The Annals of Statistics
1983, Vol. 11, No. 1. 286–295

NEGATIVE ASSOCIATION OF RANDOM VARIABLES, WITH APPLICATIONS[1]

BY KUMAR JOAG-DEV AND FRANK PROSCHAN

University of Illinois and Florida State University; and Florida State University

Random variables, $X_1, \cdots, X_k$ are said to be negatively associated (NA) if for every pair of disjoint subsets A_1, A_2 of $\{1, 2, \cdots, k\}$, $\mathrm{Cov}[f(X_i, i \in A_1), g(X_j, j \in A_2)] \le 0$, for all nondecreasing functions f, g. The basic properties of negative association are derived. Especially useful is the property that nondecreasing functions of mutually exclusive subsets of NA random variables are NA. This property is shown not to hold for several other types of negative dependence recently proposed.

One consequence is the inequality $P(X_i \le x_i, i = 1, \cdots, k) \le \prod_1^k P(X_i \le x_i)$ for NA random variables $X_1, \cdots, X_k$, and the dual inequality resulting from reversing the inequalities inside the square brackets. In another application it is shown that negatively correlated normal random variables are NA. Other NA distributions are the (a) multinomial, (b) convolution of unlike multinomials, (c) multivariate hypergeometric, (d) Dirichlet, and (e) Dirichlet compound multinomial. Negative association is shown to arise in situations where the probability measure is permutation invariant. Applications of this are considered for sampling without replacement as well as for certain multiple ranking and selection procedures. In a somewhat striking example, NA and positive association representing quite strong opposing types of dependence, are shown to exist side by side in models of categorical data analysis.

1. Introduction and summary. The concept of (positively) associated random variables was introduced into the statistical literature by Esary, Proschan, and Walkup (1967). Since then a great many papers have been written on the subject and its extensions, and numerous multivariate inequalities have been obtained. In this paper we introduce the notion of negatively associated (NA) random variables, derive basic theoretical properties, and develop applications in multivariate statistical analysis. The theory and application of NA are not simply the duals of the theory and application of positive association, but differ in important respects.

Actually, NA is but one qualitative version of negative dependence among random variables. Other versions are treated in Block, Savits, and Shaked (1982), Ebrahimi and Ghosh (1981), Jogdeo and Patil (1975), and Karlin and Rinott (1980). (See Section 2 for definitions of certain types of negative dependence.)

Negative association has one distinct advantage over the other known types of negative dependence. Increasing functions of disjoint sets of NA random variables are also NA. This type of closure property does not hold for the three other types of negative dependence described in Section 2.

In Section 2 we define NA and develop its basic properties. We define other types of negative dependence, such as negative upper (lower) orthant dependence, reverse regular of order two (RR_2) in pairs, conditionally decreasing in sequence (CDS), and negatively dependent in sequence (NDS), introduced by other statisticians working in negative

---

[1] Research performed under the support of Air Force Office of Scientific Research, AFSC, USAF, under Grant AFOSR-82-K-0007.

Received August 1982.

AMS 1970 *subject classifications.* Primary 62H05; secondary 62H10.

Key words and phrases. Negative association, negative dependence, negative orthant dependence, association, conditionally decreasing in sequence, negatively dependent in sequence, reverse rule of order 2, multivariate inequalities, negatively correlated normal random variables.

dependence. We prove that among these types of negative dependence, only the NA class enjoys the important property of being closed under formation of increasing functions of disjoint sets of random variables, as mentioned above.

A very useful theorem is proved in Section 2 which gives simple sufficient conditions for the conditional joint distribution of $X_1, \cdots, X_k$ given the sum $\sum_{i=1}^k X_i$, to be NA (Theorem 2.8).

We show that permutation distributions (Definition 2.10) are NA. The applications of this result yield NA for the ranks of a random sample as well as for the values of observations obtained by a random sample when sampling is done without replacement from a finite population of values. These and the following applications are given in Section 3.

An aesthetically appealing result states very simply that negatively correlated normal random variables are NA. The somewhat surprising fact that (positive) association and NA may co-exist in the same model is illustrated by contingency tables. Under the assumption of independence, it is shown that with marginal totals fixed, the cell frequencies with no pair from the same row or column exhibit positive association.

We point out that a number of well known multivariate distributions possess the NA property, such as (a) multinomial, (b) convolution of unlike multinomials, (c) multivariate hypergeometric, (d) Dirichlet, and (e) Dirichlet compound multinomial. Not only does this strengthen the negative dependence properties observed by the previous authors, but also, via closure properties, this makes their proofs almost a triviality.

In the sequel, the following two well known results about covariance will be used.

