Stochastic Inequalities
IMS Lecture Notes – Monograph Series
Volume 22 (1993)

SOME APPLICATIONS OF MONOTONE TRANSFORMATIONS IN STATISTICS[1]

By ALLAN R. SAMPSON

University of Pittsburgh

A number of results concerning monotone transformations of random variables are reviewed. Particular attention is paid to the effects of choice of monotone scaling in two settings: (a) describing and quantifying dependence between two random variables, and (b) comparing two populations with ordinal categorical responses.

Properties of the concordant and discordant monotone correlation coefficients (Kimeldorf, May and Sampson (1982)) between random variables X and Y are discussed, and computational approaches are considered.

The two sample problem is explored where responses are ordinal categories and typical statistical procedures involve the arbitrary choice of monotone scales. The effects of the choice of scaling upon the resultant analyses are examined in detail.

1. Introduction

In a variety of situations, it is of interest to consider how the results of the analyses change when we transform the relevant random variables by monotone functions. The usual purpose of this is to study the effects of monotonically rescaling the measured random quantities. Depending on our needs, we might want a statistical procedure that is invariant to monotone scale changes, or we might want to choose an appropriate scaling in situations where the natural choice of scales is not clear. The purpose of this paper is to review some results in this area focusing on a somewhat less than standard usage of monotone transformations.

Traditional concerns about monotone invariance can lead to various notions. In some settings it leads to considering statistical procedures which depend only on ranks of the data. For jointly distributed random variables, it can lead to a discussion of procedures which depend solely on the copula

[1]Research supported by National Security Agency Grant No. MDA–904–90–H–4036. Reproduction in whole or part is permitted for any purpose of the United States Government.

AMS 1991 *subject classifications.* Primary 62F03, 62H02; Secondary 62A05.

Key words and phrases. Scaling, monotone dependence, two–sample, ordinal variables, contingency table, concordant monotone correlation, correspondence analysis.

(Sklar (1959)) or uniform representation (Kimeldorf and Sampson (1975)), or in the case of data, upon the multivariate empirical rank distribution (e.g., Block, Chhetry, Fang and Sampson (1990)). For ordinal contingency tables, invariance leads to other related notions.

The focus of this paper is somewhat different than these preceding traditional concerns about invariance. We are interested in describing how certain probabilistic concepts and statistical notions depend on the choice of monotone scales, and utilizing this knowledge for assessing appropriateness of scales. This idea of sensitivity to scales and rescaling is particularly important in statistical usages where there is no natural choice of scales. Ordinal contingency tables offer such an example where, for instance, one variable might be an evaluative response such as excellent, very good, good, etc., and the other is degree of involvement such as none, some, etc. In Section 2, we discuss the effects of monotone scaling on certain measures of dependence and in Section 3 we consider how various two–sample tests are affected by choice of monotone scores.

2. Measures of Dependence and Scaling for Bivariate Random Variables

Lancaster (1969) interweaves several lines of research to present a set of techniques concerning bivariate random variables X and Y which describe their structure and measure their degree of relationship. Many of these results rely on the canonical decomposition of a bivariate p.d.f. See also Kendall and Stuart (1979, Chapter 33). We review some relevant definitions below.

DEFINITION 1

(i) Random variables X and Y are *mutually completely dependent (MCD)* if there exists a one–to–one function γ so that $Y = \gamma(X)$ w.p. 1.

(ii) The *sup–correlation* between random variables X and Y, denoted by $\rho'(X,Y)$, is defined as $\sup \rho(f(X), g(Y))$, taken over all suitable functions f and g.

(iii) The support $S \times T$ of (X,Y) is said to consist of (at least) k *disjunct pieces* if there exists partitions $S_1, \ldots, S_k$ of S and $T_1, \ldots, T_k$ of T such that

$$P((X,Y) \in S_i \times T_i) > 0, \quad i = 1, \ldots, k$$

and

$$P((X,Y) \in S_i \times T_j) = 0 \qquad \text{for all } i \neq j.$$

Intuitively MCD was thought to be an antithesis of independence and the sup–correlation was a method of measuring how dependent random variables

X and Y are. Clearly, X and Y are independent if and only if $\rho'(X,Y) = 0$; and if X and Y are MCD, then $\rho'(X,Y) = 1$. However, the converse to the latter is not true. If $\rho'(X,Y) = 1$, then under suitable regularity conditions, it can be shown that the support of X, Y consists of at least two or more disjunct pieces.

