The Annals of Statistics
1978, Vol. 6, No. 4, 895–903

MONOTONE DEPENDENCE

BY GEORGE KIMELDORF[1] AND ALLAN R. SAMPSON[2]

University of Texas at Dallas and Abbott Laboratories

Random variables X and Y are *mutually completely dependent* if there exists a one-to-one function g for which $P[Y = g(X)] = 1$. An example is presented of a pair of random variables which are mutually completely dependent, but "almost" independent. This example motivates considering a new concept of dependence, called *monotone dependence*, in which g above is now required to be monotone. Finally, this monotone dependence concept leads to defining and studying the properties of a new numerical measure of statistical association between random variables X and Y defined by $\sup\{\text{corr}[f(X), g(Y)]\}$, where the sup is taken over all pairs of suitable monotone functions f and g.

1. Introduction and summary. A random variable (rv) Y is defined (see Lancaster (1963)) to be *completely dependent* on a rv X if there exists a function g such that

$$P[Y = g(X)] = 1 \,. \tag{1.1}$$

Intuitively, Y is completely dependent on X if Y is perfectly predictable from X. The rv's X and Y are defined (see Lancaster (1963)) to be *mutually completely dependent* (MCD) if Y is completely dependent on X and X is completely dependent on Y. Equivalently, X and Y are MCD if (1.1) holds for some one-to-one function g. The concept of mutual complete dependence is, in a real sense, directly opposite to that of stochastic independence, in that mutual complete dependence entails complete predictability of either rv from the other, while stochastic independence entails complete unpredictability.

An important measure of dependence between two nondegenerate rv's X and Y is that of sup correlation, introduced by Gebelein (1941), studied among others by Rényi (1959) and Sarmanov (1958a, b), and defined by

$$\rho'(X, Y) = \sup \rho[f(X), g(Y)] \,,$$

where the supremum is taken over all Borel-measurable functions f, g, such that $0 < \operatorname{Var} f(X) < \infty$ and $0 < \operatorname{Var} g(Y) < \infty$, and where ρ represents the ordinary (Pearson product moment) correlation coefficient. The properties of sup correlation as a measure of dependence are discussed in Rényi (1959). It is clear

Received January 1977; revised September 1977.

[1] Sponsored in part by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant AFOSR-74-2581C.

[2] Sponsored in part by the Office of Naval Research under Grant No. N00014-76-C-0608.

The United States Government is authorized to reproduce and distribute reprints for governmental purposes.

AMS 1970 *subject classifications.* Primary 62H20; Secondary 62H05.

Key words and phrases. Correlation, association, monotone dependence, monotone correlation, sup correlation.

that two rv's which are MCD have sup correlation 1, but that the converse is not true. (See Lancaster (1963) for a discussion of necessary and sufficient conditions for the complete mutual dependence of random variables.)

Clearly, if a sequence $\{(X_n, Y_n)\}$ of pairs of independent rv's converges in law to a pair (X, Y) of rv's, then X and Y must be independent. It might be conjectured that if a sequence $\{(X_n, Y_n)\}$ of pairs of MCD rv's converges in law to a pair (X, Y) of rv's, then X and Y must be MCD. As is shown below, this conjecture is false. In fact, Section 2 presents a sequence of pairs of MCD rv's, all having the same marginals, which converges to a pair of independent rv's. This defect of mutual complete dependence motivates a new concept of total statistical dependence, called monotone dependence, which is defined and studied in Section 3.

When two rv's are neither totally statistically dependent nor totally independent, it is often useful to have a numerical measure, such as the correlation coefficient, to express the extent to which the rv's are related. A new numerical measure, called monotone correlation, is presented and examined in Section 4. This new measure is related in Section 5 to the concept of uniform representations of bivariate distributions.

