SIAM J. MATRIX ANAL. APPL.
Vol. 11, No. 3, pp. 361–368, July 1990
© 1990 Society for Industrial and Applied Mathematics
002

MONOTONE CORRELATION AND MONOTONE DISJUNCT PIECES*

DEVENDRA CHHETRY†, JAN DE LEEUW‡, AND ALLAN R. SAMPSON§

Abstract. Suppose X, Y are random variables taking values on the $m \times n$ lattice $\{x_1 < \cdots < x_m\} \times \{y_1 < \cdots < y_n\}$ with $Q = \{\text{Prob}(X = x_i, Y = y_j)\}$. Let $\rho_{\text{CMC}}(Q)$ and $\rho_{\text{DMC}}(Q)$ be the concordant and discordant monotone correlations defined, respectively, by the maximum and minimum of correlation $f(X)$, $g(Y)$ over all f, g increasing with nonzero variances. A number of results concerning $\rho_{\text{CMC}}(Q)$ and $\rho_{\text{DMC}}(Q)$ and their evaluations are obtained. One result shows that $\rho_{\text{CMC}}(Q) = 1$, 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 $\rho_{\text{CMC}}(Q) = \rho_{\text{DMC}}(Q)$.

Key words. maximal correlation, concordant monotone correlation, disjunct pieces, monotone disjunct pieces

AMS(MOS) subject classifications. primary 15A51; secondary 62H20

1. Introduction. Let X and Y be two discrete random variables taking values in the $m \times n$ lattice $S \times T \equiv \{x_1 < \cdots < x_m\} \times \{y_1 < \cdots < y_n\}$ 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. There is a substantial literature in statistics and probability dealing with measuring the association between the random variables X and Y (see Goodman and Kruskal (1979), Haberman (1982) or Raveh (1986)). One such measure of association introduced by Hirschfeld (1935) is the maximal correlation coefficient $\rho'(X, Y)$ (or $\rho'(Q)$) defined to be the max $\{\rho(f(X), g(Y))\}$, where ρ denotes correlation and the maximum is over all f and g with nonzero variances. Clearly, $0 \leqq \rho'(X, Y) \leqq 1$.

The properties of $\rho'(X, Y)$ have been extensively studied (e.g., Richter (1949), Rényi (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 exist partitions $S_1, \cdots, S_k$ of S and $T_1, \cdots, T_k$ of T such that

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

and

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

* Received by the editors March 30, 1988; accepted for publication (in revised form) August 16, 1989.

† Department of Mathematics, Tribhuvan University, Kirtipur Campus, Kathmandu, Nepal. The research of this author was done in part at the Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, and was supported by the Air Force Office of Scientific Research under contract AFOSR-84-0113.

‡ Psychology Department, University of California at Los Angeles, 405 Hilgarde Avenue, Los Angeles, California 90024-1563.

§ Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260. The research of this author was done in part at the Department of Statistics, Carnegie-Mellon University, Pittsburgh, Pennsylvania, and supported by the Air Force Office of Scientific Research under contract AFOSR-84-0113.

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. Richter (1949) has extended this result concerning disjunct pieces utilizing Fisher's canonical decomposition of Q. Define $Q^* \equiv D_r^{-1/2} Q D_c^{-1/2}$, where $D_r = \text{Diag}(r_1, \cdots, r_m)$ and $D_c = \text{Diag}(c_1, \cdots, c_n)$. Then, assuming here for convenience $m \leqq n$, the spectral decomposition of Q^* can be written as $Q^* = \Gamma[\text{Diag}(1, \rho_1, \cdots, \rho_{m-1}) : O_{m,n-m}]G'$, where $\Gamma = [D_r^{1/2}\mathbf{1}_m : \Gamma_1]$ and $G = [D_c^{1/2}\mathbf{1}_n : G_1]$ are orthonormal matrices, $O_{m,n-m}$ is an $m \times (n-m)$ matrix of zeros, and $1 \geqq \rho_1^2 \geqq \cdots \geqq \rho_{m-1}^2 \geqq 0$ are the eigenvalues of $Q^{*\prime}Q^*$. Based on this spectral decomposition, Fisher's (1940) canonical decomposition can be written

$$Q = \mathbf{r}\mathbf{c}' + D_r^{1/2}\Gamma_1 D_\rho (D_c^{1/2} G_1)',$$

where $D_\rho = [\text{Diag}(\rho_1, \cdots, \rho_{m-1}) : O_{m-1,n-m}]$. The values $\rho_1, \cdots, \rho_{m-1}$ are called the canonical correlations of the distribution Q, where it is known that $\rho'(X, Y) = \rho_1$. (See Lancaster (1969, Chap. 6) or Chhetry and Sampson (1987) for further discussions concerning the canonical decomposition and its interpretation.) The result obtained by Richter (1949) is that Q consists of exactly k disjunct pieces if and only if $\rho_1 = \cdots = \rho_{k-1} = 1$ and $\rho_k < 1$.

