PROPERTIES OF MEASURES OF ASSOCIATION
FOR NOMINAL VARIABLES

by

Antonio Forcina and Alessandra Giovagnoli

Technical Report No. 224
March 1986

Prepared under the Auspices
of
National Science Foundation
DMS 84-11411
Ingram Olkin, Project Director

Department of Statistics
Stanford University
Stanford, California
PROPERTIES OF MEASURES OF ASSOCIATION
FOR NOMINAL VARIABLES
Antonio Forcina and Alessandra Giovagnoli
Institute of Statistics, University of Perugia

SUMMARY

We investigate the mathematical structure underlying the concept of association (dependence) between two nominal random variables. On the set of all possible bivariate distributions for such variables ($m \times n$ tables) we define some equivalence and order relations, showing that they have a statistical interpretation as regards comparing the degree of dependence of one variable on the other. We study the interconnections of these relations and demonstrate their implications. Furthermore, a non-trivial property of measures of association, implicit in the concept of association itself, is shown to lead to a useful functional equation. These properties are then applied to two wide-ranging classes of measures of association that include the most common measures considered in the literature.

Key words: Association; chain majorization; dependence; measures of dispersion; measures of optimal prediction; measures of relative information; nominal random variables.
1. Introduction

Interest in measures of association as numerical assessments of the strength of the statistical dependence of two qualitative random variables has a long history. Goodman and Kruskal's (1959) colossal survey of historical and bibliographical material lists many contributions which originated within the context of different applied sciences. Recently the subject has enjoyed a revival, with contributions by Haberman (1982a, 1982b) and indirectly by Light and Margolin (1971), Patil and Taillie (1982), Rao (1982) and Goodman (1985).

This paper deals with measures of the dependence of $Y$ on $X$ for nominal polytomous random variables $Y$ and $X$. When the ranges of $Y$ or $X$ are ordered, i.e. for ordinal qualitative (or quantitative) variables, more specialized concepts of association can be formulated (see, e.g. Lehmann, 1966) and there is now an extensive literature for this particular case (see Kimeldorf and Sampson, 1978; Shaked, 1982; Agresti, 1984, for related references).

Association of $Y$ with $X$ will be determined by the conditional distributions of $Y$ given $X$ and, possibly, by the marginal distribution of $X$. However, except in very special cases, (see Goodman and Kruskal, 1954, p. 735) association does not correspond to a well-defined concept because usually the purpose of using a particular measure of association cannot be stated specifically: one may have a multiplicity of goals in mind or simply desire to summarize a large amount of data. The need to motivate the choice to be made among the myriad of existing measures has inspired a great deal of the research on the subject, as in the case of the search for operational interpretations of certain measures in the context of appropriate models of activity (Goodman and Kruskal, 1954).

However, interpretations of descriptive measures are sometimes based on ad hoc arguments that may, occasionally, conceal an undesirable behavior. Moreover, it is not clear in general when the choice between different measures is in fact relevant. A more fundamental investigation may be that outlined by Bickel and Lehmann (1975) in their introductory paper on descriptive statistics: one could characterize the attribute to be measured (association) by a consistent set of minimal requirements (desiderata) having a direct statistical interpretation. In other words, to clarify the question "how do we measure association?" one could try to define explicitly (a) which operations on a table $P$, expressing the joint distribution of $X$ and
do not affect its degree of association, and (b) when a cross-classification table $P$ possesses the attribute (association) more strongly than a table $Q$. In the present paper we have chosen to scrutinize the intuitive contents of the concept of association along these lines. So we look at orderings for $m \times n$ matrices. Although many orderings can be devised, because row and column sums represent marginal distributions, we keep in mind interpretability: the principles and orderings proposed arise from the particular goal of comparing measures of association of two such matrices, and appear to be particularly suited for this purpose.

After defining our basic notations in Section 2, we introduce in Section 3 some general principles of equivalence, stating that certain classes of tables all have the same degree of association, and of order, stating that certain tables have stronger association than others. We concentrate on the implications and mutual interconnections of the various principles in order to clarify their intuitive meaning and establish general conditions under which the existing measures preserve the relationships implied by each principle.

In Section 4 we deal with ways to decompose a given table: an additional requirement, which seems very reasonable, on the behavior of measures of association relative to such decompositions is shown to be equivalent to a useful functional equation for those measures.

The results of the previous sections are applied to the study of measures of association which can be thought of as measures of optimal prediction of $Y$ given $X$ (see Haberman, 1982a). We show that a large class of these measures satisfy the general requirements derived in Sections 3 and 4. Other well-known measures of association, which are not measures of optimal prediction, are shown to have an interpretation in terms of mean relative information. We establish the behavior of these measures as regards the properties introduced in Sections 3 and 4.

2. Notation

Let $X$ and $Y$ be nominal random variables with finite ranges denoted respectively by $\{x_1, \ldots, x_m\}$ and $\{y_1, \ldots, y_n\}$. Let $p_{ij} = P(X = x_i, Y = y_j)$, \((i = 1, \ldots, m; j = 1, \ldots, n)\) denote the table of joint probabilities and let $P = (p_{ij})$ be the $m \times n$ matrix of probabilities. A matrix of this description is often referred to in the literature as a cross-classification table. A measure of association will be a suitable function from the set of these matrices to $\mathcal{R}$. The
phrase "the degree of association of $Y$ with $X$ in table $P''$ will be shortened as $d(P)$. When comparing association in different tables, $d(Q) \leq d(P)$ will mean that association in $Q$ is less than or equal to that in $P$.

Denote the row sums of $P$ by the column vector $p_r = Pe = (p_1, \ldots, p_m)^T$ and the column sums by $p_0 = e^T P = (p_1, \ldots, p_n)$, which are the marginal distributions of $X$ and $Y$ respectively. (Here $e$ is a column vector of all ones). We shall also write $\Delta P = \text{diag}(p_1, \ldots, p_m)$ to denote the diagonal matrix with elements $p_1, \ldots, p_m$.

The table of conditional probabilities of $Y$ given $X$ is any $m \times n$ matrix $P^*$ with non-negative entries and rows adding up to 1 such that $\Delta P P^* = P$. We put $P^* = (p_{ij}^*)$. Thus $P^*$ is not uniquely defined: if $p_i = 0$ the $i$-th row of $P^*$ is an arbitrary distribution. However, as will be shown, no serious inconvenience is experienced as a result of this non-uniqueness.