Let (X, Y) be a pair of real random variables and Z be a real or vector valued random variable. Then

$$\operatorname{Cov}(X, Y) = E\{\operatorname{Cov}(X, Y \mid Z)\} + \operatorname{Cov}\{E(X \mid Z), E(Y \mid Z)\}. \tag{1.1}$$

The following inequality is known as Tchebycheff's inequality. Let X be a real random variable. For every pair of increasing functions f, g

$$\operatorname{Cov}\{f(X), g(X)\} \geq 0. \tag{1.2}$$

For f and g discordant functions, the inequality is reversed. (Two functions are discordant if one is increasing, the other decreasing.)

Throughout, we use increasing in place of nondecreasing, decreasing in place of nonincreasing, positive in place of nonnegative, and negative in place of nonpositive.

2. Theoretical results.

DEFINITION 2.1. Random variables $X_1, X_2, \cdots, X_k$ are said to be *negatively associated* (NA) if for every pair of disjoint subsets A_1, A_2 of $\{1, 2, \cdots, k\}$,

$$\operatorname{Cov}\{f_1(X_i, i \in A_1), f_2(X_j, j \in A_2)\} \leq 0 \tag{2.1}$$

whenever f_1 and f_2 are increasing. "NA" may also refer to the vector $\mathbf{X} = (X_1, \cdots, X_k)$ or to the underlying distribution of $\mathbf{X}$. Additionally, "NA" may denote negative association.

Clearly (2.1) holds if both f_1 and f_2 are decreasing. Also, without loss of generality, we may assume that $A_1 \cup A_2 = \{1, 2, \cdots, k\}$.

Some basic properties of NA follow.

DEFINITION 2.2. (Lehmann, 1966). Random variables X and Y are *negative quadrant dependent* (NQD) if for every real x, y,

$$P(X \leq x, Y \leq y) \leq P(X \leq x)P(Y \leq y). \tag{2.2}$$

PROPERTY P_1. For a pair of random variables, NQD is equivalent to NA.

This follows immediately from Definitions 2.1 and 2.2.

PROPERTY P_2. Let $A_1, \cdots, A_m$ be disjoint subsets of $\{1, \cdots, k\}$ and $f_1, f_2, \cdots, f_m$ be increasing positive functions. Then $X_1, \cdots, X_k$ NA implies

$$E \prod_{i=1}^{m} f_i(X_j, j \in A_i) \le \prod_{i=1}^{m} Ef_i(X_j, j \in A_i). \tag{2.3}$$

This follows from the repeated application of Definition 2.1.

DEFINITION 2.3. The random variables $X_1, \cdots, X_k$ are said to be *negatively upper orthant dependent* (NUOD) if for all real $x_1, \cdots, x_k$,

$$P(X_i > x_i, i = 1, \cdots, k) \le \prod_{i=1}^{k} P(X_i > x_i), \tag{2.4}$$

and *negatively lower orthant dependent* (NLOD) if

$$P(X_i \le x_i, i = 1, \cdots, k) \le \prod_{i=1}^{k} P(X_i \le x_i). \tag{2.5}$$

Random variables $X_1, \cdots, X_k$ are said to be *negatively orthant dependent* (NOD) if they are both NUOD and NLOD.

PROPERTY P_3. An immediate consequence of Property P_2 is that for A_1, A_2 disjoint subsets of $\{1, 2, \cdots, k\}$, and $x_1, \cdots, x_k$ real,

$$P(X_i \le x_i, i = 1, \cdots, k) \le P(X_i \le x_i, i \in A_1)P(X_j \le x_j, j \in A_2), \tag{2.6}$$

and

$$P(X_i > x_i, i = 1, \cdots, k) \le P(X_i > x_i, i \in A_1)P(X_j > x_j, j \in A_2). \tag{2.7}$$

In particular, $X_1, \cdots, X_k$ are NOD. The following three properties are obvious from the definitions.

PROPERTY P_4. A subset of two or more NA random variables is NA.

PROPERTY P_5. A set of independent random variables is NA.

PROPERTY P_6. Increasing functions defined on disjoint subsets of a set of NA random variables are NA.

PROPERTY P_7. The union of independent sets of NA random variables is NA.