Another related concept is the following one. If $m = n$ and Q consists of m disjunct pieces, then X and Y are called mutually completely dependent (Lancaster (1969)), and there exists a one-to-one function h such that the random variables X and Y are completely related by $Y = h(X)$.

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 $u \in U$ and all $v \in V$.

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

$$Q = \text{Diag}(Q_1, \cdots, Q_k),$$

where Q_i is an $m_i \times n_i$ matrix and $\sum m_i = m$, $\sum n_i = n$. This also can be viewed as Q being the direct sum $Q_1 \oplus \cdots \oplus Q_k$, when direct sum in this context is analogous to the direct sum of square matrices (see MacDuffee (1949, p. 114)). If $m = n$ and Q consists of m increasing (decreasing) disjunct pieces the notion of X and Y being mutually completely dependent can be refined. In this case X and Y are related by h strictly increasing (decreasing) and the probability matrix corresponds to a special class of probability distributions called the upper (lower) Fréchet bounds (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 $\rho_{\text{CMC}}(Q)$), defined by

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

where the maximum is taken over all increasing f and g with nonzero variances. Also introduced by KMS is the discordant monotone correlation $\rho_{\text{DMC}}(Q)$ defined by (1.3) where "max" is replaced by "min." KMS show that $-1 \leqq \rho_{\text{DMC}} \leqq \rho_{\text{CMC}} \leqq 1$, and

$\rho_{\text{DMC}} = \rho_{\text{CMC}} = 0$ is equivalent to X and Y being independent random variables. Also they provide an example where $\rho_{\text{DMC}} < \rho_{\text{CMC}} = 0$ and yet X and Y are dependent random variables. It is also direct to show that $\rho_{\text{DMC}} \geqq 0$ ($\rho_{\text{CMC}} \leqq 0$) if and only if X and Y are positively (negatively) quadrant dependent (Lehmann (1966)), i.e., Prob $(X \leqq x,\ Y \leqq y) \geqq (\leqq)$ Prob $(X \leqq x)$ Prob $(Y \leqq y)$ for all x, y.

The purpose of this paper is to obtain some 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 used for the correlation between $f(X)$ and $g(Y)$ is

$$\rho_Q(\mathbf{f},\mathbf{g}) = (\mathbf{f}'(D_r - \mathbf{r}\mathbf{r}')\mathbf{f})^{-1/2}(\mathbf{g}'(D_c - \mathbf{c}\mathbf{c}')\mathbf{g})^{-1/2}(\mathbf{f}'(Q - \mathbf{r}\mathbf{c}')\mathbf{g}),$$

where

$$\mathbf{r} = (r_1, \cdots, r_m)', \qquad \mathbf{c} = (c_1, \cdots, c_n)',$$

$$\mathbf{f} = (f(x_1), \cdots, f(x_m))', \qquad \mathbf{g} = (g(y_1), \cdots, g(y_n))',$$

and the denominator is nonzero.

Throughout we say the vector $(w_1, \cdots, w_p)'$ is nondecreasing if $w_1 \leqq \cdots \leqq w_p$; and use $\mathbf{e}_k$ to denote the kth coordinate unit vector of the appropriate dimension. Often we use the simple fact that for every $m \times n$ probability matrix Q, there uniquely corresponds an $m \times n$ cumulative distribution matrix defined by

$$F \equiv \{F_{ij}\} = \{\text{Prob}\,(X \leqq x_i,\ Y \leqq y_j)\},$$

i.e., $F_{ij} = \sum_{k=1}^{i} \sum_{l=1}^{j} q_{kl}$.

THEOREM 2.1. *A necessary and sufficient condition for*

$$\rho_{\text{CMC}}(Q) = 1\ (\rho_{\text{DMC}}(Q) = -1)$$

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)).