Lancaster (1969) discusses in great detail the above notions as well as many other related ones. The sup–correlation has been discussed in further detail as a measure of dependence, particularly for contingency tables when it is readily computable. Additionally, these notions form, in part, the basis for correspondence analysis. More recently, the notion of the canonical expansion of a joint p.m.f. has been explored statistically by Gilula (1984).

Kimeldorf and Sampson (1978) and others, including Vitale (1990), have shown that MCD is not an appropriate antithesis to independence. In fact, there exists $\{X_n, Y_n\}$ all with the same respective univariate marginals such that X_n, Y_n are MCD for each n and yet, X_n, Y_n converge in distribution to independent random variables X, Y. Obviously, for large enough n, if one were to take a random sample of such X_n, Y_n, this sample for all intents and purposes would look like a sample from independent random variables.

To counter this difficulty Kimeldorf and Sampson (1978) introduce the notion of X and Y being *monotone dependent* if in Definition 1(i), γ is a monotone function. (If γ is increasing we say X and Y are *increasing dependent*, and similarly for γ decreasing). Moreover, they show that if X_n, Y_n are monotone dependent for each n, and X_n, Y_n converge in distribution to X, Y, then X, Y are monotone dependent. This, and other reasons, suggest that monotone dependence serves as a suitable antithesis to independence.

To measure the degree of monotone dependence between a pair of random variables X and Y, Kimeldorf and Sampson (1978) introduce the notion of monotone correlation $\rho^*(X,Y)$, where f, g are required in Definition 1(ii) to be monotone functions. Additionally, Kimeldorf and Sampson (1978) observe the following straightforward properties: (a) $\rho^*(X,Y) = 0$, if and only if X and Y are independent; (b) $|\rho(X,Y)| \leq \rho^*(X,Y) \leq \rho'(X,Y)$; and (c) for the bivariate normal, the inequalites in (b) become equalities.

The monotone correlation is further refined by Kimeldorf, May and Sampson (1982). If in Definition 1(ii), f and g are both required to be increasing (or both decreasing) the resulting measure of dependence is called the *concordant monotone correlation (CMC)*. Motivated to study the monotone correlation notion when f is increasing and g is decreasing (or equivalently, vice versa) Kimeldorf, May and Sampson (1982) observe that

$$\sup_{f\uparrow, g\downarrow} \rho(f(X), \quad g(Y)) = \sup_{f\uparrow, g\uparrow} \rho(f(X), \quad -g(Y) = - \inf_{f\uparrow, g\uparrow} \rho(f(X), g(Y)).$$

This observation leads them to define the *discordant monotone correlation (DMC)* by $\inf \rho(f(X), g(Y))$, taken over all f, g both increasing or both decreasing. Clearly there is the following relationship: $\rho^*(X,Y) = \max(CMC(X,Y), -DMC(X,Y))$.

Kimeldorf, May and Sampson (1982) also introduce the *iso-CMC (ICMC)* and *iso-DMC (IDMC)* when f is restricted to equal to g in the preceding definitions of the CMC and DMC, respectively. (Interestingly they show that if X and Y have an exchangeable distribution, it is *not* necessarily true that $ICMC(X,Y) = CMC(X,Y)$.)

While both the CMC and DMC are measures of association, respectively, for increasing monotone dependence and decreasing monotone dependence, they serve another very useful purpose: for any increasing functions f, g

$$DMC(X,Y) \leq \rho(f(X), g(Y)) \leq CMC(X,Y). \tag{1}$$

The implications of (1) will be discussed shortly.