PROOF. Let $\mathbf{X}$, $\mathbf{Y}$ be independent vectors, each NA. We shall show that the vector $(\mathbf{X}, \mathbf{Y})$ is NA. Let $(\mathbf{X}_1, \mathbf{X}_2)$ and $(\mathbf{Y}_1, \mathbf{Y}_2)$ denote arbitrary partitions of $\mathbf{X}$ and $\mathbf{Y}$ respectively. Let f and g be arbitrary increasing functions. Note that $E\{f(\mathbf{X}_1, \mathbf{Y}_1) \mid \mathbf{Y}_1\}$ is a $\mathbf{Y}_1$ measurable function, so that

$$E\{f(\mathbf{X}_1, \mathbf{Y}_1) \mid \mathbf{Y}_1, \mathbf{Y}_2\} = E\{f(\mathbf{X}_1, \mathbf{Y}_1) \mid \mathbf{Y}_1\} \tag{2.8}$$

almost surely. A similar result holds for $E\{g(\mathbf{X}_2, \mathbf{Y}_2) \mid \mathbf{Y}_2\}$. Denote these conditional expectations by $h_1(\mathbf{Y}_1)$ and $h_2(\mathbf{Y}_2)$ respectively, and note that h_1, h_2 are increasing. Thus

$$\begin{aligned} E\{f(\mathbf{X}_1, \mathbf{Y}_1)g(\mathbf{X}_2, \mathbf{Y}_2)\} &= E[E\{f(\mathbf{X}_1, \mathbf{Y}_1)g(\mathbf{X}_2, \mathbf{Y}_2) \mid \mathbf{Y}_1, \mathbf{Y}_2\}] \\ &\le E\{h_1(\mathbf{Y}_1)h_2(\mathbf{Y}_2)\} \le E\{h_1(\mathbf{Y}_1)\}E\{h_2(\mathbf{Y}_2)\} \\ &= E\{f(\mathbf{X}_1, \mathbf{Y}_1)\}E\{g(\mathbf{X}_2, \mathbf{Y}_2)\}, \end{aligned}$$

where the first inequality follows from the fact that $(\mathbf{X}_1, \mathbf{X}_2)$ is independent of $(\mathbf{Y}_1, \mathbf{Y}_2)$ and hence NA is preserved under conditioning. The second inequality holds since $(\mathbf{Y}_1, \mathbf{Y}_2)$ is NA. □

REMARK 2.4. Properties P_6 and P_7 broaden the scope of application of NA considerably. For example, to verify NA for distributions that arise as convolutions of relatively simple distributions, we need only verify NA for the simple distributions. We shall give examples in Section 3.

REMARK 2.5. Neither NUOD nor NLOD implies NA. We present an example in which $\mathbf{X} = (X_1, X_2, X_3, X_4)$ possesses NOD, but does not possess NA.

Let X_i be a binary random variable such that $P(X_i = 1) = .5$ *for* $i = 1, 2, 3, 4$. Let (X_1, X_2) and (X_3, X_4) have the same bivariate distributions, and let (X_1, X_2, X_3, X_4) have joint distribution as shown in Table 1.

TABLE 1

		(X_1, X_2)				
		(0, 0)	(0, 1)	(1, 0)	(1, 1)	marginal
(X_3, X_4)	(0, 0)	.0577	.0623	.0623	.0577	.24
	(0, 1)	.0623	.0677	.0677	.0623	.26
	(1, 0)	.0623	.0677	.0677	.0623	.26
	(1, 1)	.0577	.0623	.0623	.0577	.24
	marginal	.24	.26	.26	.24	

It can be verified that all the NLOD and all the NUOD conditions hold (with strict inequalities in some cases). However,

$$P(X_i = 1, i = 1, 2, 3, 4) > P(X_1 = X_2 = 1)P(X_3 = X_4 = 1),$$

violating NA.

In some applications negative dependence is created when the random variables are subjected to conditioning, as in the following theorem.

THEOREM 2.6. *Let* $X_1, \cdots, X_k$ *be independent and suppose that the conditional expectation* $E\{f(X_i, i \in A) \mid \sum_{i \in A} X_i\}$ *is increasing in* $\sum_{i \in A} X_i$, *for every increasing function* f *and every proper subset* A *of* $\{1, 2, \cdots, k\}$. *Then the conditional distribution of* $X_1, \cdots, X_k$ *given* $\sum X_i$, *is* NA *almost surely.*

PROOF. Let A_1, A_2 be an arbitrary proper partition of $\{1, 2, \cdots, k\}$. Let $S_1 = \sum_{i \in A_1} X_i$, $S_2 = \sum_{j \in A_2} X_j$, $S = S_1 + S_2$, and f_1, f_2 be a pair of increasing functions. Using (1.1), where the conditioning vector is taken as (S_1, S_2), it follows that

$$\operatorname{Cov}\{f_1(X_i, i \in A_1), f_2(X_j, j \in A_2) \mid S\} = \operatorname{Cov}\{E(f_1 \mid S_1, S_2), E(f_2 \mid S_1, S_2) \mid S\}.$$

With $S = S_1 + S_2$, the two terms inside the brackets on the right side are discordant functions of S_1 and hence by (1.2) the covariance is negative. □

Theorem 2.7 takes on added interest when considered in conjunction with the following theorem.