It can be argued on intuitive grounds that rows of $P$ which are made up of zeros are irrelevant in assessing the association of $Y$ and $X$. On the other hand, artificially adding null rows may be a useful technical device for comparing tables of different dimensions. This can be stated as a principle:

(2.1) *Null rows may be inserted or deleted in a cross-classification table without affecting its degree of association.*

Because of (2.1) it is possible in some cases to assume that $P$ has no null rows. This simplifies the calculations and does not affect the results.

The random variables $X$ and $Y$ are independent if $p_{i,j} = p_{i,j}$ for all $i, j$. The class of all $m \times n$ cross-classification tables with this property will be denoted by $\mathcal{H}$. Complete dependence of $Y$ on $X$ is not so clearly defined in the literature: we shall take it to mean the existence of a function $k : \{x_1, \ldots, x_m\} \rightarrow \{y_1, \ldots, y_n\}$ such that $Y = k(X)$. Denote by $\mathcal{K}$ the class of all $m \times n$ tables with at most one non-zero element in each row. The geometric interpretation of $P$ and $P^*$ is the following: the rows of $P^*$ can be thought of as points of the fundamental simplex

$$S_n = \{(z_1, \ldots, z_n), \ x_i \in \mathbb{R}, z_i \geq 0, \Sigma z_i = 1\}.$$ 

Observe that $p_0^T P^* = p_0$ so that $p_0$ is the center of gravity of the rows of $P^*$ with weights $p_1, \ldots, p_m$. Independence means that each row of $P^*$ coincides with $p_0$, and complete depen-
idence means that each row is a vertex of $S_n$. All the other cases are somewhat intermediate.

A case in which the geometry is useful in understanding the probabilistic contents, as will be seen later, occurs when the rows of $P^*$ (with non-null weights) are linearly independent.

**Definition:** A table $P$ with linearly independent non-null rows will be called *non-degenerate*.

For $m = 2$ clearly all tables not in $\mathcal{H}$ are non-degenerate.

**Lemma 1:** If $P$ is non-degenerate, its non-null rows are uniquely determined by $p_0$ and $P^*$.

**Proof:** The proof is straightforward since $p_i^T P^* = p_0$ and $P^*$ without the rows corresponding to $p_i = 0$ is right-invertible.\|

3. Equivalences and Orderings

By the very nature of $X$ and $Y$ (nominal random variables) the labelling of their respective categories is immaterial. We can assume row and column exchangeability as equivalence principles:

(3.1) *(Row exchangeability).* Association is invariant under permutations of the rows.

(3.2) *(Column exchangeability).* Association is invariant under permutations of the columns.

Principles (3.1) and (3.2) state that we can pre- or post-multiply a cross-classification table $P$ by a permutation matrix $\Pi$ without altering the degree of association. Suppose now that the conditional distribution of $Y$ given $X = x_i$ is the same as that of $Y$ given $X = x_{i'}$, for some $1 \leq i, i' \leq m$. Then the value of $X$ provides no information on $Y$ once it is known that either $x_i$ or $x_{i'}$ has occurred. This leads to the formulation of a further equivalence principle:

(3.3) *(Row aggregation and row splitting).* Association is unaltered by replacing one of two proportional rows of a cross-classification table by their sum and the other one by a row of zeros.

This property has been described as "splitativity" by Good (1981) who shows it to hold for a particular class of association measures.

Due to the lack of symmetry there does not seem to be any obvious reason why the analogue of (3.3) for columns rather than for rows should be postulated and indeed no such
property holds for the most common measures of association, as discussed in Section 5, unless they are symmetric, which can be seen by counterexamples.

We now discuss some properties involving inequalities among the degree of association of different tables and their links with the above equivalence principles. Two fundamental properties are assumed in the literature (see Goodman and Kruskal, 1954, p. 738).

(3.4) The association of tables of $\mathcal{H}$ is minimal.

(3.5) The association of tables of $\mathcal{K}$ is maximal.

When a measure of association is 0 on $\mathcal{H}$ and 1 on $\mathcal{K}$, we say that the measure is normalized. For non-normalized measures it can be shown under such mild assumptions as (2.1) and (3.3) that the degree of association of tables in either $\mathcal{H}$ or $\mathcal{K}$ depends at most on the Y-marginal distribution. Indeed for some measures such dependence does occur for $\mathcal{K}$, so that a weaker form of (3.5) is the postulate:

(3.5)(W) Among all tables with fixed margin $p_0$, those in $\mathcal{K}$ have maximum association.

By symmetry an analogous weak form, say (3.4)(W), of (3.4) is also possible.

The next step is to introduce a pre-order relation among the $m \times n$ tables and discuss its relevance for comparing the degree of association of tables not necessarily belonging to $\mathcal{H}$ or $\mathcal{K}$. This will lead to the formulation of another principle.

Assume that $P$ and $Q$ are $m \times n$ tables such that the points of $S_n$ corresponding to the rows of $Q^*$ with non-null weights lie in the convex hull of those of $P^*$, i.e. there is a stochastic matrix $\tilde{S}$ such that

(i) $Q^* = \tilde{S} P^*$,

and assume also that the centers of gravity are the same

(ii) $p_0 = q_0$.

The probabilistic interpretation of (i) and (ii) is that each statement $X = x_i$ ($i = 1, \ldots, m$) in $Q$ gives the same information about $Y$ as the set of statements $X = x_h$, with probabilities $\tilde{s}_{ih}$ ($h = 1, \ldots, m$) respectively, in $P$. However conditions (i) and (ii) are not sufficient to guarantee a decrease in association because their effect can be counterbalanced by appropriate
shifts in the X-margin, as the following example shows:

\[ P = \begin{pmatrix} 0.1 & 0 \\ 0.49 & 0.49 \\ 0 & 0.1 \end{pmatrix}, \quad Q = \begin{pmatrix} 0.48 & 0.1 \\ 0.1 & 0.1 \\ 0.1 & 0.48 \end{pmatrix}. \]

Here the second conditional distribution of Y in P coincides with the marginal distribution and has most of the weight, whereas in Q most of the weight is on the first and last conditional distributions of Y, which are in every way very different from the marginal. So we need an extra condition on the X-marginal distribution. When P is non-degenerate, such a condition follows automatically from (i) and (ii) since, because of Lemma 1

(iii) \[ p_r = S^T q_r. \]

We argue that this condition is appropriate in general. Consider first the special case where

\[ S = \begin{pmatrix} \alpha & 1 - \alpha & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \quad 0 \leq \alpha \leq 1. \]

Each of the rows of \( Q^* \) contains the same information about Y as \( P^* \), apart from the first row, which is less informative, and this row has more weight than in P. Altogether there is a loss of information and hence a loss of association. The argument can be extended to a general matrix \( \tilde{S} \). Observe that (i) and (iii) imply (ii).