To show necessity, suppose $\rho_{\text{CMC}}(Q) = 1$. Then, there exist two nondecreasing vectors $\mathbf{f}_0$ and $\mathbf{g}_0$, such that $\rho_Q(\mathbf{f}_0, \mathbf{g}_0) = 1$ and thus, Q consists of at least two disjunct pieces. Assume that Q consists of exactly t disjunct pieces, where $t \geqq 2$. 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., $Q^* = \text{Diag}\,(Q_1^*, \cdots, Q_t^*)$, where Q_k^* is an $m_k \times n_k$ matrix, such that $\sum m_k = m$ and $\sum n_k = n$. It then follows (see Richter (1949) or Bastin et al. (1980)) that $\rho_{Q^*}(\mathbf{f}_0^*, \mathbf{g}_0^*) = 1$ if and only if $\mathbf{f}_0^* = \sum_{s=1}^{t} \lambda_s \mathbf{u}_s$, where $\mathbf{u}_s = \mathbf{e}_{m_1+\cdots+m_{s-1}+1} + \cdots + \mathbf{e}_{m_1+\cdots+m_s}$, and $\mathbf{g}_0^* = \sum_{s=1}^{t} (\alpha\lambda_s + \beta)\mathbf{v}_s$, where $\mathbf{v}_s = \mathbf{e}_{n_1+\cdots+n_{s-1}+1} + \cdots + \mathbf{e}_{n_1+\cdots+n_s}$, and where there exists $i < j$ such that $\lambda_i \neq \lambda_j$ and $\alpha > 0$. It is direct to show that $\rho_Q(\mathbf{f}_0, \mathbf{g}_0) = 1$ if and only if $\mathbf{f}_0 = P_1'\mathbf{f}_0^*$ and $\mathbf{g}_0 = P_2'\mathbf{g}_0^*$ for any $\mathbf{f}_0^*, \mathbf{g}_0^*$, which satisfies $\rho_{Q^*}(\mathbf{f}_0^*, \mathbf{g}_0^*) = 1$. For each vector $\mathbf{f}_0^*, \mathbf{g}_0^*$ of the preceding form, let $i^* \geqq 2$ be the first value such that $\lambda_{i^*} \neq \lambda_1$; the existence of i^* follows from $\lambda_i \neq \lambda_j$ for some $i < j$. Because $\mathbf{f}_0$ is nondecreasing and $\mathbf{f}_0 = P_1'\mathbf{f}_0^*$, it follows that $P_1 = \text{Diag}\,(P_1^{(1)}, P_1^{(2)})$, 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 $\rho_{\text{DMC}}(Q) = -1$. Use the preceding argument and the fact that $\rho_{\text{DMC}}(Q) = -\rho_{\text{CMC}}(Q^*)$ where $Q^* = Q(\mathbf{e}_n, \cdots, \mathbf{e}_1)$ to get the result. $\square$

KMS show that monotone correlation $\rho^*(Q)$, introduced by Kimeldorf and Sampson (1978), is also given by $\rho^*(Q) = \max\,\{\rho_{\text{CMC}}(Q), -\rho_{\text{DMC}}(Q)\}$. 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 $\rho'(Q) = \rho_{\mathrm{CMC}}(Q) = 1$, more generally we have $\rho'(Q) \geqq \rho_{\mathrm{CMC}}(Q)$. However, in some cases Schriever (1983) shows that $\rho'(Q) = \rho_{\mathrm{CMC}}(Q)$ without their necessarily being unity. We observe that $\rho'(Q) = \rho_{\mathrm{CMC}}(Q)$ means that there exists at least one pair of nondecreasing functions 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 $\mathrm{Prob}(X > x \mid Y = y)$ is nondecreasing in y for all x.

In terms of the probability matrix Q, the condition that X is PRD on Y can be written as follows: For all $i = 2, \cdots, m-1, j < j'$ implies $\sum_{l=i}^{m} q_{lj}/c_j \leqq \sum_{l=i}^{m} q_{lj'}/c_{j'}$.

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

We note that it is easily shown if Q corresponds to Y being PRD on X (X being PRD on Y), then every $\tilde{Q}$ has the same property, where $\tilde{Q}$ is obtained from Q by adding together (which is equivalent to statistically collapsing data categories) any sets of adjacent rows or adjacent columns. As a consequence of this fact and of Theorem 2.2, it follows that Q corresponding to Y is PRD on X and X is PRD on Y implies that $\rho'(\tilde{Q}) = \rho_{\mathrm{CMC}}(\tilde{Q})$ for every collapsed $\tilde{Q}$. However, Chhetry and Sampson (1987) provide an example that the conditions of Theorem 2.2 are not necessary for $\rho'(Q) = \rho_{\mathrm{CMC}}(Q)$.