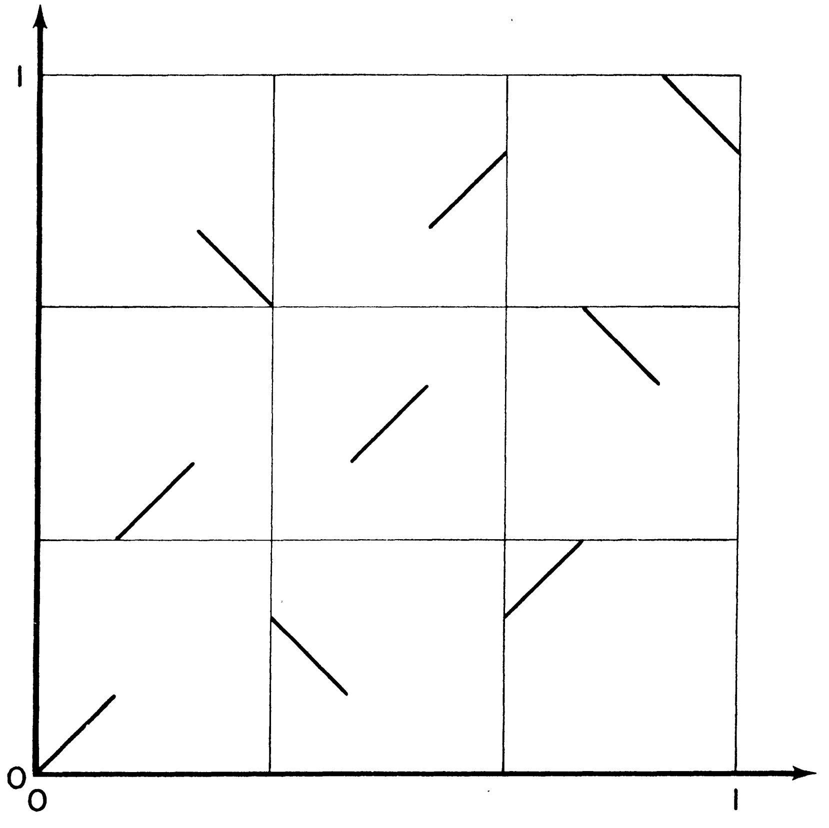

FIG. 1. Support of the distribution of (U_3, V_3).

2. MCD rv's which are almost independent. This section presents sequences $\{U_n\}$ and $\{V_n\}$ of rv's all having a uniform distribution on (0, 1) such that for each n, U_n and V_n are MCD, but that the pairs (U_n, V_n) converge in law to a pair (U, V) of independent rv's each having a uniform distribution on (0, 1).

Partition the unit square into n^2 congruent squares and denote by (i, j) the square whose upper right corner is the point with coordinates $x = i/n$, $y = j/n$. Similarly, partition each of these n^2 squares into n^2 subsquares and let (i, j, p, q) denote subsquare (p, q) of square (i, j). Now let the bivariate rv (U_n, V_n) distribute mass n^{-2} uniformly on either one of the diagonals of each of the n^2 subsquares of the form (i, j, j, i) for $1 \leqq i \leqq n$, $1 \leqq j \leqq n$. Figure 1 illustrates the case $n = 3$.

THEOREM 1. *Each of the* rv's U_n, V_n *has a uniform distribution on* (0, 1). *For each n the* rv's U_n *and* V_n *are* MCD. *The sequence* $\{(U_n, V_n)\}$ *converges in law to a pair* (U, V) *of independent uniform* rv's.

