The Annals of Statistics
1986, Vol. 14, No. 4, 1523–1543

CONDITIONAL ASSOCIATION AND UNIDIMENSIONALITY IN MONOTONE LATENT VARIABLE MODELS

BY PAUL W. HOLLAND AND PAUL R. ROSENBAUM

Educational Testing Service and University of Pennsylvania

Latent variable models represent the joint distribution of observable variables in terms of a simple structure involving unobserved or latent variables, usually assuming the conditional independence of the observable variables given the latent variables. These models play an important role in educational measurement and psychometrics, in sociology and in population genetics, and are implicit in some work on systems reliability. We study a broad class of latent variable models, namely the monotone unidimensional models, in which the latent variable is a scalar, the observable variables are conditionally independent given the latent variable and the conditional distribution of the observables given the latent variable is stochastically increasing in the latent variable. All models in this class imply a new strong form of positive dependence among the observable variables, namely *conditional* (positive) association. This positive dependence condition may be used to test whether any model in this class can provide an adequate fit to observed data. Various applications, generalizations and a numerical example are discussed.

1. Introduction. Latent variable models for multivariate distributions arise in a wide variety of applications. Examples include: factor analysis models for the multivariate normal (e.g., Lawley and Maxwell, 1971), binary response models for dichotomously scored (1 = correct, 0 = incorrect) exam questions (e.g., Rasch, 1960, Birnbaum, 1968, Bock and Lieberman, 1970, Goldstein, 1980, Lord, 1980, Bartholomew, 1980, Tjur, 1982 and Cressie and Holland, 1983), latent trait models for graded responses (Samejima, 1969 and Andersen, 1980), factor analytic models under exponential family distributions (Bartholomew, 1984), latent structure models for discrete data (e.g., Goodman, 1974) and certain genetic models relating (observable) phenotypes to (latent or unobservable) genotypes (e.g., Crow and Kimura, 1970 and Elandt-Johnson, 1971).

All of these latent variable models involve an observable (or manifest) random vector

$$\mathbf{X} = (X_1, \dots, X_J) \tag{1.1}$$

and an unobservable (or latent) variable $\mathbf{U}$, which may be either unidimensional or vector valued. We show that certain general classes of latent variable models imply that the manifest variables $\{X_j\}$ must exhibit strong forms of positive association. This result often leads to simple tests of the goodness-of-fit of classes of latent variable models. Our strongest results concern one-dimensional

Received April 1985; revised December 1985.
AMS 1980 *subject classifications.* Primary 60E15, 62P15; secondary 62H99.
Key words and phrases. Latent variable models, item response theory, population genetics, latent class models, factor analysis, systems reliability, conditional association, associated random variables, TP_2, MTP_2, SPOD, probability inequalities.

monotone latent variable models (defined in Section 2.2) and may be used to test the hypothesis that the observable distribution of $\mathbf{X}$ may be represented by a unidimensional latent variable model—i.e., informally, to test for the existence of a one-dimensional latent variable U that underlies a set of data.

The paper is organized as follows. Section 2 defines various classes of latent variable models. Section 3 reviews the relevant notions of positive association needed in this paper. Section 4 gives our main results on the types of positive association that must obtain among the manifest variables for various types of latent variable models. In Section 5 we consider the case of dichotomous manifest variables of special relevance to item response models used in educational tests. Section 6 gives both theoretical and practical examples to illustrate the results.

2. Latent variable models. In any latent variable model, the *manifest* variables, $\mathbf{X} = (X_1, \ldots, X_J)$, and the *latent* variable, $\mathbf{U}$, are assumed to have a joint distribution over a sample space (or population). The manifest variables, which are real or integer valued, can be observed directly while the latent variable is unobserable. In general, $\mathbf{U}$ may be either a vector or a scalar; however, when we restrict attention to a scalar latent variable, we will allow the notation to emphasize the restriction by writing U in place of $\mathbf{U}$.

In this section, we shall discuss several general classes of latent variable models defined by the conditions that the joint distribution of $(\mathbf{X}, \mathbf{U})$ is assumed to satisfy. Our purpose here is to give a precise definition of the models that concern us.

2.1. *Latent conditional independence.* A basic condition on $(\mathbf{X}, \mathbf{U})$ is that of *latent conditional independence*. (In item response theory, this condition is called *local independence*; e.g., Lord, 1980.) The conditional distribution function of $\mathbf{X}$ given $\mathbf{U}$ is given by

$$F(x_1, \ldots, x_J|\mathbf{u}) = P(X_1 \le x_1, \ldots, X_J \le x_J|\mathbf{U} = \mathbf{u}), \tag{2.1}$$

and the condition of latent conditional independence states that $X_1, \ldots, X_J$ are conditionally independent given $\mathbf{U}$ or

$$F(x_1, \ldots, x_J|\mathbf{u}) = \prod_{j=1}^{J} F_j(x_j|\mathbf{u}) \tag{2.2}$$

for all $x_1, \ldots, x_J$ and $\mathbf{u}$.

In (2.2) $F_j(x_j|\mathbf{u})$ is the conditional distribution function of X_j given $\mathbf{U} = \mathbf{u}$, i.e.,

$$F_j(x_j|\mathbf{u}) = P(X_j \le x_j|\mathbf{U} = \mathbf{u}). \tag{2.3}$$

A variable $\mathbf{U}$ for which latent conditional independence holds is often said to completely "explain" the association structure between the manifest variables, $X_1, \ldots, X_J$. Presumably, the sense in which $\mathbf{U}$ "explains" this association is that, given the value of $\mathbf{U}$, there is *no* association between the X_j's. While this notion of a "latent explanation" for observed association has a long history, by itself latent conditional independence is a vacuous assumption as the following theorem illustrates.

THEOREM 1 (Suppes and Zanotti, 1981). *If* $\mathbf{X}$ *has only a finite number of possible values then there always exists a one-dimensional latent random variable* U *such that* $(\mathbf{X}, U)$ *satisfies latent conditional independence.*

The proof of Theorem 1 when $\mathbf{X}$ is a $0/1$ vector is given in Suppes and Zanotti (1981) and it is easily generalized to the discrete case covered by Theorem 1. Any continuous distribution on R^J may be approximated arbitrarily well by a discrete distribution of R^J, and Theorem 1 applies to each such discrete approximation. Furthermore, in Theorem 1, U is not unique. In practical terms, latent conditional independence taken alone is neither a mathematical assumption—since for some U it is, in effect, always satisfied—nor a scientific hypothesis—since it places no testable restrictions on the behavior of observable data. These considerations emphasize the importance of other conditions in addition to latent conditional independence. Conditions such as linearity, monotonicity or functional form are not incidental conveniences, but rather they are the features of latent variable models that give them testable consequences in observed data. In Section 2.2, we discuss an important class of such conditions.

An interesting open question concerns the degree of smoothness or regularity that can be assumed for the conditional distributions $F_j(x_j|u)$ in Theorem 1 (regarded as functions of u). The construction given by Suppes and Zanotti is not smooth at all but the nonuniqueness of U suggests that smooth representations may be possible.

2.2. *Latent monotonicity.* We say that a latent variable model satisfies the condition of *latent monotonicity* if the functions

$$1 - F_j(x|\mathbf{u}) = P(X_j > x|\mathbf{U} = \mathbf{u})$$

are all nondecreasing functions of $\mathbf{u}$ for all values of x and for $j = 1, \ldots, J$. In case $\mathbf{U}$ is a vector, latent monotonicity requires that $1 - F_j(x|\mathbf{u})$ be nondecreasing in each coordinate of $\mathbf{u}$.