In the study of bivariate dependence concepts, it oftentimes is of interest to consider $P(\mathbf{r}, \mathbf{c})$, the class of all $m \times n$ probability matrices with fixed row and column marginals, $\mathbf{r}$ and $\mathbf{c}$, respectively. It is well known that (see Schriever (1985, Ex. 4.2.3)) $\rho_{\mathrm{CMC}}(Q^+) \geqq \rho_{\mathrm{CMC}}(Q)$ for all $Q \in P(\mathbf{r}, \mathbf{c})$, where Q^+ is the probability matrix uniquely corresponding to the cumulative distribution matrix of the upper Fréchet bound, which has $F^+ = \{(\min(F_i, G_j))\}$, where $F_i = \sum_{k=1}^{i} r_k$ and $G_j = \sum_{k=1}^{i} c_k$. If the random variables X and Y are both continuous, the CMC for the correspondingly defined upper Fréchet bound is one (see Kimeldorf and Sampson (1978)). However, in the discrete situation it is not always the case that $\rho_{\mathrm{CMC}}(Q^+)$ is one. In the following theorem we provide a necessary and sufficient condition for $\rho_{\mathrm{CMC}}(Q^+) = 1$ in terms of the marginal row and column sums.

THEOREM 2.3. *A necessary and sufficient condition for $\rho_{\mathrm{CMC}}(Q^+) = 1$ 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 $Q^+ = \mathrm{Diag}(Q_1^+, Q_2^+)$ 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, $Q^+ = \mathrm{Diag}(Q_1^+, Q_2^+)$ 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

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

This implies that the corresponding 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 $\rho_{\mathrm{CMC}}(Q) = \rho_{\mathrm{DMC}}(Q)$; additionally, $\rho_{\mathrm{CMC}}(Q) = -1$ ($\rho_{\mathrm{DMC}}(Q) = 1$) 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 $\rho_{\mathrm{CMC}}(Q) = \rho_{\mathrm{DMC}}(Q)$ if and only if X and Y are independent.*

Proof. Suppose $\rho_{\mathrm{CMC}}(Q) = \rho_{\mathrm{DMC}}(Q) = \eta \neq 0$ (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) $\operatorname{Var}[a_1(X)] = \operatorname{Var}[a_2(X)] = \operatorname{Var}[b(Y)] = 1$. Then, by the assumption that $\rho_{\mathrm{CMC}}(Q) = \rho_{\mathrm{DMC}}(Q)$,

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

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

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

The proof of Corollary 2.5 is obvious.

3. Some results concerning evaluation. While the quantities $\rho'(Q)$ and $\rho_{\mathrm{CMC}}(Q)$ are of interest in their own right as measures of association, the vectors at which these maxima occur play an important role in rescaling of the values of the random variables. These notions are particularly useful in statistically analyzing both nominal and ordinal contingency tables (e.g., Nishisato (1980)). The vectors that maximize $\rho'(Q)$ can be derived from certain results of statistical correspondence analysis (e.g., Benzecri (1973) and Hill (1974)). The increasing vectors that yield $\rho_{\mathrm{CMC}}(Q)$ can be interpreted as either providing dual scalings for ordinal contingency tables or a form of ordinal correspondence analysis. However, their evaluation is substantially more complicated than the nonordinal case (e.g., see KMS, or Breiman and Friedman (1985), and the comments of Buja and Kass (1985)). Chhetry and Sampson (CS) (1987) provide an approach that simplifies somewhat the calculation of $\rho_{\mathrm{CMC}}(Q)$ and the maximizing vectors. We briefly discuss that approach and then detail how to employ it effectively when the ordinal table is collapsed, i.e., when neighboring row or columns are added. The latter issue is important for the statistical modeling using hierarchies 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