We shall be interested in the pre-ordering on \( m \times n \) cross-classification tables defined by (i) and (iii). The following theorem provides a characterization of these properties.

**Theorem 1:** Conditions (i) and (iii) are equivalent to the existence of a stochastic matrix \( S \) such that \( Q = S^T P \).

**Proof:** Take \( S = \Delta_P^+ S \Delta_Q^T \), with \( \Delta_P^+ \) the Moore-Penrose inverse.||

The above provides intuitive support for the following definition.

**Definition:** \( Q \preceq P \) if there exists a stochastic matrix \( S \) such that \( Q = S^T P \),

and to the following principle:

(3.6) If \( Q \preceq P \), then \( d(Q) \leq d(P) \).

6
It is of interest to consider in what way (3.6) is related to the equivalence principles stated above.

**Theorem 2:** (i) (3.6) implies (3.1), (ii) (3.6) implies (3.3).

**Proof:** (i) If \( Q = \Pi P \), with \( \Pi \) a permutation matrix, then \( P = \Pi^T Q \), with \( \Pi, \Pi^T \) stochastic. By (3.6) the association of \( Q \) cannot exceed that of \( P \) and vice versa, so (3.1) holds.

(ii) Let \( P \) be a table with two proportional rows, say \( p_{1j} = \theta p_{2j}, \theta > 0, j = 1, \ldots, n \) and define \( P' = (p'_{ij}) : p'_{1j} = p_{1j} + p_{2j}, p'_{2j} = 0, p'_{ij} = p_{ij} \) for all \( j = 1, \ldots, n; i = 3, \ldots, m \). Let

\[
S_1 = \begin{pmatrix}
1 & 1 & : & 0 \\
0 & 0 & : & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \vdots & I_{m-2}
\end{pmatrix}, \quad S_2 = \begin{pmatrix}
1/(1+\theta) & \theta/(1+\theta) & : & 0 \\
\theta/(1+\theta) & 1/(1+\theta) & : & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \vdots & I_{m-2}
\end{pmatrix}.
\]

Both \( S_1^T \) and \( S_2^T \) are stochastic and \( P' = S_1 P, P = S_2 P'. \) By (3.6) the association of \( P \) and of \( P' \) is the same.||

**Theorem 3:**

(i) If \( P \in \mathcal{K} \) and \( Q \preceq_P P \), then \( Q \in \mathcal{K} \).

(ii) If \( Q \in \mathcal{K} \) and \( Q \preceq_P P \), then \( P \in \mathcal{K} \).

(iii) (3.6) implies (3.4)(W) and (3.5)(W).

**Proof:** Implications (i) and (ii) follow immediately from the geometric interpretations of \( \mathcal{K} \) and \( \mathcal{K} \) and of Theorem 1.

(iii) We show that if \( Q \in \mathcal{K} \) and \( P \) is such that \( e^T P = e^T Q \), then \( Q \preceq_P P \). Since \( Q \in \mathcal{K} \), \( Q = q_r \cdot p_0 = Q e e^T P = Q J P \), where \( J \) is the \( n \times m \) matrix of all ones and \( (QJ)^T \) is stochastic. So if (3.6) holds the association of \( Q \) is less than or equal to that of \( P \), for all such \( P \).

Now let \( P \in \mathcal{K} \) and \( Q \) be any table such that \( e^T P = e^T Q \). Because of Theorem 2, \( P' = (\text{diag}(p_0) : 0)^T \preceq_{	ilde{S}} P \) and, for geometric reasons, \( Q^* = \tilde{S} \ (P')^* \) with \( \tilde{S} \) stochastic. Since \( p_0 = q_0 \) and \( P' \) is non-degenerate, \( Q \preceq_{P'} P \) and, by transitivity, \( Q \preceq_{\tilde{S}} P \).||

We now investigate a different order relation between tables that is based on an ordering between matrices already considered in the literature (see Marshall and Olkin, 1979, p. 430).
although little is known about its properties.

Definition: $P$ is said to chain-majorize $Q$ (written $Q \prec P$) if there exists a finite product $D = \prod_j D_j$ of doubly stochastic matrices with exactly two non-zero off-diagonal entries such that $Q = PD$.

For brevity we shall call such a matrix a T-matrix. The relation $\prec$ is a pre-ordering. Observe that $Q \prec P$ implies $q_r = p_r$ and $q_0 = p_0 D$. Thus $Q$ has the same row margin as $P$.

Let us take $D = \alpha I + (1 - \alpha)\Pi$ with $0 \leq \alpha \leq 1$ and $\Pi$ a permutation matrix that exchanges only two elements. Then $Q = \alpha P + (1 - \alpha)P'$, with $P'$ obtained from $P$ exchanging two columns, say $j$ and $j'$. In practice this can be interpreted as a model for errors in the $Y$ variable, where category $j$ is mistaken (for example in questionnaire response) for $j'$ and vice versa with probability $(1 - \alpha)$, independently of $X$. Observing that $Q$ coincides with $P$ except for columns $j$ and $j'$, and the subtable of $Q$ made up of these two columns approaches independence as $\alpha$ tends to 0.5 (from left or right), we can argue that the degree of association of $Q$ cannot be greater than that of $P$. Repeated application of this argument, if accepted, yields the following principle:

(3.7) If $Q \prec P$, then $d(Q) \leq d(P)$.

A straightforward implication of (3.7) is

Theorem 4: If (3.7) holds, then association is invariant with respect to the exchange of columns.

Proof: Permutation matrices are T-matrices.||

Since every T-matrix is doubly stochastic, by Birkhoff's Theorem (see for instance Marshall and Olkin, 1979, p. 19) we obtain

Theorem 5: $Q \prec P$ implies $Q = \sum_h \alpha_h (P \Pi_h), h \in I$, where $I$ is a finite index set, $\alpha_h > 0, \sum_h \alpha_h = 1$ and $\Pi_h$ are permutation matrices for all $h$.