The idea behind latent monotonicity is that higher values of $\mathbf{u}$ imply stochastically larger distributions of X_j, for each j; i.e., $X_j|\mathbf{U} = \mathbf{u}$ is stochastically larger than $X_j|\mathbf{U} = \mathbf{u}'$ if $\mathbf{u} > \mathbf{u}'$. When X_j is a $0/1$ variable, latent monotonicity is equivalent to the requirement that the *item characteristic curve* (this is the term used in the psychometric literature but, in the case of a multidimensional $\mathbf{u}$, $r_j(\mathbf{u})$ defines a surface or hypersurface rather than a curve),

$$r_j(\mathbf{u}) = P(X_j = 1|\mathbf{U} = \mathbf{u}), \tag{2.4}$$

be nondecreasing in $\mathbf{u}$, so that a higher value of $\mathbf{u}$ implies a higher probability that $X_j = 1$ for each $j = 1, 2, \ldots, J$. Latent monotonicity is a natural condition when U is a scalar intended to represent an "ability" or an "attitude" that is measured by the J dichotomously scored (1 = correct or affirmative, 0 = incorrect or negative) exam or questionnaire items. Higher values of U are then associated with a greater chance of a correct or affirmative response to each item. Holland (1981) defines and studies a weaker type of monotonicity condition for the case of binary X_j's that is useful even when latent conditional independence does not hold.

If a latent variable model satisfies both latent conditional independence and latent monotonicity, then we shall call it a *monotone latent variable model*. The following lemma states that, for a monotone latent variable model, not only are the probabilities $P(X_j > x|\mathbf{U} = \mathbf{u})$ nondecreasing functions of each coordinate of $\mathbf{u}$, but so are the conditional expectations $E(g(\mathbf{X})|\mathbf{U} = \mathbf{u})$ for all nondecreasing functions of $\mathbf{X}$.

LEMMA 2 (Kamae, Krengel and O'Brien, 1977). *If* $(\mathbf{X}, \mathbf{U})$ *satisfies the conditions of latent conditional independence and latent monotonicity, then for any bounded function* $g(\mathbf{x})$ *that is nondecreasing in each coordinate, the conditional expectation*

$$E(g(\mathbf{X})|\mathbf{U} = \mathbf{u})$$

is nondecreasing in each coordinate of $\mathbf{u}$.

The proof of this lemma is given in several places. It is a special case of a general result of Kamae, Krengel and O'Brien (1977, Proposition 1). A straightforward proof is given by Ahmed, Leon and Proschan (1981, Lemma 3.3) and, in the case of binary $\mathbf{X}$, by Rosenbaum (1984, Lemma 1).

Another interpretation of Lemma 2 is in terms of the stochastic ordering of distributions. For two random vectors $\mathbf{X}$ and $\mathbf{X}'$ of equal dimension, the distribution of $\mathbf{X}$ is said to be *stochastically larger* than that of $\mathbf{X}'$ if $E(g(\mathbf{X})) \geq E(g(\mathbf{X}'))$ for all bounded functions $g(\cdot)$ that are nondecreasing in each coordinate (e.g., Lehmann, 1955, Marshall and Olkin, 1979, Tong, 1980, Section 6.3, and Eaton, 1982, Section 3). The conclusion of Lemma 2 says that if each coordinate of $\mathbf{u}$ is at least as large as the corresponding coordinate of $\mathbf{u}'$, then the conditional distribution of $\mathbf{X}$ given $\mathbf{U} = \mathbf{u}$ is stochastically larger than that of $\mathbf{X}$ given $\mathbf{U} = \mathbf{u}'$. Thus, for monotone latent variable models, not only are the distributions of each coordinate of $\mathbf{X}$ stochastically ordered by the coordinatewise partial ordering of the vector $\mathbf{u}$, but the multivariate distributions of $\mathbf{X}$ are stochastically ordered by $\mathbf{u}$.

2.3. *Latent unidimensionality.* If, in a monotone latent variable model, U is a scalar, then we call the model a *unidimensional monotone latent variable model*. Unidimensional models are of special interest for several reasons. First, they are often the most parsimonious of the latent variable models, usually involving the fewest parameters and leading to the simplest descriptions. Second, as we show in Section 4, unidimensional models lead to stronger forms of positive dependence than do multidimensional models. Third, a scalar U easily lends itself to the interpretation as an underlying "true" quantity that is fallibly measured by the observable responses in $\mathbf{X}$—e.g., the true "ability" that is fallibly measured by the exam responses $\mathbf{X}$. Though suggestive, this interpretation is often justly criticized. Rubin (1982), for example, has argued that it is at least confusing to say that the observable $\mathbf{X}$ is somehow less "true" than the unobservable and hypothetical U.

2.4. *A stronger type of latent monotonicity.* When $X_j|U$ has a density or a discrete mass function, $f_j(x|u)$, and U is unidimensional, there is a stronger form of latent monotonicity that we will use later. The latent variable model is *latent* TP_2 if for each j

$$\det\begin{bmatrix} f_j(x|u) & f_j(x|u') \\ f_j(x'|u) & f_j(x'|u') \end{bmatrix} \geq 0 \tag{2.5}$$

for all $x > x'$ and $u > u'$.

If f_j is strictly positive (2.5) can be expressed in an "odds ratio" form, i.e.,

$$\frac{f_j(x'|u')}{f_j(x'|u)} \frac{f_j(x|u)}{f_j(x|u')} \geq 1 \tag{2.6}$$

for all $x' > x$, $u' > u$.

The TP_2 condition (2.5) arises in various areas of probability and statistics (Karlin, 1968), and its multivariate generalization (Karlin and Rinott, 1980) will play an important role in Sections 3 and 4.

It is well known that (2.5) implies latent monotonicity when U is one-dimensional, and that the converse is false (e.g., Barlow and Proschan, 1975, Section 5.4). Hence, latent TP_2 is a stronger form of latent monotonicity. However, when the X_j are binary, latent TP_2 and latent monotonicity are equivalent—a fact we use in Section 5.

2.5. *Examples of monotone latent variable models.* The following examples of monotone latent variable models will be discussed again in Section 6.

(i) *Item response theory in educational testing and psychometrics.* In an examination consisting of J dichotomously scored questions or items, let $X_j = 1$ if item j is correct, and $X_j = 0$ if item j is incorrect. Write $r_j(u)$ for the item characteristic curve $P(X_j = 1|U = u)$—i.e., for the probability of a correct (or right) response to item j given the value of a *scalar* latent variable U. Rasch (1960) assumes that the item characteristic curve has the form $r_j(u) = 1/(1 + \exp(\alpha_j - u))$ with parameter α_j, whereas Lord (1980, 1982) assumes $r_j(u) = \gamma_j + (1 - \gamma_j)/\{1 + \exp(\alpha_j - \beta_j u)\}$ with parameters $0 \leq \gamma_j \leq 1$, α_j real and $\beta_j \geq 0$. Latent monotonicity is the condition that higher values of U—the latent "ability"—imply a higher probability of correctly responding to each item. Item response models represent the joint distribution of the item responses $\mathbf{X}$ in a population of examinees as

$$P(\mathbf{X} = \mathbf{x}) = \int \prod_{j=1}^{J} r_j(u)^{x_j}\{1 - r_j(u)\}^{1-x_j}\, dF(u), \tag{2.7}$$

where $F(\cdot)$ is the distribution of "ability" in the population. Expression (2.7) is the condition of latent conditional independence (2.2).

(ii) *Systems reliability theory.* In systems reliability theory (e.g., Barlow and Proschan, 1975), the X_j's indicate whether or not the jth of J components of a system is functioning, with $X_j = 1$ if the component is functioning and $X_j = 0$ if the component has failed. In this case, a monotone latent variable model might attempt to describe the pattern of association in component failures $\mathbf{X}$ with reference to certain latent stresses that have detrimental effects on all components; here, higher levels of stress are indicated by lower values of $\mathbf{U}$.