PROOF. For each n, it is clear that U_n and V_n are MCD. Also, since U_n and V_n each assign mass n^{-1} uniformly to each interval $((i-1)/n, i/n)$, it is clear that U_n and V_n have uniform distributions on (0, 1). Finally, since (U_n, V_n) assigns total mass n^{-2} to each of the n^2 large squares, $\lim_n P[U_n \leqq u, V_n \leqq v] = uv$ for each point (u, v) in the unit square. □

Now, let F and G be any pair of continuous cumulative distribution functions (cdf's). It is easy to generate sequences $\{X_n\}$ and $\{Y_n\}$ of rv's with respective marginals F and G such that X_n and Y_n are MCD for each n, yet $\{(X_n, Y_n)\}$ has joint limiting distribution $F \cdot G$ and hence is asymptotically independent. To do this, define $X_n = F^{-1}(U_n)$ and $Y_n = G^{-1}(V_n)$ where U_n and V_n are as above and where for any continuous cdf K, we define

$$K^{-1}(t) = \inf\{x : K(x) \geqq t\}. \tag{2.1}$$

This method of generating bivariate cdf's having specified continuous marginals from bivariate cdf's having uniform marginals is the method of translation. (See, for example, Mardia (1970b), Kimeldorf and Sampson (1975a).)

3. Montone dependence. The preceding example of pairs of MCD rv's which are almost independent suggests that mutual complete dependence is too broad a concept to be an antithesis of independence. We therefore propose the following concepts of total dependence.

DEFINITION. Let X and Y be continuous rv's. Then Y is *monotone dependent on* X if there exists a monotone function g for which $P[Y = g(X)] = 1$.

It is easy either to verify directly or to conclude as a corollary to Theorem 2 below that Y is monotone dependent on X if and only if X is monotone dependent on Y. We can therefore make the following definitions.

DEFINITIONS. Two continuous rv's X and Y are *monotone dependent* if there exists a monotone function g for which $P[Y = g(X)] = 1$. If g is increasing,

X and Y are said to be *increasing dependent*; if g is decreasing, X and Y are said to be *decreasing dependent*.

Before proceeding to show that monotone dependent rv's cannot be "almost" independent in the sense described in Section 2, we review some known results on Fréchet bounds. (See Fréchet (1951) and Mardia (1970a).) Let F and G be cdf's. Then

$$H^+(x, y) = \min [F(x), G(y)]$$

and

$$H^-(x, y) = \max [F(x) + G(y) - 1, 0]$$

are called the upper and lower Fréchet bounds, respectively, of the class of bivariate cdf's with marginals F and G. Both H^+ and H^- are singular bivariate distributions; H^+ assigns probability 1 to the set $\{(x, y): F(x) = G(y)\}$ and H^- to the set $\{(x, y): F(x) + G(y) = 1\}$. They are bounds in the sense that if H is any bivariate cdf with marginals F and G, then

$$H^-(x, y) \leqq H(x, y) \leqq H^+(x, y) . \tag{3.1}$$

(A proof of (3.1) appears in Johnson and Kotz ((1972), pages 22–23).)

THEOREM 2. *Let X and Y be continuous* rv's *with respective* cdf's F *and* G. *A necessary and sufficient condition that X and Y be increasing* (*decreasing*) *monotone dependent is that the joint* cdf *of* (X, Y) *is* $H^+(H^-)$.

PROOF. The sufficiency is immediate. To prove the necessity, assume that X and Y are increasing monotone dependent, so that (1.1) holds for some monotone increasing g. If $s \leqq t$, then

$$\begin{aligned} F(t) - F(s) &\leqq P[g(s) < g(X) \leqq g(t)] + P[g(X) = g(s)] \\ &= P[g(s) < Y \leqq g(t)] + P[Y = g(s)] \\ &= P[g(s) < Y \leqq g(t)] \\ &\leqq G(g(t)) - G(g(s)) . \end{aligned} \tag{3.2}$$