Thus by Theorem 5 requirement (3.7) will be met by all association measures which are matrix-convex and invariant under column permutations. A necessary and sufficient condition for a measure of association to preserve $\prec$ is given in the differentiable case by Rinott (1973).

The following theorem is the analogue of Theorem 3, part (i) and (ii):
Theorem 6: (i) If \( P \in \mathcal{K} \) and \( Q \prec P \), then \( Q \in \mathcal{K} \). (ii) If \( Q \in \mathcal{K} \) and \( Q \prec P \), then \( P \in \mathcal{K} \).

Proof: The former holds since if \( P = p_r \cdot p_0 \) and \( Q = PD \), then \( Q = p_r p_0 D = q_r q_0 \). If \( Q \in \mathcal{K} \) and \( Q \prec P \), then \( Q = \Sigma_h \alpha_h (\Pi \Pi_h) \) by Theorem 5. Because \( Q \) has at most one positive element in each row, the same must also be true for all \( \Pi \Pi_h \) such that \( \alpha_h > 0 \). This implies that \( \Pi \Pi_h \in \mathcal{K} \) for all such \( h \), i.e. \( P \in \mathcal{K} \), and also that there is only one positive \( \alpha_h \), so that \( Q = \Pi \Pi \) with \( \Pi \) a permutation matrix. ||

The following counterexample shows that the analogue of (iii) of Theorem 3 does not hold:

\[
P = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix} \in \mathcal{K}, \quad Q = \begin{pmatrix} 0.4 & 0.1 \\ 0.4 & 0.1 \end{pmatrix} \in \mathcal{K}, \quad p_r = q_r
\]

but \( Q \not\prec P \) since \( p_0 D = p_0 \) for all \( T \)-matrices \( D \).

Finally we observe that the two order relations \( \preceq_S \) and \( \prec \) can be combined together to produce a more general ordering, which we shall denote by \( \preceq_S \), as follows:

Definition: \( Q \preceq_S P \) iff there exist a stochastic matrix \( S \) and a \( T \)-matrix \( D \) such that \( Q = S^T P D \).

Clearly \( Q \preceq_S P \) or \( Q \prec P \Rightarrow Q \preceq_S P \), thus (3.6) and (3.7) together are equivalent to

(3.8) \( Q \preceq_S P \), then \( d(Q) \leq d(P) \).

4. Decompositions

Any cross-classification table \( P \) can be thought of as a mixture of different tables with the same \( Y \)-margin. The particular mixtures defined below turn out to be a useful tool for investigating new features of the concept of association.

Assume \( Z \) is another nominal random variable with finite range \( \{ z_h, h \in I \} \) and probability distribution \( P(Z = z_h) = \alpha_h \), and let \( X \) and \( Y | X = x_i \) for all \( i \) be independent of \( Z \). Let

\[
P_h = (p_{ij|h}), \quad h \in I
\]

be the tables of conditional distributions: \( p_{ij|h} = P(X = x_i, Y = y_j | Z = z_h) \). Then \( P \) is a mixture of the \( P_h \)'s with weights \( \alpha_h, h \in I \). More formally
Definition: For a given $P$ with $e^T P = p_0$, a finite set of $m \times n$ tables $\{P_h\}_{h \in I}$ together with a set of “weights” $\{\alpha_h, h \in I : \alpha_h \geq 0, \Sigma_h \alpha_h = 1\}$ is called a decomposition of $P$ if

$$P = \sum_{h \in I} \alpha_h P_h$$

1) $e^T P_h = p_0$

2) it is possible to choose $P_h^*$ and $P^*$ so that $P_h^* = P^*$; in other words $\Delta_h P^* = P_h \ \forall h \in I$.

In view of the fact that $Z$ contains no additional information on $Y|X$ and $Y$, it appears natural to state the principle

(4.1) If $P$ and $Q$ are cross-classification tables and $\{P_h, \alpha_h\}, \{Q_h, \alpha_h\}, h \in I$ are respective decompositions such that $d(P_h) = d(Q_h)$ for all $h$, then $d(P) = d(Q)$.

The following theorem demonstrates that a functional equation holds for all measures of association satisfying (4.1), namely they are “associative means”, in the sense defined by de Finetti (1931), of the associations of the tables of the decomposition.

**Theorem 7:** If $\phi$ is a measure of association which is continuous in the entries of $P$, and satisfies (3.4) and (3.5), then (4.1) implies

(4.2) There exists a continuous strictly increasing function $g$ such that for any decomposition $\{P_h, \alpha_h\}$ of $P$

$$\phi(P) = g^{-1}(\Sigma_h \alpha_h g(\phi(P_h))).$$

The converse, i.e. (4.2) implies (4.1), holds trivially for any measure of association.

**Proof:** To show that (4.1) implies (4.2) the following tables play a fundamental role: for fixed $p_0$ define the $(n + 1) \times n$ matrix

$$P(\lambda) = \begin{pmatrix} \lambda \text{ diag}(p_0) \\ \hline \\ (1 - \lambda) p_0 \end{pmatrix}, \quad 0 \leq \lambda \leq 1.$$ 

Thus $P(\lambda)$ can be decomposed as $P(\lambda) = (1 - \lambda) P_H + \lambda P_K$ with

$$P_H = \begin{pmatrix} 0 \\ p_0 \end{pmatrix} \in \mathcal{K}, \quad P_K = \begin{pmatrix} \text{diag}(p_0) \\ 0^T \end{pmatrix} \in \mathcal{K}.$$ 

The function $\psi(\lambda) = \phi(P(\lambda))$ is independent of $p_0$: this follows from (3.4) and (3.5) which give that $\phi(P_H)$ and $\phi(P_K)$ are independent of $p_0$, and from (4.1). Since $\phi$ is continuous
in the \( p_{ij} \)'s, \( \psi(\lambda) \) is continuous in \( \lambda \). We want to show that it is strictly monotone. This is essentially the argument in de Finetti (1931, p. 383). We show that the existence in \([0, 1]\) of \( \lambda_1 < \lambda_2 \) such that \( \phi(\lambda_1) = \phi(\lambda_2) \) implies that \( \phi(\cdot) \) is constant in \([0, 1]\). First let \( \lambda = r\lambda_1 + (1 - r)\lambda_2, \ r \in [0, 1] \). This defines a decomposition of \( P(\lambda) \) in \( P(\lambda_1) \) and \( P(\lambda_2) \), and hence by (4.1) \( \phi(\cdot) \) must be constant in \([\lambda_1, \lambda_2]\). If we now let \( s = \max\{0, \lambda_1 - (\lambda_2 - \lambda_1)\} \), because of (4.1), \( \phi((\lambda_1 + s)/2) = \phi((\lambda_2 + s)/2) \). Moreover since \((\lambda_2 + s)/2 \in [\lambda_1, \lambda_2], \phi(\cdot) \) is constant in \([(\lambda_1 + s)/2, \lambda_2]\). The conclusion follows by repeated application of the previous argument and by letting \( s = \min\{\lambda_2 + (\lambda_2 - \lambda_1), 1\} \).