THEOREM 2.7 (Efron, 1965). *Let* $X_1, \cdots, X_k$ *be* k *independent random variables with* PF_2 (*log concave*) *densities, let* $S = \sum_{k=1}^{k} X_i$, *and let* $\phi(x_1, \cdots, x_k)$ *be increasing in each argument. Then*

$$E\{\phi(X_1, \cdots, X_k) \mid S_k = s\}$$

is increasing in (*almost every*) s.

As a consequence of Theorems 2.6 and 2.7, we immediately obtain:

THEOREM 2.8. *Let $X_1, \cdots, X_k$ be k independent random variables with PF_2 densities. Then the joint conditional distribution of $X_1, \cdots, X_k$ given $\sum X_i$ is* NA (a.s.).

REMARK 2.9. The above theorem when specialized to independent identically distributed random variables generates a conditional distribution which is permutation invariant. Let us denote these conditioned random variables by $Y_1, \cdots, Y_n$; then according to Diaconis and Freedman (1980) these are called "finitely exchangeable." The above authors studied the approximation of the distributions of finitely exchangeable sequences by a distribution of a finite piece embedded in an (infinite) exchangeable sequence. In particular, they show that for a finitely exchangeable sequence of length n, the "variation distance" between the distribution of a subset of k random variables and the approximating distribution of k random variables chosen from an (infinite) exchangeable sequence, can be bounded by a (universal) constant times k/n. It is well known that the random variables from an exchangeable sequence possess positive correlations and hence exhibit positive dependence. In view of Theorem 2.8, at least for the upper or lower orthants, the independent random variables lie in the middle of finitely exchangeable but negatively associated random variables and exchangeable random variables with positive dependence. Thus approximation by independent random variables seems to be suitable. This is verified by Diaconis and Freedman (1980) in Theorem 4 where random sampling without replacement is considered. We comment on this example in Section 3.

NA also arises naturally via permutation distributions described below.

DEFINITION 2.10. Let $\mathbf{x} = (x_1, \cdots, x_k)$ be a set of k real numbers. A *permutation distribution* is the joint distribution of the vector $\mathbf{X} = (X_1, \cdots, X_k)$ which takes as values all $k!$ expermutations of $\mathbf{x}$ with equal probabilities, each being $1/k!$, where $k > 1$.

THEOREM 2.11. *A permutation distribution is* NA.

PROOF. For $k = 2$, the assertion is easily verified. Thus assume that it is true for $k - 1$. Let $\mathbf{X}_1, \mathbf{X}_2$ be an arbitrary partition of $\mathbf{X}$, a vector of k components. Let f_1 and f_2 be increasing functions. We want to show that

$$\mathrm{Cov}\{f_1(\mathbf{X}_1), f_2(\mathbf{X}_2)\} \leq 0.$$

Without loss of generality, we may assume that f_1 and f_2 are permutation invariant. Further, suppose that x_1 is one of the minimum values of x_i, $i = 1, 2, \cdots, k$. Let I be the random variable indicating which component of $\mathbf{X}$ assumes the value x_1. Thus I takes values $1, 2, \cdots, k$ with equal probabilities. Now

$$\mathrm{Cov}\{f_1(\mathbf{X}_1), f_2(\mathbf{X}_2)\} = E\{\mathrm{Cov}(f_1, f_2 \mid I)\} + \mathrm{Cov}\{E(f_1 \mid I), E(f_2 \mid I)\}. \tag{2.9}$$

The first term on the right side of (2.9) is negative by the induction hypothesis. Further, due to the permutation invariance of f_1, $E\{f_1(\mathbf{X}) \mid I\}$ takes only two values. It is smaller when I corresponds to one of the indices of $\mathbf{X}_1$ than when I does not. Thus $E\{f_1(\mathbf{X}_1) \mid I\}$ and $E\{f_2(\mathbf{X}_2) \mid I\}$ are discordant functions of a binary random variable and hence by (1.2) their covariance is negative. This shows that the second term on the right side (2.9) is also negative. □

Finally, in this section, we want to compare some other concepts of negative dependence developed by Karlin and Rinott (1980) and modified by Block, et al (1982).

Let μ be a probability measure on the Borel sets in R^k. For intervals $I_1, \cdots, I_k$ in R^1, define a set function $\tilde{\mu}(I_1, \cdots, I_k) = \mu(I_1 \times \cdots \times I_k)$; for convenience, write μ instead of $\tilde{\mu}$. For intervals I, J in R^1, write $I < J$ if $x \in I$, $y \in J$ implies $x < y$.