Let $t \to \infty$ and $s \to -\infty$ in (3.2) to derive $1 \leqq G(g(\infty)) - G(g(-\infty))$ and hence $G(g(\infty)) = 1$ and $G(g(-\infty)) = 0$. Let $s \to -\infty$ and set $t = x$ in (3.2) to derive

$$F(x) \leqq G(g(x)) . \tag{3.3}$$

Let $s = x$ and $t \to \infty$ in (3.2) to derive the inequality

$$1 - F(x) \leqq 1 - G(g(x)) , \tag{3.4}$$

which, together with (3.3), implies that $F(x) = G(g(x))$. Now, if H is the joint cdf of (X, Y), then

$$\begin{aligned} H(x, y) &= P[F(X) \leqq F(x), G(Y) \leqq G(y)] \\ &= P[G(g(X)) \leqq F(x), G(Y) \leqq G(y)] \\ &= P[G(Y) \leqq F(x), G(Y) \leqq G(y)] \\ &= \min \{F(x), G(y)\} . \end{aligned}$$

A similar argument is used if g is decreasing. $\square$

Theorem 2 is a partial justification for the interpretation of monotone dependence as an opposite to stochastic independence. The theorem implies that among all pairs of rv's with prescribed marginals, those which are as dependent as possible in the sense of (3.1) are exactly those which are monotone dependent. Section 2 presented a sequence of pairs of MCD continuous rv's which converges in law to a pair of independent rv's. The following theorem shows that this cannot happen for pairs of monotone dependent continuous rv's by showing that the property of monotone dependence is preserved under weak convergence.

THEOREM 3. *If* $\{(X_n, Y_n)\}$ *is a sequence of pairs of monotone dependent continuous* rv's *which converge in law to a pair* (X, Y) *of continuous* rv's, *then* X *and* Y *are monotone dependent.*

PROOF. Denote by H_n and H the respective bivariate cdf's of (X_n, Y_n) and (X, Y), and denote by F_n, G_n, F, and G the cdf's of X_n, Y_n, X, and Y, respectively. Since $\{(X_n, Y_n)\}$ converges in law to (X, Y), it follows that $\{F_n(x)\}$ converges to $F(x)$, $\{G_n(y)\}$ converges to $G(y)$ and there exists a subsequence $\{(X_{n_k}, Y_{n_k})\}$ such that either X_{n_k} and Y_{n_k} are increasing monotone dependent for all k or decreasing monotone dependent for all k. It follows in the former case by Theorem 2 that $H_{n_k}(x, y) = \min\{F_{n_k}(x), G_{n_k}(y)\}$, which converges to $H(x, y) = \min\{F(x), G(y)\}$. Therefore, X and Y are increasing monotone dependent. A similar argument holds if X_{n_k} and Y_{n_k} are decreasing monotone dependent for each k. □

4. Monotone correlation. Two continuous rv's X and Y are monotone dependent if there exists a perfect monotone relation between them. If the rv's are not perfectly monotonically related, it may be useful to measure numerically the degree of monotone dependence between them. One such measure, called monotone correlation, can be defined as follows:

DEFINITION. The *monotone correlation* ρ^* between two nondegenerate rv's X and Y is

$$\rho^*(X, Y) = \sup \rho[f(X), g(Y)], \tag{4.1}$$

where the supremum is taken over all monotone functions f, g, for which $0 < \operatorname{Var} f(X) < \infty$ and $0 < \operatorname{Var} g(Y) < \infty$.

It is clear that if two rv's are monotone dependent, then their monotone correlation is 1. To see that the converse implication fails, let (X, Y) have a uniform distribution over the region $[(0, 1) \times (0, 1)] \cup [(1, 2) \times (1, 2)]$ so that X and Y are not monotone dependent, although $\rho^*(X, Y) \geqq \rho[I_{(0,1)}(X), I_{(0,1)}(Y)] = 1$, where I denotes the indicator function.

It is obvious that

$$|\rho(X, Y)| \leqq \rho^*(X, Y) \leqq \rho'(X, Y). \tag{4.2}$$

For bivariate normal rv's (X, Y), it is well known that $|\rho(X, Y)| = \rho'(X, Y)$, in which case the inequalities in (4.2) are equalities. On the other hand, it can be easily seen that ρ^* is not in general equal to ρ'. For example, let (X, Y) have a uniform distribution on the region $[(0, 1) \times (0,1)] \cup [(0, 1) \times (2, 3)] \cup [(1, 2) \times (1, 2)] \cup [(2, 3) \times (0, 1)] \cup [(2, 3) \times (2, 3)]$ and let $f = I_{(0,1)} + I_{(2,3)}$, so that $\rho^*(X, Y) < 1$, but $\rho'(X, Y) \geqq \rho[f(X), f(Y)] = 1$.