When X and Y are jointly discrete random variables with finite support, a number of additional useful results can be obtained. If in Definition 1(iii), $S \leq S_2 \leq \cdots \leq S_k$ and $T_1 \leq (\geq) T_2 \leq (\geq) \cdots \leq (\geq) T_k$, where $U \leq V$ means for all $u \in U$, $v \in V$, $u \leq v$, we say the support consists of *increasing (decreasing) disjunct pieces.* (Obviously, this definition extends beyond the discrete case.) Chhetry, DeLeeuw and Sampson (1990) show that the $CMC = 1$, if and only if the support of X and Y consists of two or more increasing disjunct pieces (and equivalently $DMC = -1$, if and only if the support consists of two or more decreasing disjunct pieces). To compare $\rho'(X,Y)$ and $CMC(X,Y)$, we recall Lehmann's (1966) notion of Y being *positively regressive dependent (PRD)* on X if $P(Y > y \mid X = x)$ is nondecreasing in x for all y. Schriever (1983) showed that if Y is PRD on X and X is PRD on Y, then $\rho'(X,Y) = CMC(X,Y)$. However, Chhetry and Sampson (1987) give an example where X and Y are not mutually PRD, yet $\rho'(X,Y) = CMC(X,Y)$. Lastly, under the assumption of finite discrete support (with the number of support points of X being three or more, or of $Y \geq 3$) we have $CMC(X,Y) = DMC(X,Y)$, if and only if X and Y are independent random variables.

To explore the uses of the inequalites in (1) for arbitrary X and Y, we begin by noting that various measures of dependence are defined as correlations between certain increasing functions. Spearman's rho, ρ_s, can be defined by $\rho(F(X), G(Y))$, where F and G are, respectively, the marginal c.d.f.'s of X and of Y. Another possible measure is given by $\rho(\Phi^{-1}F(X), \Phi^{-1}G(Y))$, where Φ is the standard normal c.d.f. Let $m(X,Y)$ generically denote a measure of monotone dependence between X and Y of the form $\rho(\phi(X), \psi(Y))$, where ϕ, ψ are particular increasing functions. Such measures are discussed by Agresti (1984) and Williams (1952). Then from (1), any such $m(X,Y)$ must lie in the interval $[DMC(X,Y), CMC(X,Y)]$. If in addition we require $\phi = \psi$, denoting the resulting measure by $m_I(X,Y)$, then $IDMC \leq m_I \leq$

$ICMC$. Now consider a situation where we are not certain which measure of monotone dependence is most meaningful for a problem being analyzed. If the CMC is relatively close to the DMC, the problem is rendered moot, in that the inequality of (1) yields that *all* measures of monotone dependence must be close to each other.

This observation is more pertinent when measuring dependence for a bivariate ordinal contingency table (or equivalently dealing with bivariate discrete random variables with finite support). Here there is an extensive literature for measures of positive dependence (e.g., Agresti (1984) or Schriever (1985)), with a number in the form of a measure of monotone dependence. In this situation, oftentimes appropriate scalings for the ordinal X and Y variables are not available. Thus, having the knowledge that the CMC and DMC are close indicates little sensitivity of a measure of monotone dependence to the choice of scalings.

To illustrate these notions, we consider in Table 1 the "father–son British social mobility data" (Glass and Hall (1954)). In this case since the variable being measured for both father and for son is the same occupational status variable which is ordinal in nature, it is appropriate to use *iso–scaling*, (that is, requiring $\phi = \psi$). In this table Status S1 is professional, and high administrative; Status S2 is managerial, executive and higher grade supervisory; Status S3 is lower grade supervisory; Status S4 is skilled manual; and Status S5 is semi–skilled and unskilled manual. A son in Status j whose father is in Status i is said to be upwardly mobile if $j < i$ (e.g., Bishop, Fienberg, and Holland (1975, p. 321)).

Father's Occupational Status	Son's Occupational Status				
	S1	S2	S3	S4	S5
S1	50	45	8	18	8
S2	28	174	84	154	55
S3	11	78	110	223	96
S4	14	150	185	714	447
S5	3	42	72	320	411

Table 1. British mobility data (3,500 father–son data values). (Glass and Hall (1954))

In this case the $ICMC = .496$ and $IDMC = .242$, (see Kimeldorf, May and Sampson (1982)) indicating that regardless of the monotone scaling for these five ordinal categories, the resulting correlation is between .242 and .496. (The monotone scales corresponding to the $ICMC$ assign S5, S4, S3, S2 and S1, respectively, the values 0, .077, .158, .373, and 1. The monotone scales corresponding to the $IDMC$ assign to S5, S4, S3, S2, and S1, respectively, 0, 1, 1, 1, and 1.)