DEFINITION 2.12. (a) Let μ be a probability measure on R_2. We say that μ is *reverse regular of order two* (RR_2) if

$$\mu(I_1, I_2)\mu(I_1', I_2') \leq \mu(I_1, I_2')\mu(I_1', I_2) \tag{2.10}$$

TABLE 2
The Joint Density g of (Y_1, Y_2)

	Y_1		
	0	1	2
0	$f(0, 3, 0) + f(3, 0, 0)$	0	$f(1, 2, 0) + f(2, 1, 0)$
1	$f(0, 2, 1) + f(2, 0, 1)$	$f(1, 1, 1)$	0
2	$f(0, 1, 2) + f(1, 0, 2)$	0	0
3	$f(0, 0, 3)$	0	0

for all intervals $I_1 < I_1'$, $I_2 < I_2'$ in R^1.

(b) Let μ be a probability measure on $R^k (k \geq 2)$. We say that μ is RR_2 in I_i, I_j when (2.10) holds for this pair with the remaining variables held fixed, for all $1 \leq i < j \leq k$. We also say that the random variables $X_1, \cdots, X_k$ (or the random vector $\mathbf{X}$ or its distribution function F) is RR_2 in pairs if its corresponding probability measure on R^k is.

DEFINITION 2.13. The random variables $X_1, \cdots, X_k$ are said to be: (a) *conditionally decreasing in sequence* (CDS) if for $i = 1, 2, \cdots, k - 1$,

$$[X_{i+1} \mid X_1 = x_1, \cdots, X_i = x_i] \text{ is decreasing stochastically in } x_1, \cdots, x_i,$$

(b) *negatively dependent in sequence* (NDS) if for $i = 1, \cdots, k$, $[X_1, \cdots, X_{i-1} \mid X_i = x_i]$ is decreasing stochastically in x_i.

THEOREM 2.14. (Block, et al, 1982). *Let $X_0, X_1, \cdots, X_k$ be independent random variables and let each have a* PF_2 *density or probability function. Then for fixed s, the conditional random variables $(X_1, \cdots, X_k \mid X_0 + X_1 + \cdots + X_k = x)$ are* RR_2 *in pairs and consequently* CDS *and* NOD.

We next show that Property P_5 enjoyed by the NA class is not enjoyed by the classes of RR_2 in pairs, NDS, or CDS.

RESULT 2.15. Neither RR_2 in pairs, CDS, nor NDS are closed under the formation of increasing functions of disjoint sets, while NA is.

Let $\mathbf{X} = (X_1, X_2, X_3)$ be a random vector with a trivariate multinomial frequency function f having strictly positive cell probabilities p_1, p_2, p_3, and satisfying $X_1 + X_2 + X_3 = 3$. Consider the induced random vector $\mathbf{Y} = (Y_1, Y_2)$, where $Y_1 = X_1 X_2$ and $Y_2 = X_3$. Denote the frequency of (Y_1, Y_2) by g. We now show that although f is RR_2 in pairs, CDS and NDS, g is neither RR_2, CDS, nor NDS.

The multinomial is shown to be RR_2 in pairs in Block, et al (1982) and Karlin and Rinott (1980). The former authors show also that the multinomial is CDS and NDS. Thus the multinomial f above has all three negative dependence properties.

To see that the induced g is not RR_2, we obtain the joint density g of (Y_1, Y_2) as shown in Table 2.

It follows that

$$\begin{vmatrix} g(0, 0) & g(0, 1) \\ g(1, 0) & g(1, 1) \end{vmatrix} > 0,$$

so that g cannot be RR_2.

To verify that (Y_1, Y_2) is not CDS, we note that:

$$P(Y_2 > 0 \mid Y_1 = 0) = 1 - P(Y_2 = 0 \mid Y_1 = 0) < 1,$$

while

$$P(Y_2 > 0 \mid Y_1 = 1) = 1 - P(Y_2 = 0 \mid Y_1 = 1) = 1.$$

Thus $\mathbf{Y}$ is *not* CDS.

For a bivariate vector, CDS and NDS are the same. It follows that $\mathbf{Y}$ is *not* NDS. Using Theorem 2.8, we conclude that $\mathbf{X}$ is NA. It follows by Property P_6 that $\mathbf{Y}$ is also NA. The powerful closure property possessed by NA but not by RR_2 in pairs, CDS, or NDS gives NA a great advantage over the other three classes of negative dependence.

REMARK 2.16. NA *does not imply* RR_2 *in pairs*, CDS, *or* NDS. This assertion may be verified directly from the preceding example. Specifically, $\mathbf{Y}$ in that example is NA, but is not RR_2, CDS, or NDS.