Thus \( \psi \) is invertible and \( g = \psi^{-1} \) exists in \([a, b]\) where \( a = \phi(H) \) and \( b = \phi(K) \) for \( H \in \mathcal{H}, K \in \mathcal{K} \). So for any \( P \), there is a \( \lambda = g(\phi(P)) \). Let \( D = \{P_h, \alpha_h\}_{h \in I} \). Define \( P' = \Sigma_h \alpha_h P(\lambda_h) \) with \( P(\lambda_h) \) as \( P(\lambda) \) above, with respect to the same \( p_0 = e^T P \). Thus \( D' = \{P(\lambda_h), \alpha_h\}_{h \in I} \) is a decomposition of \( P' \) and \( \phi(P_h) = \phi(P(\lambda_h)) \) for all \( h \in I \).

By (4.1)
\[
\phi(P) = \phi(P') = \phi((1 - \Sigma H \alpha_h \lambda_h)P_H + \Sigma \alpha_h \lambda_h P_K) \\
= \psi(\Sigma \alpha_h \lambda_h) = \psi(\Sigma \alpha_h \lambda_h \phi(P_h)).
\]

Corollary: Under the assumptions of the Theorem,

(i) \( \phi(P_h) \leq \phi(Q_h) \) for all \( h \) implies \( \phi(P) \leq \phi(Q) \);

(ii) if strict inequality holds on the left-hand side of (i) for at least one \( h \), then strict inequality holds on the right-hand side.

From the proof of Theorem 7 it follows that if (3.4) and (3.5) are replaced by the weaker forms (3.4)(W) and (3.5)(W) then (4.1) implies a weaker version of (4.2), namely

(4.2)(W). Given \( \phi \), for any \( P \) there exists a continuous, strictly increasing function \( g \) which depends on \( p_0 \) such that \( \phi(P) = g^{-1}(\Sigma \alpha_h \lambda_h \phi(P_h)) \) for any decomposition \( \{P_h, \alpha_h\} \) of \( P \).

Theorem 7 provides an easy way of checking whether (4.1) holds for a given association measure \( \phi \). It is easy to see for instance that if

(4.3) \[
\phi(P) = \varphi(\Sigma p_i, d(p_i, p_0))
\]

with \( \varphi \) increasing and invertible, and \( d(u, v) \) a function of two row-vector variables defined
on $S_n \times S_n$, which denotes their distance (divergence, dissimilarity) (see below), then (4.2) holds for $\phi$. On the other hand assuming (4.1) rules out some measures of association discussed in Goodman and Kruskal (1959) such as $\phi_1(P) = \Sigma p_i.p_r.d(p_i^*, p_r^*)$ and also $\phi_2(P) = \text{median}\{d(p_i^*, p_0)|p_r\}$.

**Proposition 1:** Any association measure of the form (4.3) satisfies (3.1) and (3.3).

The proof of Proposition 1 is straightforward.

We shall explore some of the properties of the association measures which satisfy (4.3).

**Theorem 8:** Let $\phi$ be of the form (4.3).

(i) If $d(u, v)$ is convex in $u$ for any given $v$, then $\phi$ satisfies (3.6).

(ii) If $d(uD, vD) \leq d(u, v)$ for all $(u, v) \in S_n \times S_n$ and any $T$-matrix $D$, then $\phi$ satisfies (3.7).

**Proof:** It is sufficient to consider $\tilde{\phi} = \varphi^{-1} \circ \phi$.

(i) Let $Q \preceq P$, that is, $Q^* = SP^*, p_r = S^T q_r, S$ stochastic, and let $S = (s_{hi})$. Then

$$\tilde{\phi}(Q) = \Sigma h q_h.d(q_h^*, q_0) = \Sigma h q_h.d(\Sigma_i s_{hi}p_i^*, p_0)$$

$$\leq \Sigma_i (\Sigma_h s_{hi} q_h) d(p_i^*, p_0)$$

$$= \Sigma_i p_i.d(p_i^*, p_0) = \tilde{\phi}(P)$$

(ii) Let $Q \preceq P$. This implies $q_i = p_i$ and $p_i^* D = q_i^*$ for all $1 \leq i \leq m$. Furthermore $q_0 = p_0 D$. Then clearly $\tilde{\phi}(Q) \leq \tilde{\phi}(P)$.

**Corollary:** If (3.2) holds and $\phi$ is of the form (4.3) with $d(u, v)$ a convex function on $S_n \times S_n$, then $\phi$ satisfies both (3.6) and (3.7).

**Proof:** The proof follows from Theorems 7 and 5.||

Finally we establish a connection between decompositions and the idea of a non-degenerate table, and between decompositions and the relation $\prec$, which we hope will further clarify the latter.

**Theorem 9:** Given two decompositions of $P$ and $Q$ respectively, with the same set of weights, say $P = \Sigma h \alpha_h P_h$ and $Q = \Sigma h \alpha_h Q_h, (h = 1, \ldots, t)$, if $Q_h \preceq P_h$ for all $h$, then $Q \preceq P$.

**Proof:** For simplicity, let the rows of $P$ be non-null. From $Q_h = S^T_h P_h$, $S_h$ stochastic
\[(h = 1, \ldots, t), \text{ we obtain} \]
\[
Q = \Sigma_h \alpha_h S^T h P_h = \Sigma_h \alpha_h S^T h \Delta P_h P^* = \Sigma_h \alpha_h S^T h \Delta P_h \Delta P^{-1} P
\]
and \((\Sigma_h \alpha_h S^T h \Delta P_h \Delta P^{-1})^T\) is stochastic.||

5. Applications to Measures of Optimal Prediction

Some measures of association have an interpretation in terms of optimal prediction: we can compare the "goodness" of the prediction of \(Y\) — under an optimal strategy—in the two cases (a) that the value of \(X\) is known, (b) without any such information. This approach is outlined in Goodman and Kruskal (1954) for some special cases and has been developed by Haberman (1982a and b) and Rao (1982); it is based on a measure of dispersion \(\Delta\) of a random variable \(Z\), i.e. on a function \(\Delta : S_n \rightarrow \mathbb{R}\), which is assumed non-negative and has the property \(\Delta(p) = 0\) for degenerate distributions \(p\). Measures of dispersion are also called heterogeneity indices or measures of randomness (see Forcina and Giovagnoli, 1982, and Hickey, 1982), and must be Schur-concave functions.