(iii) *Factor analysis for multivariate normal* $\mathbf{X}$. A linear factor analysis model (Lawley and Maxwell, 1971) for continuous $\mathbf{X}$ involves the assumption that

$$\mathbf{X} = \Lambda\mathbf{U} + \mathbf{e}, \tag{2.8}$$

where $\mathbf{U}$ is a multivariate normal vector of latent factors, $\mathbf{e}$ is a J-dimensional vector of $\mathbf{J}$ mutually independent normal random deviates which are also independent of $\mathbf{U}$, and Λ is a fixed matrix of factor loadings. When all of the factor loadings are nonnegative, (2.8) is a monotone latent variable model; when, in addition, U is a scalar and $\Lambda \geq \mathbf{0}$ is a vector, then (2.8) is a unidimensional monotone latent variable model. If U is a scalar and Λ is a vector of 1's, then (2.8) reduces to the "true score theory" for composite exams (Lord and Novick, 1968, Section 4). More generally, a nonlinear unifactor model of the form

$$\mathbf{X} = \mathbf{f}(U) + \mathbf{e} \tag{2.9}$$

for scalar U, independent deviations $\mathbf{e}$, and monotone increasing vector valued function $\mathbf{f}(\cdot)$, is a unidimensional monotone latent variable model.

(iv) *Population genetics: segregation analysis.* For convenience, consider sibships of fixed size J—i.e., J offspring of common parents. Let X_j, $j = 1, 2, \ldots, J$, indicate whether ($X_j = 1$) or not ($X_j = 0$) the jth sib has a specific (observable) phenotype which is completely determined by the presence (AA or Aa) or absence (aa) of a single dominant allele (A). As in Table 1, let the scalar U be the conditional probability of the trait given parents' unobservable genotypes; i.e., $U = P(X_j = 1 \mid \text{Parent's Genotypes})$. Then the X_j's are conditionally independent given U and, with this coding of the genotype, satisfy a monotone unidimensional latent variable model, both in the population as a whole, and in any subpopulation defined by the parents phenotype. Indeed, the distribution of $\mathbf{X}$ is exchangeable (Kingman, 1978):

$$P(\mathbf{X} = \mathbf{x}) = \sum_u u^{\mathbf{x}^T\mathbf{1}}\{1 - u\}^{J-\mathbf{x}^T\mathbf{1}}P(U = u). \tag{2.10}$$

While Table 1 is confined to one of the simplest models of inheritance, condition (2.10) holds quite generally for sibships, providing only that an offspring's manifest phenotype depends only on its parents genotypes: the unidimensional latent variable U is the "segregation parameter" (cf. Elandt-Johnson, 1971, Chapters 17 and 18).

(v) *Latent class models for dichotomous responses.* In latent class models for dichotomous responses, the binary manifest variables $\mathbf{X} = (X_1, X_2, \ldots, X_J)$ are assumed to be conditionally independent within each latent class, that is,

TABLE 1
A unidimensional monotone latent variable model for traits in sibships

Parents' Phenotypes (Observable)		Parents' Genotypes (Unobservable)		Probability of the Trait Given Parents' Genotypes
Mother	**Father**	**Mother**	**Father**	$U = P(X_j = 1 \mid U)$ for $j = 1, 2, \ldots, J$
A?	A?	AA	AA	1
A?	A?	AA	Aa	1
A?	aa	AA	aa	1
A?	A?	Aa	AA	1
A?	A?	Aa	Aa	3/4
A?	aa	Aa	aa	1/2
aa	A?	aa	AA	1/2
aa	A?	aa	Aa	1/2
aa	aa	aa	aa	0

conditionally independent given a discrete *nominal* variable U taking values in an unordered finite set S (e.g., Goodman, 1974), so

$$P(\mathbf{X} = \mathbf{x}) = \sum_{u \in S} \prod_{i=1}^{J} \gamma_{ui}^{x_i} \{1 - \gamma_{ui}\}^{1 - x_i} P(U = u) \tag{2.11}$$

with $0 \leq \gamma_{ui} \leq 1$ for all $u \in S$ and $i = 1, 2, \ldots, J$. Such a variable U always exists if we permit S to contain 2^J elements. If S contains only two elements, then we may always number them 0 and 1, and then relabel some coordinates of $\mathbf{X}$ by replacing some X_i by $1 - X_i$, to obtain a unidimensional monotone latent variable model.

3. Some types of positive multivariate association. In this section we first review three important types of positive association that apply to multivariate distributions. Since we will be applying these ideas to the manifest variables of a latent variable model, random vectors will be denoted by $\mathbf{X} = (X_1, \ldots, X_J)$. After this review, we define three parallel but stronger types of positive association that will be used in our discussion of unidimensional monotone latent variable models in Section 4.

3.1. *A review of three types of positive multivariate association.* Esary, Proschan and Walkup (1967) define the notion of *associated* random variables as follows.

DEFINITION 3.1. *Associated random variables* (Esary, Proschan and Walkup, 1967). *The distribution of a random vector* $\mathbf{X}$ *is (positively) associated* (A) *if*

$$\operatorname{Cov}(f(\mathbf{X}), g(\mathbf{X})) \geq 0 \tag{3.1}$$

for all nondecreasing, bounded functions, $f(\cdot)$ *and* $g(\cdot)$.

When Definition 3.1 holds for the distribution of $\mathbf{X}$, then $\mathbf{X}$ is said to be associated.

In Definition 3.1, $f(\mathbf{x})$ and $g(\mathbf{x})$ are nondecreasing in each of their J coordinates. It is well known that if $\mathbf{X}$ is one-dimensional then $\mathbf{X}$ is associated, and if the coordinates of $\mathbf{X}$ are independent then $\mathbf{X}$ is associated. See Esary, Proschan and Walkup (1967) for proofs of these and other properties of associated random variables.

Karlin and Rinott (1980) define a second type of multivariate positive association, MTP_2, for a random vector $\mathbf{X}$ with a density or discrete mass function which we shall denote by $p(\mathbf{x})$. For the history of this type of positive association, see Karlin and Rinott (1980) and the references given there. For two J-vectors $\mathbf{x}$ and $\mathbf{x}^*$ let $\max(\mathbf{x}, \mathbf{x}^*)$ and $\min(\mathbf{x}, \mathbf{x}^*)$ be defined by

$$\max(\mathbf{x}, \mathbf{x}^*) = (\max(x_1, x_1^*), \dots, \max(x_J, x_J^*))$$

and

$$\min(\mathbf{x}, \mathbf{x}^*) = (\min(x_1, x_1^*), \dots, \min(x_J, x_J^*)). \tag{3.2}$$

Hence, $\max(\mathbf{x}, \mathbf{x}^*)$ and $\min(\mathbf{x}, \mathbf{x}^*)$ are the *coordinatewise* max and min of the two vectors, $\mathbf{x}$ and $\mathbf{x}^*$.

DEFINITION 3.2. MTP_2 (Karlin and Rinott, 1980). *The distribution of a random vector* $\mathbf{X}$ *is multivariate totally positive of order* 2 (MTP_2) *if* $\mathbf{X}$ *has a density or a mass function* $p(\mathbf{x})$ *such that for all* $\mathbf{x}, \mathbf{x}^*$,

$$p(\max(\mathbf{x}, \mathbf{x}^*))p(\min(\mathbf{x}, \mathbf{x}^*)) \geq p(\mathbf{x})p(\mathbf{x}^*). \tag{3.3}$$

If $\mathbf{X}$ is two-dimensional, then MTP_2 and TP_2 defined in Section 2.4 are equivalent. Examples of MTP_2 random vectors $\mathbf{X}$ include:

(i) independent variables,

(ii) multivariate normal $\mathbf{X}$ in which the partial correlation between each pair of coordinates of $\mathbf{X}$ given the remaining coordinates is positive (Karlin and Rinott, 1983, Fact 3 and Theorem 3),

(iii) binary variables with an attractive interaction potential (Preston, 1974, Proposition 8.1, page 60),

(iv) binary variables that satisfy a monotone unidimensional latent variable model (Holland, 1981, Theorem 3).

Later, in Corollary 11 and the counterexample in Section 6.1, we strengthen (iv) above; specifically, we show that all monotone unidimensional latent variable models for binary $\mathbf{X}$ imply a form of positive dependence for $\mathbf{X}$ that is strictly stronger than MTP_2.