Another focus of this development is examining and interpreting the resulting monotone scales for the CMC. In dual scaling (Nishisato (1980)) or in correspondence analysis (Benzecri (1973)), the scalings derived for contingency tables are those which maximize $\rho(a(X), b(Y))$ over all a, b, whether or not they are monotone. Obviously these $a(\ )$ and $b(\ )$ are the functions which yield the sup–correlation between the row and column classifications. If the classifications are categorical, these resulting scalings have standard interpretations. However, when the classifications are ordinal, a natural requirement is that the scales be monotone. There is no guarantee for this in the usual dual scalings, whereas, our scales obviously guarantee monotonicity.

Using a random generation procedure described in their paper, Kimeldorf, May and Sampson (1982) generated the following random 10×10 probability matrix with a slight amount of positive dependence.

		Y									
		b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
	a_1	0.0331	0.0111	0.0092	0.0049	0.0016	0.0028	0.0009	0.0108	0.0096	0.0007
	a_2	0.0101	0.0361	0.0057	0.0081	0.0133	0.0062	0.0121	0.0066	0.0003	0.0020
	a_3	0.0102	0.0059	0.0347	0.0027	0.0055	0.0020	0.0104	0.0046	0.0069	0.0056
	a_4	0.0144	0.0018	0.0065	0.0342	0.0006	0.0071	0.0055	0.0066	0.0084	0.0113
X	a_5	0.0006	0.0016	0.0087	0.0132	0.0435	0.0061	0.0100	0.0046	0.0044	0.0053
	a_6	0.0022	0.0035	0.0151	0.0015	0.0056	0.0427	0.0062	0.0035	0.0089	0.0125
	a_7	0.0002	0.0084	0.0026	0.0020	0.0005	0.0086	0.0387	0.0007	0.0034	0.0111
	a_8	0.0084	0.0100	0.0079	0.0036	0.0100	0.0128	0.0044	0.0303	0.0121	0.0065
	a_9	0.0028	0.0079	0.0141	0.0008	0.0133	0.0077	0.0064	0.0139	0.0402	0.0068
	a_{10}	0.0009	0.0149	0.0042	0.0108	0.0022	0.0144	0.0130	0.0151	0.0146	0.0438

Table 2. Random 10×10 probability matrix.
(Kimeldorf, May and Sampson (1982))

For this joint distribution the $CMC = .443$ and the monotone scales are given in Table 3.

	1	2	3	4	5	6	7	8	9	10
X	0.	.461	.461	.461	.872	.872	.872	.872	.873	1.
Y	0.	.537	.541	.541	.842	.842	.842	.842	.842	1.

Table 3. Monotone Scales for Table 2.

It is interesting to observe that for the X variable, there are only 5 distinct scores for the 10 ordinal categories and for the Y variable, the same is true. This suggests that for cross–prediction purposes the appropriately collapsed 5×5 distribution with the noted monotone scales allows for the best linear predictability. Although this approach to obtaining monotone scales requires further development, it appears to provide a notion of dual scalings for ordinal contingency tables or may form a basis for "ordinal correspondence analysis."

For bivariate discrete distributions on an $m \times n$ lattice $\{u_1, \ldots, u_m\} \times \{v_1, \ldots, v_n\}$, the dual scalings can be computed from a spectral decomposition of the matrix $Q^* = D_r^{-1/2} Q D_c^{-1/2}$, where $Q = \{q_{ij}\} \equiv \{\text{Prob}(X = u_i, Y = v_j)\}$, $D_r = \text{Diag}(q_{1+}, \ldots, q_{m+})$ and $D_c = \text{Diag}(q_{+1}, \ldots, q_{+n})$. For instance, the sup–correlation is the square root of the second largest eigenvalue of $Q^* Q^{*\prime}$. On the other hand, to compute the CMC and the corresponding monotone scores is a more difficult computational problem. An analogy of this increased difficulty, is the comparative difficulty of the following two optimization problems.

$$\sup_{x'x=1} x'Sx \tag{2}$$

and

$$\sup_{\substack{x'x=1 \\ x_1 \leq \cdots \leq x_p}} x'Sx, \tag{3}$$

where S is a symmetric $p \times p$ matrix.