3. Applications.

3.1. *Standard multivariate distributions possessing the* NA *property.* In the papers listed earlier, a number of well known multivariate distributions are shown to possess the NOD properties, i.e., they satisfy inequalities (2.4) and (2.5). In some cases stronger properties are shown such as RR_2 in pairs, NDS, etc., which imply NOD. Specifically, distributions shown to enjoy the NOD properties are (a) Multinomial, (b) Convolution of unlike multinomials, (c) Multivariate hypergeometric, (d) Dirichlet, (e) Dirichlet compound multinomial, (f) Multinormals having certain covariance matrices.

We illustrate the ease with which NA can be established by considering (a) and (c).

(a) Let $\mathbf{Z} = (Z_1, \cdots, Z_k)$ be a vector having a multinomial distribution, obtained by taking only one observation. Thus only one Z_i is 1 while the rest are zero. The NA property for $\mathbf{Z}$ trivially follows from Definition 2.1. Since the general multinomial is the convolution of independent copies of $\mathbf{Z}$, the closure property (P_6) establishes NA in this case.

An alternative way to derive the multinomial is to condition independent Poisson random variables by fixing their sum. By Theorem 2.8, we see that the multinomial is NA.

(c) An urn contains N balls, each having a different color. Suppose a random sample of n balls is chosen (without replacement) and X_i, $i = 1, \cdots, N$, be random variables indicating the presence of a ball of the ith color in the sample. Clearly, the random variables X_i, $i = 1, \cdots, N$, possess a permutation distribution and hence are NA. More generally, suppose N_i balls are of the ith color, $i = 1, \cdots, k$, with $\sum_{i=1}^{k} N_i = N$, and let Y_i be the number of balls of the ith color in the sample. Then Y_i can be viewed as the sum of N_i indicators in the simple model above, where the ith color is obtained by pooling N_i colors. Since the Y_i are sums over nonoverlapping sets of random variables, by Property P_6 the NA property is transmitted.

3.2. *Further applications of permutation distributions.*

(a) *Random sampling without replacement.* The multivariate hypergeometric discussed above is a special case of the following model. Suppose a finite population consists of N values: $x_1, \cdots, x_N$. Let $X_1, \cdots, X_n$ represent the sample values obtained by random sampling without replacement. Then $X_1, \cdots, X_n$ may be considered as a subset of $X_1, \cdots, X_N$, which has a permutation distribution and hence, by Property P_4, is NA. In most textbooks the fact that $\operatorname{Cov}(X_i, X_j) \le 0$ is pointed out. However, NA is a much stronger property.

(b) *Joint distribution of ranks.* Let $X_1, \cdots, X_k$ be a random sample from a population. Let R_i be the rank of X_i, $i = 1, \cdots, k$. Then clearly $(R_1, \cdots, R_k)$ has a permutation distribution and hence is NA.

Next we present two applications discussed by Lehmann (1966) in the framework of the bivariate case. We prove that the relevant vectors are NA, a fact which can be used for multiple decision rules.

(c) *Selection procedures based on multiple rankings.* Suppose m judges independently rank n individuals. Let R_{ik} be the rank received by the ith individual from the kth judge, and $R_i = \sum_{k=1}^{m} R_{ik}$. Under the assumption that there is no preference among the individuals, Lehmann (1966) proves negative quadrant dependence for a pair (R_i, R_j).

Note that for k fixed, the vector $(R_{1k}, \cdots, R_{nk})$ is NA. Further, $(R_{11}, \cdots, R_{n1}), \cdots, (R_{1m}, \cdots, R_{nm})$ are mutually independent, implying that $(R_1, \cdots, R_n)$ is NA. Thus the Lehmann result is strengthened and extended to the multivariate case.

Suppose now that several "good" candidates are to be selected. The criterion of goodness is set by requiring that the rank R_i be at least as large as a constant $c(m, n)$. For this procedure, the probability of selecting k undesirable candidates is similar to the probability of an error of the first kind. If the least favorable distribution *happens* to correspond to the case of no preference among the candidates (that this may not always correspond to the least favorable distribution was shown by Rizvi and Woodworth, 1970), then NA provides an upper bound for the above probability:

$$P\{R_i \geq c(m, n); i = 1, \cdots, k\} \leq \sum_{i=1}^{k} P\{R_i \geq c(m, n)\}. \tag{3.1}$$

Usually, computing the upper bound on the right is much easier than computing the exact probability on the left.