In the statement of Theorem 7 in Section 4 we will need to observe that the definition of MTP_2 in (3.3) makes no use of the fact that $p(\mathbf{x})$ is a density and can be applied to any nonnegative function of several variables. In Theorem 5 we will apply it to the function $p_j(x_j|\mathbf{u})$ regarded as a function of $(x_j, \mathbf{u})$.

The third type of positive association was defined by Joag-Dev (1983).

DEFINITION 3.3. SPOD (Joag-Dev, 1983). *The distribution of a random vector* $\mathbf{X}$ *is strongly positively orthant dependent (SPOD) if and only if the following three conditions are satisfied for any set of constants* $c_1, \dots, c_J$ *and any partition* $(B, \bar{B})$ *of* $A = \{1, \dots, J\}$:

(i) $P\{X_j \geq c_j,\ j \in A\} \geq P\{X_j \geq c_j,\ j \in B\}P\{X_j \geq c_j,\ j \in \bar{B}\}$,

(ii) $P\{X_j \leq c_j,\ j \in A\} \geq P\{X_j \leq c_j,\ j \in B\}P\{X_j \leq c_j,\ j \in \bar{B}\}$,

(iii) $P\{X_i \geq c_i,\ i \in B;\ X_j \leq c_j,\ j \in \bar{B}\}$
$\leq P\{X_i \geq c_i,\ i \in B\}P\{X_j \leq c_j,\ j \in \bar{B}\}$.

SPOD is a generalization to multivariate $\mathbf{X}$ of the concept of positive quadrant dependence for bivariate $\mathbf{X}$ as discussed by Lehmann (1966). A characterization of binary SPOD distributions in terms of latent variable models is given by Holland (1981, Theorem 2).

The following theorem summarizes the well-known relationships between these three types of positive dependence. The first implication of Theorem 3 is given by Karlin and Rinott (1980, Corollary 4.1), and is related to the FKG inequality (Fortuin, Kasteleyn and Ginibre, 1971). The second implication in Theorem 3 is obvious.

THEOREM 3 (Karlin and Rinott, 1980). *If a random vector* $\mathbf{X}$ *is* MTP_2 *then* $\mathbf{X}$ *is associated, which in turn implies that* $\mathbf{X}$ *is SPOD.*

The implications in Theorem 3 are strict in general; counterexamples to their reversal appear in Esary, Proschan and Walkup (1967).

3.2. *Three stronger forms of positive association.* Rosenbaum (1984) defines a stronger form of Definition 3.1, which we call *conditionally associated* random variables. Parallel stronger forms of Definitions 3.2 and 3.3 are conditional MTP_2 and conditional SPOD. It is these stronger forms of MTP_2, association and SPOD that are relevant to the study of unidimensional monotone latent variable models.

The general form of each of these stronger "conditional" types of positive association is the same. In each case we will consider all possible ways of partitioning (and rearranging) the random vector $\mathbf{X}$ into two components which we denote by

$$\mathbf{X} = (\mathbf{Y}, \mathbf{Z}). \tag{3.4}$$

Then the positive association condition is required to hold (almost surely) for the conditional distribution of $\mathbf{Y}$ given any (measurable) function $\mathbf{h}(\mathbf{Z})$ of $\mathbf{Z}$. We give the formal definitions of $CMTP_2$, CA and CSPOD now.

DEFINITION 3.4 (CA). *The distribution of a random vector* $\mathbf{X}$ *is conditionally associated if, for any partition* $(\mathbf{Y}, \mathbf{Z})$ *of* $\mathbf{X}$ *and any function* $\mathbf{h}(\mathbf{Z})$, *the conditional distribution of* $\mathbf{Y}$ *given* $\mathbf{h}(\mathbf{Z})$ *is associated.*

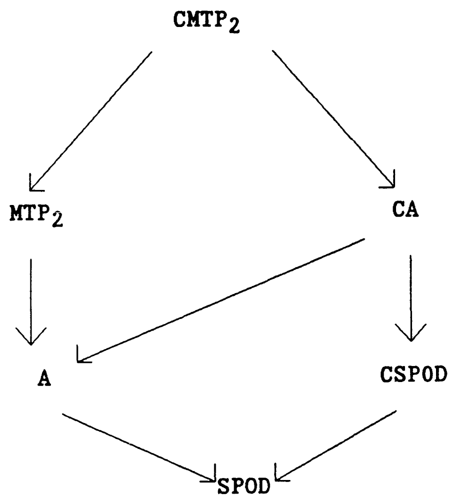

FIG. 1. *Summary of interrelationships among six definitions of positive multivariate association.*

DEFINITION 3.5 (CMTP_2). *The distribution of a random vector* $\mathbf{X}$ *is conditionally* MTP_2 *if, for any partition* $(\mathbf{Y}, \mathbf{Z})$ *of* $\mathbf{X}$ *and any function* $\mathbf{h}(\mathbf{Z})$, *the conditional distribution of* $\mathbf{Y}$ *given* $\mathbf{h}(\mathbf{Z})$ *is* MTP_2.

DEFINITION 3.6 (CSPOD). *The distribution of a random vector* $\mathbf{X}$ *is conditionally SPOD if, for any partition* $(\mathbf{Y}, \mathbf{Z})$ *of* $\mathbf{X}$ *and any function* $\mathbf{h}(\mathbf{Z})$, *the conditional distribution of* $\mathbf{Y}$ *given* $\mathbf{h}(\mathbf{Z})$ *is SPOD.*

It is evident that Theorem 3 implies a corresponding theorem which orders the three "conditional" positive association definitions. Again, the implications are strict. It is stated as Theorem 4 which follows.

THEOREM 4. *If a random vector* $\mathbf{X}$ *is* $CMTP_2$ *then* $\mathbf{X}$ *is CA which in turn implies that* $\mathbf{X}$ *is CSPOD.*

Figure 1 summarizes the interrelationships among the various types of positive dependence. MTP_2 and CA are not comparable in the sense that neither implies the other in general.

In Section 5, we will show that in the special case where $\mathbf{X}$ is a 1/0 random vector the three conditions, CMTP_2, CA and CSPOD, are all equivalent. No such equivalence holds for MTP_2, A or SPOD in the J-dimensional, binary case.

4. Positive association in monotone latent variable models. The principal conclusion of this section is that multidimensional latent variable models lead to positive association for $\mathbf{X}$ when $\mathbf{U}$ is itself positively associated; however, a unidimensional model leads to (stronger) forms of positive conditional association for $\mathbf{X}$. Our strongest, and probably most useful, results concern unidimensional, monotone latent variable models. For this reason we discuss the unidimensional case first.

4.1. *Conditional association and unidimensionality.* Our first result concerns the strongest of the positive association conditions, $CMTP_2$. It is given in the following theorem.

THEOREM 5. *If a latent variable model* $(\mathbf{X}, U)$ *satisfies the conditions of latent conditional independence and latent unidimensionality and, in addition, is latent* TP_2, *then the distribution of* $\mathbf{X}$ *is* $CMTP_2$.

PROOF. Let $(\mathbf{Y}, \mathbf{Z})$ be any partition of $\mathbf{X}$, let $\mathbf{h}(\mathbf{Z})$ be arbitrary, let the conditional density or mass function of $\mathbf{Y}$ given $\mathbf{h}(\mathbf{Z}) = \mathbf{t}$ be denoted by $p(\mathbf{y}|\mathbf{t})$, and let $\sigma_{\mathbf{t}}(u)$ denote the conditional probability measure of the scalar U given $\mathbf{h}(\mathbf{Z}) = \mathbf{t}$. Then by latent conditional independence,

$$p(\mathbf{y}|\mathbf{t}) = \int \prod_j f_j(y_j|u)\sigma_{\mathbf{t}}(du), \tag{4.1}$$

where $f_j(y_j|u)$ is the conditional density or mass function of Y_j given $U = u$. Now $\prod f_j(y_j|u)$ is MTP_2 in $(\mathbf{Y}, U)$ by (2.5) and the fact that products of TP_2 functions are MTP_2 (e.g., Karlin and Rinott, 1980, Proposition 3.3). That $p(\mathbf{y}|\mathbf{t})$ is MTP_2 for fixed $\mathbf{t}$ follows immediately from (4.1) and Proposition 3.4 of Karlin and Rinott (1980), with their $g(\cdot)$ identically equal to 1. □

The next theorem, which applies to a larger class of latent variable models, gives the result for CA.