Kimeldorf, May and Sampson (1982) express the monotone correlation problem as a nonlinear programming problem with linear constraints and then employ a numerical optimization algorithm of May (1979). The resulting software package, called MONCOR, is described by Kimeldorf, May and Sampson (1981) and is available from the author (written in FORTRAN and requiring IMSL routines.)

When dealing with continuous bivariate random variables, the problems are obviously compounded. To compute the sup–correlation and corresponding canonical variables requires solving a continuous eigenfunction problem, although for some bivariate distributions, certain classical bivariate expansions yield these. We are unaware of any technique like these to bring to bear for the CMC and the corresponding monotone scalings.

Based upon a random sample from a bivariate continuous distribution, the ACE algorithm of Breiman and Friedman (1985) can be applied to estimate the sup–correlation and corresponding canonical variables. There appear to be monotonicity constraints within the ACE algorithm which would allow the estimation of the monotone correlation and corresponding monotone scalings.

No explicit sampling results are applicable to the CMC and DMC, even in the case of multinomial sampling. Perhaps bootstrapping may be effective in some of these cases. An asymptotic distribution for the sample sup–correlation has been obtained by Sethuraman (1990).

No extensions of monotone dependence notions to three or more dimensions are available.

3. Two–Sample Ordinal Data

Suppose we have random samples from two populations or treatments where each observation falls into one of k levels of an ordinal categorization. There are a variety of standard statistical procedures for comparing the two populations or treatments based upon these data.

An example of this type of problem is in clinical trials, where there is an experimental procedure (E) and a control (C), and the evaluations of each patient are the physician's global rating. These ratings are typically 5–point or 7–point scales, e.g., very improved, moderately improved, no change, moderately deteriorated, and very deteriorated. The standard analysis procedure typically involves: scoring the responses, on an equal–spaced scale, or using rank scores, or even dichotomizing the response. The interpretative difficulty is that the scorings are in some sense arbitrary, as would be any dichotomization.

More specifically, let $L_1 < \cdots < L_k$ denote the ordinal categorical levels' labels where "<" denotes the underlying experimental order. The arbitrariness in many of these two sample procedures comes from the choice of increasing scores $x_1 \leq \cdots \leq x_k$ ($x_1 \neq x_k$) that one can assign to the respective levels $L_1, \ldots, L_k$. Among the standardly used scoring systems are: (i) $1, \ldots, k$, (ii) $R_1, \ldots, R_k$, where R_i is the marginal mid–rank score for level i; (iii) ridit scores (and modified ridits), and (iv) $0, \ldots, 0, 1, \ldots, 1$ which proves a dichotomization of the levels.

As Kimeldorf, Sampson and Whitaker (1992), hereafter denoted by KSW, note, the commonly used procedures: (a) Wilcoxon–Mann–Whitney, (b) χ^2 on the dichotomization, (c) scored two sample t–test, (d) Cochran–Armitage test, and (e) appropriate log–linear models all share a common feature. For all sample sizes (or asymptotically in case (e)), the resultant test statistics are monotonically increasing functions of a certain correlation $r(x_1, \ldots, x_k)$. Denote the resulting data in the form:

	Levels				
	L_1	L_2	$\cdots$	L_k	Total
C	m_1	m_2	$\cdots$	m_k	m
E	n_1	n_2	$\cdots$	n_k	n
Total	$m_1 + n_1$	$m_2 + n_2$	$\cdots$	$m_k + n_k$	N

Then $r(x_1, \ldots, x_k)$ is the Pearson correlation coefficient based on the scores $x_1, \ldots, x_k$ and the values 0 assigned to C, and 1 to E.

The issue as discussed by KSW (1992) is how does the choice of scores effect $r(x_1, \ldots, x_k)$ and, consequently, the noted standardly employed procedures.