While correlation as a measure of dependence is invariant under changes of scale and location in X and Y, monotone correlation is invariant under all order-preserving or order-reversing transformations of X and Y. Thus, monotone correlation would be a suitable measure of association for ordinal data. For a further discussion of measures of association for ordinal data, the reader is referred to Kruskal (1958) and Gibbons ((1971), Chapter 12).

Any candidate for a measure of association should have the property of being zero when the rv's are independent. Clearly, correlation, sup correlation, and monotone correlation all have this property. It would also be desirable for a measure of association to satisfy the converse implication, namely that it be zero only when X and Y are independent. Correlation clearly does not satisfy this converse property, although sup correlation does. (See Rényi (1959).) The following theorem shows that monontone correlation satisfies this converse implication. The proof of the theorem is essentially similar to that given by Rényi for sup correlation.

THEOREM 4. *If X and Y are nondegenerate* rv's *with monotone correlation zero, then X and Y are independent.*

PROOF. Suppose $\rho^*(X, Y) = 0$. For any real t, define $f_t = I_{(-\infty,t)}$. We claim that $\rho[f_s(X), f_t(Y)] = 0$. For if not, then either $\rho[f_s(X), f_t(Y)] > 0$ or $\rho[f_s(X), -f_t(Y)] > 0$, which contradicts the hypothesis. Now, $\rho[f_s(X), f_t(Y)] = 0$ implies that $P[X \leqq s, Y \leqq t] = P[X \leqq s] \cdot P[Y \leqq t]$, which implies independence. □

5. Uniform representation and monotone correlation. Let H be a continuous bivariate cdf with marginal cdf's F and G. The *uniform representation* U_H of H as defined by Kimeldorf and Sampson (1975b) is

$$U_H(u, v) = H(F^{-1}(u), G^{-1}(v)), \qquad 0 \leqq u \leqq 1, 0 \leqq v \leqq 1, \tag{5.1}$$

where F^{-1} and G^{-1} are as defined by (2.1). Observe that U_H is a cdf on the unit square with both marginal distributions being uniform on (0, 1). Thus, the class of all continuous bivariate cdf's can be decomposed into equivalence classes determined by the equivalence relation

$$H_1 \sim H_2 \quad \text{iff} \quad U_{H_1} = U_{H_2}. \tag{5.2}$$

If (X, Y) and (V, W) have continuous cdf's H and K, respectively, we write $(X, Y) \sim (V, W)$ whenever $H \sim K$.

LEMMA 1. *Let (X, Y) and (V, W) have continuous bivariate* cdf's *H and K, respectively. Then $H \sim K$* (i.e., $(X, Y) \sim (V, W)$) *if and only if there exist increasing functions A and B such that the joint* cdf *of $A(X)$ and $B(Y)$ is K.*

PROOF. Denote the marginals of H by F_X and F_Y and the marginals of K by

F_V and F_W. Suppose there exist increasing functions A and B such that $(A(X), B(Y))$ has continuous cdf K. Since $A(X)$ is a continuous rv, the marginal cdf of $A(X)$ is $F(s) = P[A(X) \leqq s] = P[X \leqq A^{-1}(s)] = (F_X \circ A^{-1})(s)$, where A^{-1} is as defined by (2.1). Similarly, the marginal cdf of $B(Y)$ is $(F_Y \circ B^{-1})(t)$. Therefore, the uniform representation of K is

$$\begin{aligned} U_K(s, t) &= K[(A^{-1})^{-1} \circ F_X^{-1}(s), (B^{-1})^{-1} \circ F_Y^{-1}(t)] \\ &= P[A(X) \leqq (A^{-1})^{-1} \circ F_X^{-1}(s), B(Y) \leqq (B^{-1})^{-1} \circ F_Y^{-1}(t)] \\ &= P[X \leqq F_X^{-1}(s), Y \leqq F_Y^{-1}(t)] \\ &= H(F_X^{-1}(s), F_Y^{-1}(t)), \end{aligned}$$

which is the uniform representation of H.