Haberman (1982a and b) and Rao (1982) assume further that \(\Delta\) is concave. Such a \(\Delta\) can be interpreted as the minimum expected loss in predicting \(Z\) under an optimal strategy. The following measure of association ensues:

\[
(5.1) \quad \phi_{Y \cdot X} = \frac{\Delta Y - \Delta Y \cdot X}{\Delta Y},
\]
where \(\Delta Y = \Delta(p_0)\) and \(\Delta Y \cdot X = \sum_{i=1}^m p_i \Delta(p_i^*)\).

Examples:

1. Let \(\Delta(p) = 1 - \max_{1 < j < n} p_j\). Expression (5.1) gives Guttman's (1941) measure, (see Goodman and Kruskal, 1954, p. 741)

\[
\lambda_b(P) = \frac{\sum_{i=1}^n \max_j p_{ij} - \max_j p_{.j}}{1 - \max_j p_{.j}}.
\]

2. Let \(\Delta(p) = 1 - \sum_{j=1}^n p_j^2 = -\sum_{j=1}^n p_j(p_j - 1)\) be Gini's measure of dispersion. The corresponding measure of association is

\[
\tau(P) = \frac{\Sigma_i \Sigma_j p_{ij}^2 / p_i - \Sigma_j p_{j}^2}{1 - \Sigma_j p_{j}^2}
\]
3. From $\Delta(p) = -\sum_{j=1}^{n} p_j \log p_j$ (Shannon's entropy) we obtain a measure of association due to Theil (1970)

$$\eta(P) = \frac{-\sum_i \sum_j p_{ij} \log(p_{ij}/p_i p_j)}{\sum_j p_j \log p_j}.$$ 

Examples (2) and (3) are special cases of a more general class of potential measures, based on an extended definition of entropy. Given a function $f : [0, 1] \to \mathbb{R}$, if for $x \in S_n$, the quantity $H_{n,f}(x) = -\sum_{i=1}^{n} f(x_i)$ is non-negative we call it $f$-entropy (see Burbea and Rao, 1982). Particularly interesting cases are listed in Burbea and Rao (1982): the entropies of degree $\alpha$ due to Havrda and Charvat (1967), the so-called paired entropies of degree $\alpha$, and others. For convex $f$, $H_{n,f}$ is concave and since it is permutation invariant, it is also Schur-concave. If $f(0) = f(1) = 0$ then $H_{n,f} = 0$ for degenerate distributions, and we can use $H_{n,f}$ as $\Delta$ in the definition of an association measure of type (5.1).

We now study the properties of such association measures.

**Theorem 10:** The association measures $\phi_{Y,X}$ defined as in (5.1) are normalized, hence they satisfy (3.4) and (3.5); they preserve the order relation $\preceq$, hence satisfying (3.6). The equivalence properties (3.1), (3.2), (3.3) are also trivially satisfied. For these measures the decomposition axioms (4.1) and (4.2) hold true.

**Proof:** If we let $d(p^*_i, p_0) = 1 - \Delta(p^*_i)/\Delta(p_0)$, $\varphi$ be the identity, we see that the $\phi_{Y,X}$ are of the form (4.3) so (4.2) and hence (4.1) are true. (3.4) and (3.5) follow from the definition, and (3.6) follows from Theorem 8(i) since for concave $\Delta$, $d(p^*_i, p_0)$ is a convex function of $p^*_i$ for given $p_0$. Since (3.6) implies (3.1) and (3.3) only (3.2) remains to be shown, but this is a consequence of the Schur-concavity of $\Delta$. 

However, (3.7) is not always satisfied by association measures of type (5.1). For example let

$$P = \begin{pmatrix} .2 & 0 & 0 \\ .2 & .2 & 0 \\ .2 & .2 & .2 \end{pmatrix}, \quad D = \begin{pmatrix} .5 & .5 & 0 \\ .5 & .5 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad Q = PD = \begin{pmatrix} .1 & .1 & 0 \\ .2 & .2 & 0 \\ .1 & .1 & .2 \end{pmatrix}.$$ 

Then $\lambda_k(P) = 0 < \lambda_k(Q) = 0.16$. We want to investigate under what further conditions on $\Delta$ does $\phi_{Y,X}$ satisfy (3.7). Let $x_1, \ldots, x_m \in S_n$ with associated weights $\pi_1, \ldots, \pi_m, \pi_i \geq 0, \Sigma_i \pi_i = 1$. We introduce the following definition:
Definition: The generalized mutual information measure of $x_1, \ldots, x_m$ relative to the weights $\pi_1, \ldots, \pi_m$ and an f-entropy $H_{n,f}$ is

$$J_{n,f}^\pi(x_1, \ldots, x_m) = H_{n,f}(\sum_{i=1}^{m} \pi_i x_i) - \sum_{i=1}^{m} \pi_i H_{n,f}(x_i).$$

The following result is due to Burbea and Rao (1982):

Lemma 2: Let $f$ be a $C^2$ function on an interval $I$. Then $J_{n,f}^\pi(x_1, \ldots, x_m)$ is convex (concave) on $I^n \times I^n \times \cdots \times I^n$ if and only if $f$ is convex (concave) and $1/f''$ is concave (convex) on $I$.

Given a table $P$, if we choose as a measure of dispersion a suitable f-entropy $H_{n,f}$, a corresponding measure of association based on optimal prediction will be

$$\phi(P) = \frac{J_{n,f}^{p^*}(p_1^*, \ldots, p_m^*)}{H_{n,f}(p_0^*)} . \tag{5.2}$$

Theorem 11: Let $\phi(P)$ be a measure of association of type (5.2) with $f$ a $C^2$ function. If $f$ is convex and $1/f''$ is concave, then $\phi(P)$ satisfies (3.7).