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

where $\bar{A} = (I_m - \mathbf{1}_m \mathbf{1}_m' D_r)\Psi_m$, $\bar{B} = (I_n - \mathbf{1}_n \mathbf{1}_n' D_c)\Psi_n$, and Ψ_p is the $p \times (p-1)$ matrix whose (i,j)th element is zero, if $i \leqq j$, and 1, otherwise. Let $\Sigma_{11} = \bar{A}' D_r \bar{A}$, $\Sigma_{12} = \bar{A}' Q \bar{B}$, $\Sigma_{22} = \bar{B}' D_c \bar{B}$, 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 $\alpha \in R^{m-1}$, $\beta \in R^{n-1}$

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

where $\alpha \neq 0$ and $\beta \neq 0$. Then CS show that the maximal correlation coefficient and the two monotone correlation coefficients can be evaluated as follows:

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

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

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

The relationships of (3.3a)–(3.3c) can be viewed as simplifying computation by reducing dimensionality. Also note that if α_0 and β_0 optimize any of (3.3a), (3.3b), or (3.3c), then the corresponding maximizing vectors $\mathbf{f}_0$ and $\mathbf{g}_0$ defining the left-hand sides are related by $\mathbf{f}_0 = \bar{A}\alpha_0$ and $\mathbf{g}_0 = \bar{B}\beta_0$. For example, if $r_Q(\alpha, \beta)$ is maximized at α_0, β_0, then $\rho_Q(\mathbf{f}, \mathbf{g})$ is maximized at $\mathbf{f}_0 = \bar{A}\alpha_0$ and $\mathbf{g}_0 = \bar{B}\beta_0$.

An additional 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 good discussion concerning traditional multivariate normal canonical correlations is given in Anderson (1984, Chap. 12). For the p-dimensional multivariate normal distribution with positive-definite covariance matrix Σ, canonical correlation analysis involves a study of the determinental roots and solutions for $\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} - \lambda^2\Sigma_{22}$, where Σ_{11}, Σ_{12}, Σ_{21}, Σ_{22} are a partitioning of Σ with the dimension of Σ_{11} being $p_1 < p$. A description of the relationship between our problem and traditional canonical correlation analysis 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 $\alpha^{(1)} \neq 0$ and $\beta^{(1)} \neq 0$ satisfy the equations*

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

and

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

then $\rho_Q(\alpha^{(1)}, \beta^{(1)}) = \rho'(Q)$. Moreover, $\rho'(Q) = \rho_{\mathrm{CMC}}(Q)$ if and only if there exist nonnegative vectors $\alpha^{(1)}$ and $\beta^{(1)}$ 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. Recent discussions on the general issue of collapsing nonordinal contingency tables are 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 $P = \{p_{ij}\}$, $m \leqq n$, 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 $e = j + 1, \cdots, k - 1$.

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 $\tilde{Q} = P_1QP_2'$, 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 obtained from Q by collapsing and with row and column marginals $\tilde{\mathbf{r}} = P_1\mathbf{r} = (\tilde{r}_1, \cdots, \tilde{r}_s)'$ and $\tilde{\mathbf{c}} = P_2\mathbf{c} = (\tilde{c}_1, \cdots, \tilde{c}_t)'$, respectively. Moreover, if $D_{\tilde{r}} = \mathrm{Diag}(\tilde{r}_1, \cdots, \tilde{r}_s)$ and $D_{\tilde{c}} = \mathrm{Diag}(\tilde{c}_1, \cdots, \tilde{c}_t)$, then $D_{\tilde{r}} = P_1D_rP_1'$ and $D_{\tilde{c}} = P_2D_cP_2'$.

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

THEOREM 3.3. *If $\tilde{Q} = P_1QP_2'$, where P_1 and P_2 are, respectively, $s \times m$ and $t \times n$ C-matrices, then*

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

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

$$(\mathbf{e}_2 - \mathbf{e}_1, \mathbf{e}_3 - \mathbf{e}_2, \cdots, \mathbf{e}_p - \mathbf{e}_{p-1}).$$

Proof. From CS (Lemma 3.2(i))

$$\Sigma_{12}(\tilde{Q}) = \Psi_s'(\tilde{Q} - \tilde{\mathbf{r}}\tilde{\mathbf{c}}')\Psi_t$$
$$= \Psi_s' P_1 (Q - \mathbf{rc}') P_2' \Psi_t.$$

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 relationship 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

$$P_1 \equiv (\mathbf{e}_1, \cdots, \mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_{m-s}) \quad \text{and} \quad P_2 \equiv (\mathbf{e}_1, \cdots, \mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_{n-t}).$$

Then, the matrices, K_m and K_n defined in Theorem 3.3 reduce to the form

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

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

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

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

and

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

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

REFERENCES

T. W. ANDERSON (1984), *An Introduction to Multivariate Statistical Analysis*, 2nd ed., John Wiley, New York.

CH. BASTIN, J. P. BENZECRI, CH. BOURGARIT, AND P. CAZES (1980), *Pratique de l'analyse de donnees*, Dunod, Paris.