THEOREM 6. *If a latent variable model* $(\mathbf{X}, U)$ *satisfies the conditions of latent conditional independence and latent unidimensionality, and, in addition, is monotone then the distribution of* $\mathbf{X}$ *is CA.*

PROOF. The proof is the same as Rosenbaum's (1984) proof of his Theorem 1. Since it is brief and provides insight we repeat it here. We need to show that

$$E(g_1(\mathbf{Y})g_2(\mathbf{Y})|\mathbf{h}(\mathbf{Z})) \geq E(g_1(\mathbf{Y})|\mathbf{h}(\mathbf{Z}))E(g_2(\mathbf{Y})|\mathbf{h}(\mathbf{Z})),$$

where $(\mathbf{Y}, \mathbf{Z})$ is any partition of $\mathbf{X}$, $\mathbf{h}(\mathbf{Z})$ is arbitrary, and $g_1(\cdot)$, $g_2(\cdot)$ are bounded and nondecreasing. The proof uses the two facts that (i) the coordinates of $\mathbf{Y}$ are independent given U and are therefore associated given U, and (ii) U is a scalar and is therefore associated. Clearly we have

$$E(g_1(\mathbf{Y})g_2(\mathbf{Y})|\mathbf{h}(\mathbf{Z})) = E(E[g_1(\mathbf{Y})g_2(\mathbf{Y})|U]|\mathbf{h}(\mathbf{Z})),$$

from latent conditional independence of $\mathbf{Y}$ and $\mathbf{Z}$ given U. Therefore,

$$E(g_1(\mathbf{Y})g_2(\mathbf{Y})|\mathbf{h}(\mathbf{Z})) \geq E(E[g_1(\mathbf{Y})|U]E[g_2(\mathbf{Y})|U]|\mathbf{h}(\mathbf{Z})),$$

from the fact that the coordinates of $\mathbf{Y}$ are independent given U and independent random variables are associated. Thus,

$$E(g_1(\mathbf{Y})g_2(\mathbf{Y})|\mathbf{h}(\mathbf{Z})) \geq E(E[g_1(\mathbf{Y})|U]|\mathbf{h}(\mathbf{Z}))E(E[g_2(\mathbf{Y})|U]|\mathbf{h}(\mathbf{Z})),$$

from Lemma 2 and the fact that U is a scalar and scalar random variables are associated. Hence

$$E(g_1(\mathbf{Y})g_2(\mathbf{Y})|\mathbf{h}(\mathbf{Z})) \geq E(g_1(\mathbf{Y})|\mathbf{h}(\mathbf{Z}))E(g_2(\mathbf{Y})|\mathbf{h}(\mathbf{Z})),$$

from latent conditional independence and the definition of conditional expectation. □

Theorems 5 and 6 show that unidimensional monotone latent variable models place strong and testable conditions on the distribution of the observable data, **X**. (See Sections 6.2 and 6.3 for examples.) An interesting open question is whether every observable distribution of **X** that is either CMTP_2 or CA can be given a monotone unidimensional latent variable representation. In the case of binary **X**, Holland (1981) has shown that the weaker form of positive dependence, SPOD, does imply that a latent variable representation exists for **X** within a larger class of latent variable models. In this larger class of unidimensional latent variable models, conditional independence is replaced by a more general condition called "local nonnegative dependence."

4.2. *Positive association and multidimensional latent variables.* If **U** is allowed to be a vector of arbitrary dimension with an arbitrary distribution $F(\mathbf{u})$, satisfying no restrictions, then any multivariate random vector **X** has a representation as a *monotone* latent variable model satisfying latent conditional independence. A referee pointed out that this is easily seen by taking **U** to be **X** itself. In contrast, when the latent vector **U** is restricted to be MTP_2 or associated, then the observable distribution of **X** itself is restricted. This section states two such results.

THEOREM 7 (Karlin and Rinott, 1980). *If a latent variable model* $(\mathbf{X}, \mathbf{U})$ (i) *satisfies the condition of latent conditional independence,* (ii) *has conditional densities* $f_j(x_j|\mathbf{u})$ *that are* MTP_2, *as functions of* $(x_j, \mathbf{u})$, *and if in addition* (iii) *the distribution of* **U** *is* MTP_2, *then the distribution of* **X** *is* MTP_2.

PROOF. Follows from applying Proposition 3.4 of Karlin and Rinott (1980) to the density of $\mathbf{X}|\mathbf{U}$ as their f and the density of **U** as their g. □

The second result was proved by Jogdeo (1978).

THEOREM 8 (Jogdeo, 1978). *If a latent variable model* $(\mathbf{X}, \mathbf{U})$ *satisfies the conditions of* (i) *latent conditional independence and* (ii) *latent monotonicity and if in addition* (iii) *the distribution of* **U** *is associated then the distribution of* **X** *is associated.*

4.3. *Unidimensionality amid multidimensionality.* If $\mathbf{U}$ is of dimension two or more, then it is possible that some subset of the coordinates of $\mathbf{X}$ depend only on a single coordinate of $\mathbf{U}$. This subset is "unidimensional," although $\mathbf{X}$ as a whole is not. Let $(\mathbf{Y}, \mathbf{Z})$ be a partition of $\mathbf{X}$ into two nonoverlapping sets of coordinates, and suppose that $(\mathbf{X}, \mathbf{U})$ has a latent variable representation which need not be monotone. Then $\mathbf{Y}$ is monotone and unidimensional if for each coordinate Y_i of $\mathbf{Y}$

$$P(Y_i > y | \mathbf{U} = \mathbf{u}) = P(Y_i > y | U_1 = u_1) \tag{4.2}$$

is nondecreasing in u_1 for each y, where U_1 is the first coordinate of $\mathbf{U}$. The following extension of Theorem 6 may be proved by a parallel argument.

THEOREM 9. *If a multidimensional latent variable model for* $(\mathbf{X}, \mathbf{U})$ *satisfies latent conditional independence, and if the subset* $\mathbf{Y}$ *of coordinates of* $\mathbf{X}$ *is monotone and unidimensional in the sense that* (4.2) *holds, then the conditional distribution of* $\mathbf{Y}$ *given* $\mathbf{h}(\mathbf{Z})$ *is CA for any choice of the function* $\mathbf{h}(\cdot)$, *where* $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$.

Informally, Theorem 9 states that a monotone, unidimensional subset, $\mathbf{Y}$, of the coordinates of a multidimensional $\mathbf{X}$ exhibit conditional association in every subpopulation defined by any function of the remaining coordinates of $\mathbf{X}$.

5. The case of binary X.

Our own work on latent variable models grows out of the binary case, i.e., the X_j are 0/1 variables. This is due to the wide applicability of such models to data from educational tests. The results of Section 4 have a simplicity in the binary case that we wish to emphasize in this section.

In the binary case, the three conditions CMTP_2, CA and SPOD are equivalent. This is proved in the next theorem, whose conclusion is summarized in Figure 2.

THEOREM 10. *If X is a binary random vector then the following three conditions are equivalent*:

(i) $\mathbf{X}$ *is* $CMTP_2$,
(ii) $\mathbf{X}$ *is CA*,
(iii) $\mathbf{X}$ *is CSPOD*.