Conversely, suppose $H \sim K$. Let $A = F_V^{-1} \circ F_X$ and $B = F_W^{-1} \circ F_Y$. Then $A(X)$ has cdf F_V and $B(Y)$ has cdf F_W. Moreover, by (5.1) the uniform representation of the joint cdf of $(A(X), B(Y))$ is

$$\begin{aligned} U(s, t) &= P[A(X) \leqq F_V^{-1}(s), B(Y) \leqq F_W^{-1}(t)] \\ &= P[F_X(X) \leqq s, F_Y(Y) \leqq t] \\ &= P[X \leqq F_X^{-1}(s), Y \leqq F_Y^{-1}(t)], \end{aligned}$$

which is the uniform representation of H, hence of K. Finally, since the joint cdf of $(A(X), B(Y))$ has the same marginals as K and also the same uniform representation as K, the joint cdf is K. □

An elementary relationship between the concepts of uniform representation and monotone correlation is that

$$(X, Y) \sim (V, W) \qquad \text{implies} \quad \rho^*(X, Y) = \rho^*(V, W).$$

This relationship follows directly from Lemma 1. A further relationship between the concepts of uniform representation and monotone correlation is expressed by the following theorem, whose proof requires an additional lemma.

THEOREM 5. *Let (X, Y) have continuous bivariate* cdf *H. Then*

$$\rho^*(X, Y) = \sup \{|\rho(V, W)| : (V, W) \sim (X, Y)\}. \tag{5.3}$$

LEMMA 2. *Given nondegenerate* rv's *X and Y, let f and g be increasing functions for which $0 < \operatorname{Var} f(X) < \infty$ and $0 < \operatorname{Var} g(Y) < \infty$. Then there exist sequences $\{f_n\}$ and $\{g_n\}$ of strictly increasing functions for which $\operatorname{Var} f_n(X) < \infty$, $\operatorname{Var} g_n(Y) < \infty$, and $\lim_n \rho[f_n(X), g_n(Y)] = \rho[f(X), g(Y)]$.*

PROOF. We use the fact that any increasing function can be uniformly approximated by a strictly increasing function. Let $\{f_n\}$ and $\{g_n\}$ be sequences of strictly increasing functions converging uniformly to f and g, respectively. Since $f(X)$ and $g(Y)$ have finite nonzero variances, we have $\rho[f_n(X), g_n(Y)] \to \rho[f(X), g(Y)]$. □

PROOF OF THEOREM. Let any number $\varepsilon > 0$ be given. By the definition of

ρ^* and Lemma 2, there exist strictly increasing functions f and g such that $|\rho^*(X, Y) - |\rho[f(X), g(Y)]|| < \varepsilon$. Thus, the pair $(V = f(X), W = g(Y))$ has a continuous joint cdf and Lemma 1 can be applied to conclude that $(V, W) \sim (X, Y)$. Hence the left side of (5.3) cannot exceed the right side. To prove the reverse inequality, suppose $(V, W) \sim (X, Y)$. Then by Lemma 1, there exist increasing functions A and B for which $\rho[A(X), B(Y)] = \rho(V, W)$. Let $A' = [\operatorname{sgn} \rho(V, W)] \cdot A$ so that $|\rho(V, W)| = \rho[A'(X), B(Y)] \leqq \rho^*(X, Y)$. □