J. P. BENZECRI (1973), *L'analyse des donnees* II, Dunod, Paris.

L. BREIMAN AND J. H. FRIEDMAN (1985), *Estimating optimal transformations for multiple regression and correlation*, J. Amer. Statist. Assoc., 80, pp. 580–598.

A. BUJA AND R. E. KASS (1985), *Some observations on* ACE *methodology*, J. Amer. Statis. Assoc., 80, pp. 602–607.

D. CHHETRY AND A. SAMPSON (1987), *A projection decomposition for bivariate discrete probability distributions*, SIAM J. Algebraic Discrete Methods, 8, pp. 501–509.

R. A. FISHER (1940), *The precision of discriminant functions*, Ann. Eugen. London, 10, pp. 422–429.

Z. GILULA (1986), *Grouping and association in contingency tables: An exploratory canonical correlation approach*, J. Amer. Statist. Assoc., 81, pp. 773–779.

Z. GILULA AND A. M. KRIEGER (1983), *The decomposability and monotonicity of Pearson's chi-square for collapsed contingency tables*, J. Amer. Statist. Assoc., 78, pp. 176–80.

L. GOODMAN AND W. KRUSKAL (1979), *Measure of Association for Cross Classifications*, Springer-Verlag, New York.

S. HABERMAN (1982), Association, measures of, in Encyclopedia of Statistical Sciences (Vol. 1), S. Kotz and N. Johnson, eds., John Wiley, New York, pp. 130–137.

W. J. HALL (1969), *On characterizing dependence in joint distributions*, in Essays in Probability and Statistics, R. Bose, I. Chakravarti, P. Mahalanobis, C. Rao, and K. Smith, eds., University of North Carolina Press, Chapel Hill, NC.

M. O. HILL (1974), *Correspondence analysis: A neglected multivariate method*, Appl. Statist., 23, pp. 340–354.

H. O. HIRSCHFELD (1935), *A connection between correlation and contingency*, Proc. Cambridge Philos. Soc., 31, pp. 520–524.

G. KIMELDORF AND A. R. SAMPSON (1978), *Monotone dependence*, Ann. Statist., 6, pp. 895–903.

G. KIMELDORF, J. MAY, AND A. R. SAMPSON (1982), *Concordant and discordant monotone correlations and their evaluation by nonlinear optimization*, in Optimization in Statistics, S. Zanakis and J. Rustagi, eds., TIMS Studies Management Sci., 19, pp. 117–130.

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

——— (1958), *The structure of bivariate distributions*, Ann. Math. Statist., 29, pp. 719–736.

E. L. LEHMANN (1966), *Some concepts of dependence*, Ann. Math. Statist., 37, pp. 1137–1153.

C. MACDUFFEE (1949), *Vectors and Matrices*, The Mathematical Association of America, Washington, DC.

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

A. RAVEH (1986), *On measures of monotone association*, Amer. Statist., 40, pp. 117–123.

A. RÉNYI (1959), *On measures of dependence*, Acta. Math. Acad. Sci. Hungar., 10, pp. 441–451.

H. RICHTER (1949), *Zur maximal correlation*, Z. Angew. Math. Mech., 19, pp. 127–128.

O. V. SARMANOV (1958a), *The maximal correlation coefficient (symmetric case)*, Dokl. Akad. Nauk. SSSR, 120, pp. 715–718. (In Russian.) Sep. Transl. Math. Statist. Probab., 4, pp. 271–275. (In English.)

——— (1958b), *The maximal correlation coefficient (non-symmetric case)*, Dokl. Akad. Nauk. SSSR, 121, pp. 52–55. (In Russian.) Sep. Transl. Math. Statist. Probab., 4, pp. 207–210. (In English.)

B. F. SCHRIEVER (1985), *Order dependence*, Ph.D. thesis, Free University of Amsterdam, Amsterdam, the Netherlands.

——— (1983), *Scaling of order dependent categorical variables with correspondence analysis*, Internat. Statist. Rev., 51, pp. 225–238.