PROOF. Because of Theorem 4 and Figure 1 it is sufficient to show that CSPOD implies CMTP_2. Select a partition $(\mathbf{Y}, \mathbf{Z})$ of $\mathbf{X}$ and a function $\mathbf{h}(\cdot)$. We need to show that CSPOD implies that for each fixed $\mathbf{y}, \mathbf{y}^*$

$$\begin{aligned} &P\{\mathbf{Y} = \max(\mathbf{y}, \mathbf{y}^*)\} | \mathbf{h}(\mathbf{Z})\} P\{\mathbf{Y} = \min(\mathbf{y}, \mathbf{y}^*) | \mathbf{h}(\mathbf{Z})\} \\ &\quad \geq P\{\mathbf{Y} = \mathbf{y} | \mathbf{h}(\mathbf{Z})\} P\{\mathbf{Y} = \mathbf{y}^* | \mathbf{h}(\mathbf{Z})\}. \end{aligned} \tag{5.1}$$

Partition and rearrange $\mathbf{Y}$ into $(\mathbf{W}_0, \mathbf{W}_1, \mathbf{W}_2)$ where (a) $\mathbf{W}_0$ contains those coordinates of $\mathbf{Y}$ for which $y_i = 0$ and $y_i^* = 1$, (b) $\mathbf{W}_1$ contains those coordinates of $\mathbf{Y}$

$$\begin{tikzcd} \mathbf{CSPOD} \arrow[r, leftrightarrow] & \mathbf{CA} \arrow[r, leftrightarrow] \arrow[d] & \mathbf{CMTP_2} \\ & \mathbf{MTP_2} \arrow[d] & \\ & \mathbf{A} \arrow[d] & \\ & \mathbf{SPOD} & \end{tikzcd}$$

FIG. 2. *The case of binary* $\mathbf{X}$: *interrelationships among definitions of positive association.*

with $y_i = 1$, $y_i^* = 0$ and (c) $\mathbf{W}_2$ contains those coordinates of $\mathbf{Y}$ with $y_i = y_i^*$. Assuming CSPOD, it follows that

$$P\{\mathbf{W}_0 = \mathbf{1}, \mathbf{W}_1 = \mathbf{1} | \mathbf{W}_2, \mathbf{h}(\mathbf{Z})\} P\{\mathbf{W}_0 = \mathbf{0}, \mathbf{W}_1 = \mathbf{0} | \mathbf{W}_2, \mathbf{h}(\mathbf{Z})\}$$
$$\geq P\{\mathbf{W}_0 = \mathbf{0}, \mathbf{W}_1 = \mathbf{1} | \mathbf{W}_2, \mathbf{h}(\mathbf{Z})\} P\{\mathbf{W}_0 = \mathbf{1}, \mathbf{W}_1 = \mathbf{0} | \mathbf{W}_2, \mathbf{h}(\mathbf{Z})\},$$

which implies (5.1). □

If we now apply Theorem 5 or 6 to the binary case, we obtain the following corollary which gives our strongest result for testing binary response models for latent unidimensionality.

COROLLARY 11. *If* $\mathbf{X}$ *is binary and* $(\mathbf{X}, U)$ *is a unidimensional, monotone latent variable model then the distribution of* $\mathbf{X}$ *is CSPOD or equivalently CA or equivalently* $CMTP_2$.

In a way that is quite different from Corollary 11, Stout (1986) also develops necessary conditions for unidimensional, monotone latent variable models when $\mathbf{X}$ is binary.

6. Examples.

6.1. *A Counterexample*: *MTP_2 does not imply CSPOD, CA or $CMTP_2$*. Table 2 displays a distribution that is MTP_2 but does not satisfy the conditions CSPOD, CA or $CMTP_2$. An even simpler example of this phenomenon assigns probability 1/4 to each of the four (X_1, X_2, X_3, X_4) vectors $(1,1,1,1)$, $(0,1,0,1)$, $(1,0,1,0)$ and $(0,0,0,0)$, so that the conditional distribution of (X_3, X_4) given $X_1 + X_2 = 1$ is perfectly negatively associated.

6.2. *Some implications of positive conditional association for the examples of Section* 2.5.

(i) *Item response theory.* By either Theorem 5 or Theorem 6, if a unidimensional monotone item response model is to describe the joint distribution of J dichotomously scored exam items, then every pair of items (X_i, X_j) must have a nonnegative (population) correlation. But not only must these first-order correlations be nonnegative, many *conditional* correlations must be nonnegative as well. For example, there must be a nonnegative correlation between X_i and X_j among

TABLE 2
A counterexample: a distribution that is MTP_2 but not CPSPOD, CA or $CMTP_2$

a. *The joint MTP_2 distribution of four dichotomous variables (to obtain probabilities, divide each count by* 288*)*

X_1	X_2	X_3	X_4 = 1	X_4 = 0
1	1	1	27	27
1	1	0	9	9
1	0	1	12	12
1	0	0	24	24
0	1	1	18	36
0	1	0	6	12
0	0	1	8	16
0	0	0	16	32

b. *The conditional distribution of (X_3, X_4) given $X_1 + X_2 = 1$ (to obtain probabilities, divide each count by* 144*)*

		X_4 = 1	X_4 = 0
X_3	1	30	48
	0	30	36

$$\text{Odds-ratio} = \frac{30 \times 36}{30 \times 48} = \frac{3}{4} < 1$$

all examinees with a given total score on the remaining items; i.e.,

$$\operatorname{Cov}\left(X_i, X_j \middle| \sum_{k \neq i, j} X_k\right) \geq 0. \tag{6.1}$$

Equivalently, the (population) odds ratio must be at least 1 in each 2×2 subtable of the $2 \times 2 \times (J-1)$ contingency table recording the values of X_i by X_j by $\sum_{k \neq i, j} X_k$. This observation has been used in Rosenbaum (1984) to test the unidimensionality/conditional-independence/monotonicity assumptions of item response theory in the College Board's 1982 Advanced Placement Examination in Biology; see also Holland (1981) for closely related procedures. Rosenbaum (1985) develops a related method for judging whether the item response patterns in two populations of examinees can be represented by a monotone unidimensional model with a higher distribution of U in one population.

(ii) *Systems reliability theory.* A pattern of association among component failures that is commonly used in deriving bounds on the reliability of a system (Esary and Proschan, 1970) does result from a monotone latent variable model of the type described here. Specifically, if (2.7) holds for a vector of latent stresses whose distribution is associated and with each $r_j(\mathbf{u})$ nondecreasing in $\mathbf{U}$, then the component failures are themselves associated (i.e., Theorem 8). When the latent stress is unidimensional so that Theorems 5 and 6 apply, similar bounds may often be obtained for system reliability conditional on the state of certain components or subsystems.

(iii) *Linear factor analysis for multivariate normal* $\mathbf{X}$. Under the unidimensional monotone linear factor model—i.e., under (2.8) with scalar U and $\Lambda \geq 0$—all partial correlations between pairs of coordinates of $\mathbf{X}$ given any set of linear functions of the remaining coordinates are nonnegative. This condition will imply that all first-order correlations among the coordinates of $\mathbf{X}$ must be nonnegative, but it is a much stronger condition than this. Violations of these conditions indicate that the monotone unidimensional model does not hold.