The following is another variant where a selection rule similar to the rule above is used. Suppose the candidates are grouped according to regions of residence. The aim is to select *at most one* candidate from every region subject to the goodness requirement. Suppose there are p regions and let $R^{(j)} = \max R_i$, where the maximum is taken over the jth region. Then it follows that $(R^{(1)}, \cdots, R^{(p)})$ is NA and again we have:

$$P\{R^{(j)} \geq c(m, n, p); j = 1, \cdots, p\} \leq \prod_{j=1}^{p} P\{R^{(j)} \geq c(m, n, p)\},$$

a bound which is easier to compute.

3.3. *Moment inequalities.* Property P_2 can be utilized to derive moment inequalities. Suppose $Y_1, \cdots, Y_k$ are NA positive random variables, $\alpha_i \geq 0$, $i = 1, \cdots, m$, with $m \leq k$. Then

$$\mu_{\alpha_1\alpha_2\cdots\alpha_m} \leq \mu_{\alpha_1}\mu_{\alpha_2} \cdots \mu_{\alpha_m},$$

where $\mu_{\alpha_1\alpha_2\cdots\alpha_m} = E(Y_1^{\alpha_1}Y_2^{\alpha_2} \cdots Y_m^{\alpha_m})$ and $\mu_{\alpha_i} = E(Y_i^{\alpha_i})$. In particular,

$$E(Y_1 Y_2 \cdots, Y_m) \leq \prod_{i=1}^{m} E(Y_i).$$

In the above inequalities, $Y_1, \cdots, Y_m$ could be replaced by any other subset of m chosen from $Y_1, \cdots, Y_k$.

3.4. *Negatively correlated normal random variables are* NA. We use the same approach as in the proof that positively correlated normal random variables are associated, given by Joag-Dev, Perlman, and Pitt (1982). The present case is actually simpler since the functions utilized are defined on disjoint sets.

PROOF. Let $X_1, \cdots, X_k$ be jointly normally distributed random variables with covariance matrix Σ. By a lemma of Slepian (1962), it follows that

$$\frac{\partial}{\partial \sigma_{12}} P\{X_1 \geq a_1(\mathbf{X}_3), X_2 \geq a_2(\mathbf{X}_3), \mathbf{X}_3 \in A\} \geq 0, \tag{3.2}$$

where $\mathbf{X}_3 = (X_3, X_4, \cdots, X_k)$, a_1, a_2 are defined on R^{k-2}, A is an arbitrary measurable set in R^{k-2}, and σ_{ij} is the element in the ith row and jth column of Σ. From (3.2) it follows that if $f_1(x_1, \mathbf{x}_3)$ and $g_1(x_2, \mathbf{x}_3)$ are increasing in x_1 and x_2 respectively, then

$$E\{f_1(X_1, \mathbf{X}_3)g_1(X_2, \mathbf{X}_3)\} \uparrow \text{ in } \sigma_{12}. \tag{3.3}$$

Suppose now that f and g are increasing in each argument and A_1, A_2 are disjoint subsets of $\{1, \cdots, k\}$. Then

$$E\{f(x_i, i \in A_1)g(x_i, i \in A_2)\} \uparrow \text{ in } \sigma_{j_1 j_2} \tag{3.4}$$

for every pair (j_1, j_2) such that $j_1 \in A_1$, and $j_2 \in A_2$. From (3.4), it follows that if $\sigma_{j_1 j_2} \leq 0$ for every pair (j_1, j_2), then

$$E\{f(X_i, i \in A_1)g(X_i, i \in A_2)\} \leq Ef(X_i, i \in A_1)Eg(X_i, i \in A_2).$$

The desired result is now proven. □

TABLE 3

	A_1	A_2	$\cdots$	A_r	Totals
B_1	X_{11}	X_{12}	$\cdots$	X_{1r}	n_1
B_2	X_{21}	X_{22}	$\cdots$	X_{2r}	n_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
B_k	X_{k1}	X_{k2}	$\cdots$	X_{kr}	n_k
Totals	T_1	T_2	$\cdots$	T_r	$n = \sum n_i$

3.5. *Dependence among the cell frequencies in categorical data.* Suppose an individual or an item can be classified according to one of the categories A_i, $i = 1, \cdots, r$, corresponding to a characteristic A and also classified as B_j, $j = 1, \cdots, k$, corresponding to characteristic B. The usual categorical analysis is made for the independence between the two characteristics A and B. To test such a hypothesis, the usual statistical model assumed is the following. Let $n_1, \cdots, n_k$ be the sizes of independent random samples taken from subpopulations formed by partitioning the population according to the categories $B_1, \cdots, B_k$. Let X_{ij} be the cell frequency corresponding to B_i and A_j, with full data table in Table 3. Under the hypothesis of independence, the model above implies that the components in each row have a multinomial distribution, and the random vectors representing the rows are independent with a common parameter vector $(p_1, \cdots, p_k)$.

