Lancaster, H. O. (1958). The structure of bivariate distributions. *Ann. Math. Statist. 29*, 719-736.

Lehmann, E. L. (1966). Some concepts of dependence. *Ann. Math. Stat. 37*, 1137-1153.

MacDuffee, C. (1949). *Vectors and Matrices*. The Mathematical Association of America.

Nishisato, S. (1980). *Analysis of Categorical Data: Dual Scaling and its Applications*. University of Toronto Press.

Raveh, A. (1986). On measures of monotone association. *The American Statistician 40*, 117-123.

Renyi, A. (1959). On measures of dependence. *Acta. Math. Acad. Sci. Hunga. 10*, 441-451.

Richter, H. (1949). Zur Maximal Correlation. *Z. Angew. Math. Mech. 19*, 127-128.

Sarmanov, O. V. (1958a). The maximal correlation coefficient (symmetric case). *Dokl. Akad. Nank. SSSR 120*, 715-718. (English translation in *Sep. Transl. Math. Statist. Probability 4*, 271-275.)

Sarmanov, O. V. (1958b). The maximal correlation coefficient (non-symmetric case). *Dokl. Akad. Nank SSR 121*, 52-55. (English translation in *Sep. Transl. Math. Statist. Probability 4*, 207-210.)

Schriever, B. F. (1985). *Order Dependence*. Ph.D. Dissertation, Free University of Amsterdam.

Schriever, B. F. (1983). Scaling of order dependent categorical variables with correspondence analysis. *International Statistical Review 51*, 225-238.