(iv) *Population genetics*: *segregation analysis.* Since many genetic disorders are comparatively rare, it is common in studies of human sibships to examine only sibships containing affected individuals. Various methods of obtaining such sibships—so-called "methods of ascertainment"—are common (e.g., Levitan and Montagu, 1977, Chapter 10 and Elandt-Johnson, 1971, Chapters 17 and 18). One such method is single selection: all children in a population containing at most one sib from each sibship (such as the fifth grade of a particular school system) are screened for the trait (or phenotype); whenever the trait is found, the entire sibship of the affected child is added to the study. In this formulation, there is at most a single identifiable child in each sibship who can lead to inclusion of the sibship; this child is called the proband or index case. Let X_1 indicate the presence (1) or absence (0) of the trait in the proband, and let $X_2, \ldots, X_J$ indicate the presence or absence of the trait in the remaining children of the sibship, arranged from youngest (X_2) to oldest (X_J). Ascertainment by single selection is, in effect, sampling conditional on $X_1 = 1$. Our results in Section 4

TABLE 3
A cross tabulation of sibships exhibiting positive partial association under single selection

Number of Other Sibs Affected	Youngest Sib Excluding the Proband (X_2)	Oldest Sib Excluding the Proband (X_J)	
		$X_J = 1$	$X_J = 0$
$\sum_{j=3}^{J-1} X_j = 0$	$X_2 = 1$	n_{110}	n_{100}
	$X_2 = 0$	n_{010}	n_{000}
	$\vdots$		
		$X_J = 1$	$X_J = 0$
$\sum_{j=3}^{J-1} X_j = i$	$X_2 = 1$	n_{11i}	n_{10i}
	$X_2 = 0$	n_{01i}	n_{00i}
	$\vdots$		
		$X_J = 1$	$X_J = 0$
$\sum_{j=3}^{J-1} X_j = J - 3$	$X_2 = 1$	$n_{11,J-3}$	$n_{10,J-3}$
	$X_2 = 0$	$n_{01,J-3}$	$n_{00,J-3}$

Notes: (a) Let $m_{abc} = E(n_{abc})$, where $E(\cdot)$ denotes expectation under (2.10). (b) Positive dependence implies $m_{11i} \cdot m_{00i} \geq m_{10i} \cdot m_{01i}$. (c) Symmetry implies $m_{10i} = m_{01i}$.

show that under the model (2.10)

(a) $\mathbf{X}$ is CMTP_2 in the population, and
(b) the distribution of $X_2, X_3, \ldots, X_J$ given that $X_1 = 1$ is CMTP_2 in the sample obtained by single selection.

Thus, single selection does not eliminate the strong positive dependence within sibships. In particular, (2.10) implies nonnegative association (and symmetry) in each of the 2×2 contingency tables in Table 3: a departure from nonnegative association and symmetry would indicate a departure from a strictly genetic model (2.10), perhaps indicating effects of:

(a) parental age of conception (e.g., Stene and Stene, 1979), possibly resulting in asymmetry (e.g., $m_{01i} > m_{10i}$),
(b) an exposure over a short time period of the population to an environmental hazard affecting prenatal development, possibly resulting in few sibships with both oldest and youngest affected (i.e., $m_{11i} \cdot m_{00i} \leq m_{10i} \cdot m_{01i}$), or
(c) in studies of behavioral traits, the effects of being an older or younger sib.

With sibships of randomly varying sizes—that is, with X's of varying dimensions, say $J = 3, \ldots, M$ with $P(J = j | X_1, X_2, \ldots, X_i) = P(J = j)$ for $j \geq i$—we may concatenate the one table for $J = 3$, the two tables for $J = 4$, the three tables for $J = 5$, etc., forming a single $2 \times 2 \times R$ table, in which each 2×2 slice is positively dependent and symmetric. Clearly, the above argument does not require ordering $(X_2, X_3, \ldots, X_J)$ by age within sibships; any order involving nongenetic factors may be used to test (2.10). Sibships would not have randomly varying sizes if the manifestation of the trait reduced the chance of having additional children.

(v) *Latent class models for dichotomous responses.* To obtain a unidimensional monotone latent variable model from a latent class model with two classes—i.e., a model with $|S| = 2$—the coordinates of $\mathbf{X}$ must be relabeled so $P(X_j = 1 | U = 1) \geq P(X_j = 1 | U = 0)$ for all j. By Theorem 6, with a sufficiently large sample, this relabeling may be carried out in practice even though U is unobserved: simply label the X_j's so that the part–whole covariance, $\operatorname{cov}(X_j, \sum_{i=1}^{J} X_i)$, is positive for all j. If no such relabeling exists, or if, with such a relabeling, there are other violations of the positive dependence implied by Theorems 5 and 6, then no latent class model with two classes can describe the observable distribution of $\mathbf{X}$. (Some ambiguity would result even in large samples if the population part–whole covariance is exactly zero for some j, but this is not a serious problem in the applications with which we are familiar.)

6.3. *Do multiple choice items and an essay score measure a unidimensional latent variable?* As an illustration, we apply the results of Section 4 to a practical example for which available methods are inadequate. There is, in educational measurement, a debate concerning the use of essays in national testing programs. On one side is the view that essays and multiple-choice items measure different skills and abilities, that the ability to organize thoughts and write about them cannot be tested using multiple choice items, and there is even the fear in some circles that if testing programs stopped using essays, some teachers would stop teaching students to write. Against this is the uneasiness of individuals observing the operational aspects of scoring essays, who question the reliability of the essay scores given by armies of part-time essay readers who score tens of thousands of essays in a couple of weeks. Even the detailed scoring instructions and training given to essay readers do not necessarily remove the question of the reliability of essay grading.

The question, then, may be posed as follows: Do essays and multiple choice items measure the same thing? Or, alternatively, must we use essays, despite their operational difficulties, to measure certain abilities that are measured by essays but not by multiple choice items? This question can be formalized by asking whether, within some family of monotone latent variable models for both item responses and essay scores, a unidimensional model provides an adequate description of empirical distributions. If it does, then one would tend to question the added value of measures that involve essays.

To illustrate our results within this context, we examined the joint distributions of responses to the 40 dichotomous multiple choice items (1 = correct, 0 = incorrect) on the population biology subscore of the College Board's 1982 Advanced Placement Examination in Biology, together with the ordinal 15-point-scale response to that exam's essay #6, also on population biology. A total of 11,533 examinees wrote essay #6. In this case, $\mathbf{X} = (X_1, X_2, \ldots, X_{41})$ where $X_1, X_2, \ldots, X_{40}$ are binary valued, X_{41} takes values in the set $\{1, 2, \ldots, 15\}$, and

11,533 observations on $\mathbf{X}$ are available. Although there are no widely used latent variable models for combining essay and multiple-choice responses of this sort, we may nonetheless check whether $\mathbf{X}$ violates the condition of conditional association: if it does, then no *monotone* unidimensional latent variable model could describe the data no matter what parametric form it takes. Monotonicity is a natural condition to apply both to dichotomous multiple choice items and to scores on the essay. The IRT models of Bock (1972) and Masters (1982) for nominal and ordered responses can be applied to mixtures of essays and multiple choice tests. However, these models assume specific parametric forms and therefore might be rejected in an empirical test because of the inadequacy of the functional forms assumed rather than for lack of unidimensionality. Masters' model is an example of one that is latent TP_2. Bock's model is not TP_2 unless the item parameters are restricted. Samejima's graded response model (Samejima, 1969) is a monotone model.

To this end, we constructed 120 contingency tables of dimensions $2 \times 2 \times 40$, recording for $i = 1, 2, \dots, 40$ and $j = 4, 6, 8$ the joint distribution of $\{X_i, [X_{41} \le j], \Sigma_{k=1, k \neq i}^{40} X_k\}$, where [event A] denotes the indicator of the event A. If $\mathbf{X}$ were conditionally associated, the *population* odds ratio in each 2×2 slice of each of the 120 $2 \times 2 \times 40$ tables would be greater than or equal to one. For each of the 120 tables, we calculated the Mantel–Haenszel weighted combination of odds ratios (cf., Breslow, 1981), producing one combined odds ratio per table. Of the 120 combined odds ratios, nine were less than one. Three of those nine had individual p-values less than .1 in a test of the null hypothesis of positive association, with p-values of 0.008, 0.03 and 0.007, for item $i = 5$ and cutpoints $j = 4, 6, 8$, respectively. Given that 120 statistical tests have been performed, this is rather marginal evidence that at most item 5 and the essay are not measuring a unidimensional variable.