Proof: Suppose $Q \prec P$. Then $Q = PD$ with $D$ a T-matrix and $p_r = q_r, p_0D = q_0$.

We shall prove that $\phi(Q) \leq \phi(P)$ in a slightly more general case, namely for any $D$ of the form $D = \Sigma_k \alpha_k \Pi_h$, with $\alpha_k \geq 0, \Sigma \alpha_k = 1$ and $\Pi_h$ permutation matrices. If $f$ is convex, $H_{n,f}$ is Schur-concave; it follows that $H_{n,f}(p_0) \leq H_{n,f}(q_0)$. Also $q_i^* = p_iD, 1 \leq i \leq m$ and the function $J_{n,f}^{p^*}(p_1^*, \ldots, p_m^*)$ is invariant for the same permutation of the coordinates of the vectors $p_1^*, \ldots, p_m^*$. Moreover if $f$ is convex and $1/f''$ is concave, because of Lemma 2

$$J_{n,f}^{q^*}(q_1^*, \ldots, q_m^*) = J_{n,f}^{p^*}(\Sigma \alpha_k(p_1^* \Pi_h, \ldots, p_m^* \Pi_h))$$

$$\leq \Sigma \alpha_k J_{n,f}^{p^*}(p_1^* \Pi_h, \ldots, p_m^* \Pi_h) = (\Sigma \alpha_k) J_{n,f}^{p^*}(p_1^*, \ldots, p_m^*)$$

and hence $\phi(P) \geq \phi(Q)$.

Corollary: (3.7) holds for the measures of association $\tau$ and $\eta$ of examples (2) and (3).

Proof: Put $f(x) = x^2 - x$ in the first case and $f(x) = x \log x$ in the second.

6. Applications to Measures of Mean Relative Information

Intuitively, the degree of association in a table $P$ could be assessed by the difference between each conditional distribution of $Y|X = x_i$ ($i = 1, \ldots, m$) and the marginal distribution
of \( Y \). This approach, suggested by Gini (1914-15, p. 924), is discussed by Fréchet (1947) and Goodman and Kruskal (1959, p. 145).

In general let \( I(u; v) \) be a function of two row vectors \( u, v \in S_n \) that can be interpreted as a distance or a measure of divergence between \( u \) and \( v \), and let \( g : \mathbb{R} \to \mathbb{R} \), continuous and strictly increasing, be chosen in order to provide an appropriate general mean value (in the sense of Kolmogorov-Nagumo, see Rényi, 1961, p. 554) of the quantities \( I(p_i^*; p_0) \) with weights \( p_i \) \((i = 1, \ldots, m)\). A general expression for a class of measures of association is

(6.1)

\[
\phi(P) = g^{-1}\left(\sum_{i=1}^{m} p_i g(I(p_i^*; p_0))\right)
\]

In general \( I(u; v) \) need not be symmetric or satisfy the triangular inequality. Its properties can be conveniently discussed in the light of the concept of relative information due to Rényi (1961). Let \( u \) be the distribution of a random variable \( \xi \) conditionally on an event \( E \) and let \( v \) be the unconditional distribution of \( \xi \). Then we may define a function \( I(u; v) \) that measures the amount of information about \( \xi \) that is obtained by observing the event \( E \).

Two obvious requirements for this function are:

(6.2a) \( I(u; v) \geq 0 \) and \( I(u; u) = 0 \) for all \( u, v \in S_n \).

(6.2b) \( I(u\Pi; v\Pi) = I(u; v) \) for any permutation matrix \( \Pi \).

The following properties also deserve some consideration:

(6.2c) \( I(u; v) \) is a convex function of \( u \) for any fixed \( v \).

(6.2d) \( I(uD; vD) \leq I(u; v) \) for any \( n \times n \) T-matrix \( D \).

(6.2e) \( I(u; v) \) is convex on \( S_n \times S_n \).

Obviously (6.2d) implies (6.2b); (6.2e) implies (6.2c), and (6.2b) and (6.2e) imply (6.2d).

Let \( E = \bigcup_{h=1}^{t} E_h \) with \( \alpha_h = P(E_h) \) \((h = 1, \ldots, t)\); then (6.2c) is equivalent to the statement that the information gained by the knowledge that \( E \) has occurred cannot exceed the expectation of the information gain due to the knowledge that event \( E_h \) has occurred. Inequality (6.2d) can be seen as a weak extension of (6.2b) in the direction of (6.2e): (6.2e) says that the gain in information of \( \lambda u_1 + (1 - \lambda)u_2 \) relative to \( \lambda v_1 + (1 - \lambda)v_2 \), \( \lambda \in [0, 1] \) cannot exceed the expected gain in information of \( u_1 \) relative to \( v_1 \) and \( u_2 \) relative to \( v_2 \).
with probabilities \( \lambda \) and \((1 - \lambda)\); (6.2d) says only that this inequality must hold if there is a permutation matrix \( \Pi \) such that \( u_2 = u_1 \Pi \) and \( v_2 = v_1 \Pi \). In particular, for \( v = n^{-1}e^T \), (6.2d) implies the Schur-convexity of \( I(\cdot; n^{-1}e^T) \) with respect to the first vector.

Note that though the measures of association discussed in Section 5 can be seen as the expected value of \((\Delta(p_0) - \Delta(p_1))/\Delta(p_0)\), the quantity \(1 - \Delta(u)/\Delta(v)\) cannot be interpreted as a measure of relative information since it is certainly negative for some choices of \( u \) and \( v \).

Various types of functions have been considered in the literature more or less explicitly as measures of relative information. The following are the best-known examples

\[(I) \quad I_r(u; v) = \left(\sum_{i=1}^{n} |u_i - v_i|^r \right)^{1/r}, \quad r \geq 1.\]

For \( r = 1 \), \( I_1(u; v) \) is proportional to Gini's measure of dissimilarity (Gini, 1914-15).

\[(II) \quad I_{R, \alpha}(u|v) = (\alpha - 1)^{-1} \log \left[ \sum_{j=1}^{n} u_j (u_j/v_j)^{\alpha - 1} \right], \quad \alpha > 1,\]

\[I_{R,1}(u|v) = \lim_{\alpha \to 1} I_{R, \alpha}(u|v) = \sum_{j} u_j \log(u_j/v_j).\]

These are Rényi's measures of information of order \( \alpha \). \( I_{R,1}(u|v) \) is also called the Kullback-Liebler distance of \( u \) and \( v \).