The approach taken by KSW (1992) is to find

$$r_{\mathrm{MAX}} = \max_{\substack{x_1 \leq \cdots \leq x_k \\ x_1 \neq x_k}} r(x_1, \ldots, x_k)$$

and

$$r_{\mathrm{MIN}} = \min_{\substack{x_1 \leq \cdots \leq x_k \\ x_1 \neq x_k}} r(x_1, \ldots, x_k)$$

and, thus, by monotonicity obtain the max and min of any of the related test statistics.

For example, suppose that our goal is to test $H_0 : E = C$ versus the alternative that E produces "larger" values than C, and we plan to use a scored t–test and the t distribution as the approximate null hypothesis distribution. If the resulting minimum t–statistic, t_{MIN}, is greater than t^{α}, the appropriate one–sided α–level critical value, then *all* scoring systems produce a statistically siginificant result. Similarly if $t_{\mathrm{MAX}} < t^{\alpha}$, then *no* scoring system can produce a significant result. If we face the situation $t_{\mathrm{MIN}} < t^{\alpha} < t_{\mathrm{MAX}}$, which we call the "straddling" case, then the results depend on the choice of scoring system. In this case, much care must be taken in the choice of scales and in justifying them.

The computational approach of MONCOR could be used to compute r_{MIN} and r_{MAX}. However, an analytical solution is possible in this setting. We now briefly describe our approach to computing r_{MIN} and r_{MAX}, noting that since r is location and scale invariant, we can, if convenient, assume that $x_1 = 0$ and $x_k = 1$.

RESULT 1 *The empirical distribution for treatment E is stochastically greater (smaller) than that of C, if and only if* $r(X_1, \ldots, X_k) \geq 0$ *(*≤ 0*) for all possible* $x_1 \leq \cdots \leq x_k$ *(*$x_1 \neq x_k$*).*

The applications of Result 1 are immediate. If treatments E and C are not stochastically comparable, then there exist scores, so that the one–sided or two–sided t–test, etc. will *not* reject $H_0 : E = C$. In fact, in this situation there exist scorings which yield both *positive* and *negative* t–values.

RESULT 2A *If E is stochastically incomparable to C, then* r_{MAX} *occurs at the scores* $y_1^* \leq \cdots \leq y_k^*$ *that minimize*

$$\sum_{i=1}^{k} (m_i + n_i)\{n_i/(m_i + n_i) - y_i\}^2 \tag{4}$$

among $y_1 \leq \cdots \leq y_k$*; and* r_{MIN} *occurs at the scores* $z_1^* \leq \cdots \leq z_k^*$ *that minimize*

$$\sum_{i=1}^{k} (m_i + n_i)\{m_i/(m_i + n_i) - y_i\}^2 \tag{5}$$

among $y_1 \leq \cdots \leq y_k$.

RESULT 2B *If E is stochastically greater than C,* r_{MAX} *occurs at the* $y_1^* \leq \cdots \leq y_k^*$ *given in (2A) and* r_{MIN} *occurs at one of the* $k-1$ *monotone extreme points, namely,* $(0,1,\ldots,1),(0,0,1,\ldots,1),\ldots,(0,\ldots,0,1)$.

RESULT 2C *If E is stochastically smaller than C, then* r_{MAX} *occurs at a monotone extreme point and* r_{MIN} *occurs at the* $z_1^* \leq \cdots \leq z_k^*$ *given in (2A).*

The solutions to (4) and (5) can be obtained, respectively, from the isotonic regression of $n_i/(m_i+n_i)$ and of $m_i/(m_i+n_i)$, both with weights (m_i+n_i). Robertson, Wright and Dykstra (1988) give a variety of algorithms to compute these isotonic regressions, including the simple Pool–Adjacent–Violators–Algorithm (PAVA). KSW (1992) illustrate the application of the PAVA technique to solve a number of examples. For moderate k, it is straightforward to directly compute r_{MIN} and r_{MAX}.

KSW (1992) provide proofs for Result 1 and Result 2. They also discuss the further interpretation of these results in data analysis but no distributional results have been obtained for the statistics: r_{MIN} and r_{MAX}. Further research on this class of problems is considered by Gautam (1991) in his dissertation.