6. Remarks. If X and Y are univariate rv's with respective cdf's F and G, then the *grade correlation* (see, for example, Gibbons (1971)), which is the population analog of Spearman's rank correlation coefficient, is defined as $\rho_g = \rho[F(X), G(Y)]$. Thus, the grade correlation is the (ordinary) correlation coefficient of the uniform representation, and

$$\rho_g(X, Y) \leqq \rho^*(X, Y) \leqq \rho'(X, Y) .$$

Note that the probability integral transform can be used to standardize an ordinal scale by devising ranges that are equal in terms of probability. In computing relationships between two such ordinal variables, therefore, the rank correlation (or grade correlation) is a useful device. However, it might be argued that the scaling should be done in an absolute fashion, rather than relative to some sort of population distribution. What the monotone correlation measures is the maximal correlation that might be achieved under any such monotone scaling.

A continuous bivariate distribution can be decomposed into two components: its structure, by which is meant the equivalence class determined by the equivalence relation $\sim$ (defined by (5.2)) in which the distribution belongs, and its marginal distributions. Conversely, given any equivalence class and any pair of continuous univariate distributions, there exists a unique bivariate distribution with these two components. In this context, Whitt (1976) posed the following problem: If the marginals are fixed, for what structure is the correlation maximized? Whitt showed that the maximum correlation is achieved when the bivariate distribution is the upper Fréchet bound, and the correlation is minimized when the distribution is the lower Fréchet bound.

One can just as well pose the reverse problem: If the structure is fixed, for what pair of marginals is the correlation maximized? In general, there will not be any pair of marginals for which the maximum is achieved; on the other hand, Theorem 5 states that the supremum of the correlations is exactly the monotone correlation.

REFERENCES

[1] FRÉCHET, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon Sect. A Ser.* 3 **14** 53-77.

[2] GEBELEIN, H. (1941). Das statitistiche Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichungsrechnung. *Z. Angew. Math. Mech.* **21** 364-379.

[3] GIBBONS, J. D. (1971). *Nonparametric Statistical Inference.* McGraw-Hill, New York.

[4] JOHNSON, N. L. and KOTZ, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions.* Wiley, New York.

[5] KIMELDORF, G. and SAMPSON, A. (1975 a). One-parameter families of bivariate distributions with fixed marginals. *Comm. Statist.* **4** 293–301.

[6] KIMELDORF, G. and SAMPSON, A. (1975 b). Uniform representations of bivariate distributions. *Comm. Statist.* **4** 617–627.

[7] KRUSKAL, W. H. (1958). Ordinal measures of association. *J. Amer. Statist. Assoc.* **53** 814–861.

[8] LANCASTER, H. O. (1963). Correlation and complete dependence of random variables. *Ann. Math. Statist.* **34** 1315–1321.

[9] MARDIA, K. V. (1970a). A translation family of bivariate distributions and Fréchet's bounds. *Sankhyā, Ser. A* **32** 119–122.

[10] MARDIA, K. V. (1970 b). *Families of Bivariate Distributions.* Hafner, Darien.

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

[12] SARMANOV, O. V. (1958 a). The maximal correlation coefficient (symmetric case). *Dokl. Akad. Nauk SSSR* **120** 715–718. (English translation in *Sel. Transl. Math. Statist. Probability* **4** 271–275.)

[13] SARMANOV, O. V. (1958 b). The maximal correlation coefficient (nonsymmetric case). *Dokl. Akad. Nauk SSSR* **121** 52–55. (English translation in *Sel. Transl. Math. Statist. Probability* **4** 207–210.)

[14] WHITT, W. (1976). Bivariate distributions with given marginals. *Ann. Statist.* **4** 1280–1289.

PROGRAMS IN MATHEMATICAL SCIENCES
UNIVERSITY OF TEXAS—DALLAS
BOX 688
RICHARDSON, TEXAS 75080

DEPARTMENT OF BIOSTATISTICS
ABBOTT LABORATORIES
NORTH CHICAGO, ILLINOIS 60064