(III) \( I_f(u\|v) = \sum_{j=1}^{n} v_j f(u_j/v_j) \), with \( f(\cdot) \) convex and strictly convex at 1, \( f(1) = 0 \). These measures are studied in Csiszár (1972) and are called \( f \)-divergences.

\[(IV) \quad J_n,f(u, v) = \sum_{j=1}^{n} \left[ \frac{1}{2} f(u_j) + f(v_j) - f\left(\frac{1}{2}(u_j + v_j)\right) \right] \]

with \( f \) convex, \( 1/f'' \) concave. This is a special case of the generalized mutual information measures defined in Section 5 (see Burbea and Rao, 1982). Note that \( I_{R, \alpha}(u|v) = [\log I_f(u\|v)]/(\alpha - 1) \) with \( f(x) = x^\alpha \) for \( \alpha > 1 \) and \( I_{R,1}(u|v) = I_f(u\|v) \) with \( f(x) = x \log x \).

Well-known examples of measures of association of the form (6.1) are listed below:

\[(1) \quad \phi(P) = \{\Sigma_i p_i \Sigma_j |p_{ij} - p_{.j}|^r\}^{1/r}.\]

This is an extension of Gini's index of association obtained by putting \( I(\cdot; \cdot) = I_r \) and \( g(x) = x^r \).
(see for example Leti, 1983);

$$\phi(P) = \frac{1}{\alpha - 1} \log \left\{ \Sigma_{ij} \frac{p_{ij}^\alpha}{(p_i \cdot p_j)^{\alpha-1}} \right\} = \log(J(\alpha-1) + 1), \quad \alpha > 1,$$

where $J_{\alpha}, \lambda > 0$, are association measures described in Good (1981). For $\lambda = 1$ we obtain the well-known expression for $\chi^2$. If we let $I(\cdot; \cdot) = I_{R,\alpha}$, $g(x) = \exp x(\alpha - 1)$, it can be easily verified that this is of the form (6.1);

$$\phi(P) = \Sigma_{ij} \ p_{ij} \log p_{ij} - \Sigma_{ij} \ (p_i \cdot p_j) \log(p_i \cdot p_j).$$

This measure, which apparently was first introduced by Halphen (1957) and is denoted by $I$ in Good (1981), can be interpreted as a descriptive analogue of the log-likelihood ratio test of independence in a two-way table. It reduces to the form (6.1) with $I(\cdot; \cdot) = I_{R,1}$, $g(x) = x$.

We now establish which of the properties (6.2) are satisfied by the measures of information listed as (I), (II), (III), (IV), with a view to analyzing the above association measures in the light of the order and equivalence principles of Section 3.

Lemma 3: Properties (6.2a), (6.2b), (6.2e) (and thus also (6.2c) and (6.2d) are satisfied by

$$I_r(u; v), \ I_{R,1}(u|v), \ I_f(u||v) \text{ and } J_{n,f}(u, v).$$

Proof: (6.2a) is obvious for $I_r$ and is well-known for $I_{R,1}$ (see Rényi, 1961) whereas for $I_f$ (see Csiszar, 1972, p. 194) and for $J_{n,f}$ it follows from the convexity of $f(\cdot)$. (6.2b) is trivially true in all cases. As regards (6.2e), for $I_r$ this property follows from the fact that $I_r, r \geq 1$, is a convex function of $|u - v|$ because of Minkowski’s inequality; for $I_f$ it has been demonstrated by Csiszar (1972); for $J_{n,f}$ it follows from Lemma 2 of Section 5 and for $I_{R,1}$ it follows from the convexity of the function $x \log(x/y)$.

Lemma 4: $I_{R,\alpha}(u|v), \alpha > 1,$ satisfies (6.2a), (6.2b) and (6.2d).

Proof: (6.2b) is obvious. It is easily shown that the function $x^\alpha/y^{\alpha-1}, \text{ with } \alpha > 1,$ is convex on $\{x \geq 0, y > 0\}$. Thus $\exp\{I_{R,\alpha}(u|v)\}$ is convex on $S_n \times S_n$ and this implies (6.2d). From $I_{R,\alpha}(u|v) \geq I_{R,\alpha}(uD|vD)$, with $D = n^{-1}(ee^T)$, it follows that (6.2a) holds.

We can now establish some general properties of measures of association of the form (6.1).

Theorem 12: Let $\phi$ be a measure of association defined as in (6.1), then
(a) \( \phi \) satisfies (3.1), (3.3), (4.2);

(b) from property 6.2a of \( I(\cdot;\cdot) \) it follows that \( \phi \) is \( \geq 0 \) and satisfies (3.4); (6.2b) yields (3.2);

(c) if (6.2c) holds for \( I(\cdot;\cdot) \) and \( g \) is convex, then \( \phi \) satisfies (3.6);

(d) if (6.2d) holds for \( I(\cdot;\cdot) \), then \( \phi \) satisfies (3.7).

**Proof:** Because \( \phi \) is of the form (4.3) defined in Section 4, (4.2) holds and (3.1) and (3.3) follow from Proposition 1. If \( I(u;v) \geq I(u;u) = 0 \) for all \( u, v \), since for \( H \in \mathcal{X}, p_i^* = p_0 \) for all \( i \) then \( \phi(H) = 0 \) and \( \phi(P) \geq 0 \) for all \( P \). The fact that (3.2) follows from (6.2b) is obvious. Statements (c) and (d) follow from Theorem 8.||

**Corollary:** The measures of association given in the examples (1), (2), (3) above satisfy the requirements of principles (3.1), (3.2), (3.3), (4.2), (3.4), (3.6), (3.7).

**Proof:** Since it has been shown that these measures are of the form (6.1), (3.1), (3.3) and (4.2) follow from Theorem 12. For the measures of examples (1) and (3) the proof of (3.6) and (3.7) is obtained by combining Theorem 12 with the results of Lemma 3. Property (3.6) for example (2) follows from Theorem 8 if we put \( d(u,v) = \sum_i u_i^2 / v_i^{q-1} \), which is convex on \( S_n \times S_n \). Property (3.7) for example (2) follows from Lemma 4 and Theorem 12; (3.2) is a consequence of (3.7).||

**Acknowledgments**

We wish to thank Professor Henry Wynn for providing some very useful bibliographical material, and Professor Ingram Olkin for interesting conversations, and many comments.
REFERENCES