In the following we show that positive as well as negative association is manifested when the marginal totals are fixed. Note that under the hypothesis, the vector of column totals $(T_1, \cdots, T_r)$ is a sufficient statistic, so that the joint (discrete) distribution of $\{X_{ij}\}$ given $T_j = t_j$, $j = 1, \cdots, r$, is

$$P(X_{ij} = x_{ij},\ i = 1, \cdots, k;\ j = 1, \cdots, r) = \frac{\prod_1^r t_j! \prod_1^k n_i!}{n! \prod_{j=1}^r \prod_{i=1}^k x_{ij}!},$$

independent of $(p_1, \cdots, p_k)$.

We sketch the proof of the following two assertions.

A_1: The marginal distributions of row (column) vectors possess negative association.

A_2: The marginal distribution of a set of cell frequencies such that no pair of cells is in the same row or column (for example the set of diagonal cells) possesses (positive) association.

To verify these properties, first consider the special case where all row totals are 1. Thus the table represents the results of n independent copies of multinomial trials, each of size 1, and in the previous notation, $k = n$ and $n_i = 1$, $i = 1, \cdots, n$. It is clear that the conditional marginal distribution of the column vector $(X_{11}, X_{21}, \cdots, X_{n1})$ when the column totals are fixed at $T_i = t_i$, $i = 1, \cdots, r$, is the same as that of the conditional distribution of n Bernoulli random variables given that the sum is t_1. However, the Bernoulli random variable has a distribution which is PF_2, so that by Theorem 2.8, the vector is NA. Similarly, the conditional distribution of a row vector is multinomial with sample size 1 and probability vector $(t_1/n, \cdots, t_r/n)$.

To verify A_2, consider X_{11}, X_{22}, and X_{33}. We will show that these satisfy a regression condition called "stochastically increasing in sequence", which in this case amounts to showing that for every c,

$$P(X_{11} \geq c \mid X_{22}, X_{33}) \tag{3.5}$$

is increasing in X_{22} and in X_{33} with probability one and showing a similar property for

$$P(X_{22} \geq c \mid X_{33}). \tag{3.6}$$

Since X_{ij} are binary variables, the only nontrivial value of c is 1. It is easy to see that

$$P(X_{22} = 1 \mid X_{33} = 1) = \frac{t_2}{n-1},\ P(X_{22} = 1 \mid X_{33} = 0) = \frac{t_2}{n-1} \cdot \frac{n-1-t_3}{n-t_3}.$$

A similar computation establishes (3.5).

Next, to verify these properties for the case where row sums are $n_1, n_2, \cdots, n_k$, observe that the first row vector can be viewed as the sum of n_1 rows of the simple model, the second as the sum of the next n_2 rows, and so on. NA for the columns follows by the closure properties. The row vectors have a multinomial distribution and hence are NA. Positive association for cell frequencies from distinct rows and columns follows by an argument similar to that for the simple case.

REFERENCES

BLOCK, H. W., SAVITS, T. H. and SHAKED, M. (1982). Some concepts of negative dependence. *Ann. Probab.* **10** 765–772.

DIACONIS, P. and FREEDMAN, D. (1980). Finite exchangeable sequences. *Ann. Probab.* **8** 745–764.

EBRAHIMI, N. and GHOSH, M. (1981). Multivariate negative dependence. *Comm. Statist.* **A10**(4) 307–337.

EFRON, B. (1965). Increasing properties of Pólya frequency functions. *Ann. Math. Statist.* **36** 272–279.

ESARY, J. D., PROSCHAN, F. and WALKUP, D. W. (1967). Association of random variables, with applications. *Ann. Math. Statist.* **44** 1466–1474.

JOGDEO, K. and PATIL, G. P. (1975). Probability inequalities for certain multivariate discrete distributions. *Sankhyā* **37** B 158–164.

JOAG-DEV, K., PERLMAN, M. and PITT, L. (1983) Association of normal random variables and Slepian's inequality. *Ann. Probab.* **11**. To appear.

KARLIN, S. and RINOTT, Y. (1980). Classes of orderings of measures and related correlation inequalities. II. Multivariate reverse rule distributions. *J. Multivariate Analysis* **10** 499–516.

LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **43** 1137–1153.

MALLOWS, C. L. (1968). An inequality involving multinomial probabilities. *Biometrika* **55** 422–424.

RIZVI, M. H. and WOODWORTH, G. G. (1970). On selection procedures based on ranks: counterexamples concerning least favorable configurations. *Ann. Math. Statist.* **41** 1942–1951.

SLEPIAN, D. (1962). The one sided barrier problem for Gaussian noise. *Bell System Tech. J.* **41** 463–501.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS
URBANA ILLINOIS 61801

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306