We then repeated this process for the $\binom{40}{2} = 780$ pairs of items in the population biology subscore, grouping by the total score on the remaining items; i.e., we looked for negative partial associations in the 780 $2 \times 2 \times 39$ tables recording $\{X_i, X_j, \Sigma_{k=1, k \neq i, j}^{40} X_k\}$. A negative partial association among these variables in the *population* would indicate that $\{X_1, X_2, \dots, X_{40}\}$ is not conditionally associated, and therefore that no monotone unidimensional latent variable model could adequately describe the observable distribution of the item responses $\{X_1, X_2, \dots, X_{40}\}$. Here we found a number of negative partial associations, including five with p-values less than $0.000013 = 0.01/780$ and an additional four with p-values below $0.000013 = 0.05/780$. Using the Bonferroni inequality (e.g., Miller, 1980, page 8), it is reasonable to judge a number of the p-values to be surprisingly small, despite the large number of significance tests. There is, then, rather strong evidence that the item responses themselves are not unidimensional. Thus, this analysis suggests that there is more evidence in these data that the multiple choice items are not all measuring the same thing than there is that the essay measures something different than do the multiple choice items. This example was chosen to illustrate our method and, of course, does not constitute a definite resolution of the essay/multiple choice debate. It does, however, suggest that quantitative methods can be brought to bear on the question.

An interesting open question is whether better methods exist for testing the CA condition. Two possibilities deserve consideration. First, is there a better choice for $\mathbf{h}(\mathbf{Z})$ than $\Sigma_{k \neq i,j} X_k$? Stout (1986), for example, suggests grouping examinees on the basis of an index from a preliminary factor analysis. Second, can a single superior test for CA be obtained without applying many separate hypothesis tests followed by a correction for multiplicity?

Acknowledgments. We would like to thank Robert Mislevy, Frank Proschan and William Stout for helpful comments on an earlier draft.

REFERENCES

AHMED, A., LEON, R. and PROSCHAN, F. (1981). Generalized association, with applications in multivariate statistics. *Ann. Statist.* **9** 168–176.

ANDERSEN, E. B. (1980). *Discrete Statistical Models with Social Science Applications.* North-Holland, Amsterdam.

BARLOW, R. E. and PROSCHAN, F. (1975). *Statistical Theory of Reliability and Life Testing: Probability Models.* Holt, Rinehart and Winston, New York.

BARTHOLOMEW, D. J. (1980). Factor analysis for categorical data (with discussion). *J. Roy. Statist. Soc. Ser. B* **42** 293–321.

BARTHOLOMEW, D. J. (1984). The foundations of factor analysis. *Biometrika* **71** 221–232.

BIRNBAUM, A. (1968). Some latent trait models and their use in inferring an examinee's ability (Part 5). In *Statistical Theories of Mental Test Scores* (F. Lord and M. Novick). Addison-Wesley, Reading, Mass.

BOCK, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika* **37** 29–51.

BOCK, R. D. and LIEBERMAN, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika* **35** 179–197.

BRESLOW, N. E. (1981). Odds ratio estimators when the data are sparse. *Biometrika* **68** 73–84.

CRESSIE, N. and HOLLAND, P. (1983). Characterizing the manifest probabilities of latent trait models. *Psychometrika* **48** 129–141.

CROW, J. F. and KIMURA, M. (1970). *An Introduction to Population Genetics Theory.* Burgess, Minneapolis, Minn.

EATON, M. L. (1982). A review of selected topics in multivariate probability inequalities. *Ann. Statist.* **10** 11–43.

ELANDT-JOHNSON, R. (1971). *Probability Models and Statistical Methods in Genetics.* Wiley, New York.

ESARY, J. D. and PROSCHAN, F. (1970). A reliability bound for systems of maintained, interdependent components. *J. Amer. Statist. Assoc.* **65** 329–338.

ESARY, J. D., PROSCHAN, F. and WALKUP, D. W. (1967). Association of random variables, with applications. *Ann. Math. Statist.* **38** 1466–1474.

FORTUIN, C. M., KASTELEYN, P. W. and GINIBRE, J. (1971). Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.* **22** 89–103.

GOLDSTEIN, H. (1980). Dimensionality, bias, independence and measurement scale problems in latent test score models. *British J. Math. Statist. Psych.* **33** 234–246.

GOODMAN, L. (1974). Exploratory latent-structure analysis using both identifiable and unidentifiable models. *Biometrika* **61** 215–231.

HOLLAND, P. W. (1981). When are item response models consistent with observed data? *Psychometrika* **46** 79–92.

JOAG-DEV, K. (1983). Independence via uncorrelatedness under certain dependence structures. *Ann. Probab.* **11** 1037–1041.

JOGDEO, K. (1978). On a probability bound of Marshall and Olkin. *Ann. Statist.* **6** 232–234.

KAMAE, T., KRENGEL, U. and O'BRIEN, G. L. (1977). Stochastic inequalities on partially ordered spaces. *Ann. Probab.* **5** 899–912.

KARLIN, S. (1968). *Total Positivity.* Stanford Univ. Press, Stanford, Calif.

KARLIN, S. and RINOTT, Y. (1980). Classes of orderings of measures and related correlation inequalities, I. Multivariate totally positive distributions. *J. Multivariate Anal.* **10** 467–498.

KARLIN, S. and RINOTT, Y. (1983). M-matrices as covariance matrices of multinormal distributions. *Linear Algebra Appl.* **52/53** 419–438.

KINGMAN, J. F. C. (1978). Uses of exchangeability. *Ann. Probab.* **6** 183–197

LAWLEY, D. N. and MAXWELL, A. E. (1971). *Factor Analysis as a Statistical Method.* Elsevier, New York.

LEHMANN, E. L. (1955). Ordered families of distributions. *Ann. Math. Statist.* **26** 399–419.

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

LEVITAN, M. and MONTAGU, A. (1977). *A Textbook of Human Genetics.* Oxford, New York.

LORD, F. (1980). *Applications of Item Response Theory to Practical Testing Problems.* Lawrence Erlbaum, Hillsdale, N. J.

LORD, F. (1982). Item response theory and equating—A technical summary. In *Test Equating* (P. W. Holland and D. B. Rubin, eds.) 141–148. Academic, New York.

LORD, F. and NOVICK, M. (1968). *Statistical Theories of Mental Test Scores.* Addison-Wesley, Reading, Mass.

MARSHALL, A. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Prentice-Hall, Englewood Cliffs, N. J.

MASTERS, G. (1982). A Rasch model for partial credit scoring. *Psychometrika* **47** 149–174.

MILLER, R. (1980). *Simultaneous Statistical Inference*, 2nd ed. Springer, New York.

PRESTON, C. J. (1974). *Gibbs States on Countable Sets.* Cambridge, New York.

RASCH, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests.* Neilson and Lydiche, Copenhagen.

ROSENBAUM, P. R. (1984). Testing the conditional independence and monotonicity assumptions of item response theory. *Psychometrika* **49** 425–436.

ROSENBAUM, P. R. (1985). Comparing distributions of item responses for two groups. *British J. Math. Statist. Psych.* **38** 206–215.

RUBIN, D. B. (1982). Discussion of "Observed-score test equating: A mathematical analysis of some ETS equating procedures." In *Test Equating* (P. W. Holland and D. B. Rubin, eds.) 51–54. Academic, New York.

SAMEJIMA, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement* **17**.

STENE, J. and STENE, E. (1979). Down's syndrome and father's age. *Scand. J. Statist.* **6** 180–181.

STOUT, W. (1986). A theory-based nonparametric approach for assessing latent trait unidimensionality in psychological testing. *Psychometrika.* To appear.

SUPPES, P. and ZANOTTI, M. (1981). When are probabilistic explanations possible? *Synthese* **48** 191–199.

TJUR, T. (1982). A connection between Rasch's item analysis model and a multiplicative Poisson model. *Scand. J. Statist.* **9** 23–30.

TONG, Y. L. (1980). *Probability Inequalities in Multivariate Distributions.* Academic, New York.

RESEARCH STATISTICS GROUP
EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY 08541

DEPARTMENT OF STATISTICS
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104