4. Discussion

For both the problems, correlation between ordinal variables and testing two ordinal populations, we have considered the effects due to arbitrary monotone scorings. In each case we obtain min and max bounds on the appropriate statistics. The correlation problem requires extensive computation and the two–sample problem is solved simply.

Further discussion of the problematic "straddling" case for the two sample problem is given by KSW (1992). The extension of the two sample case to the one–way analysis of variance and the multivariate setting are being established by Guatam, Kimeldorf and Sampson.

Acknowledgement

The author wishes to thank both referees for their helpful comments.

REFERENCES

AGRESTI, A. (1984). *Analysis of Ordinal Categorical Data.* Wiley, New York.

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

BISHOP, Y., FIENBERG, S. E. AND HOLLAND, P. (1975). *Discrete Multivariate Analysis: Theory and Practice.* MIT Press, Cambridge, MA.

BLOCK, H. W., CHHETRY, D., FANG, Z., AND SAMPSON, A. R. (1990). Partial orderings on permutations and dependence orderings on bivariate empirical distributions. *Ann. Statist.* **18** 1840–1850.

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

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

CHHETRY, D., DE LEUUW, J. AND SAMPSON, A. R. (1990). Monotone correlation and monotone disjunct pieces. *SIAM J. Matrix Anal. Appl.* **11** 361–368.

GAUTAM, S. P. (1991). Application of the t, T^2 and F statistics to ordinal categorical data. Ph.D. Dissertation, Department of Mathematical Sciences, University of Texas at Dallas.

GILULA, Z. (1984). On some similarities between canonical correlation models and latent class models for two-way contingency tables. *Biometrika* **71** 523–529.

GLASS, D. V. AND HALL, J. R. (1954). Social Mobility in Britain: A study of inter-generation changes in status. In *Social Mobility in Britain,* D. V. Glass, ed. Routledge and Kogan Paul, London.

KENDALL, M. G. AND STUART, A. (1979). *The Advanced Theory of Statistics* Vol. 2, 4th edition, Griffen, London.

KIMELDORF, G. AND SAMPSON, A. R. (1975). Uniform representations of bivariate distributions. *Commun. Statist.* **4** 617–627.

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

KIMELDORF, G., MAY, J. H., AND SAMPSON, A. R. (1981). A program to compute concordant and other monotone correlations. In *Proceedings of Computer Science and Statistics: 13th Symposium on the Interface,* W. F. Eddy, ed. Springer-Verlag, 348-351.

KIMELDORF, G., MAY, J. H. AND SAMPSON, A. R. (1982). Concordant and discordant monotone correlations and their evaluation using nonlinear optimization. In *Optimization in Statistics,* S. H. Zanakis and J. S. Rustagi, eds., North-Holland, Amsterdam. 117–130.

KIMELDORF, G., SAMPSON, A. R. AND WHITAKER, L. (1992). Min and max scorings for two-sample ordinal data. *J. Amer. Statist. Assoc.* **87** 241–247.

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

LEHMANN, E. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37** 1137–53.

MAY, J. H. (1979). Solving nonlinear programs without using analytic derivatives. *Oper. Res.* **27** 457–484.

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

ROBERTSON, T., WRIGHT, F. T. AND DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference.* Wiley, New York.

SCHRIEVER, B. F. (1983). Scaling of order dependent categorical variables with correspondence analysis. *Intern. Statist. Rev.* **51** 225–238.

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

SETHURAMAN, J. (1990). The asymptotic distribution of the Renyi maximal correlation. *Commun. Statist. - Theo. Meth.* **19** 4291–4298.

SKLAR, A. (1959). Fonctions de répartition a n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris 8,* 229–231.

VITALE, R. A. (1990). On stochastic dependence and class of degenerate distributions. In *Topics of Statistical Dependence*, H. W. Block, T. Savits, and A. R. Sampson, eds. Institute of Mathematical Statistics, Hayward, CA. 459–469.

WILLIAMS, E. J. (1952). Use of scores for the analysis of association in contingency tables. *Biometrika* **39** 274–289.

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF PITTSBURGH
PITTSBURGH, PA 15260