Conditional Association, Essential Independence and Monotone Unidimensional Item Response Models

Brian W. Junker

Department of Statistics
Carnegie Mellon University
Pittsburgh PA 15213

July 17, 1992

Research supported by ONR grants N00014-87-K-0277 and N00014-90-J-1984, and NIMH grant MH15758.

AMS 1991 subject classifications. Primary 62E10; secondary 62P15, 60E15, 62G20.

Key words and phrases. strict unidimensionality, essential unidimensionality, useful latent variable models, conditional association, negative association, empirical response curves, monotonicity.

Abstract

We consider two recent approaches to characterizing the manifest probabilities of a strictly unidimensional latent variable representation (one satisfying local independence and response curve monotonicity with respect to a unidimensional latent variable) for binary response variables, such as those arising from the dichotomous scoring of items on standardized achievement and aptitude tests. Holland and Rosenbaum (1986) show that *conditional association* is a necessary condition for strict unidimensionality; and Stout (1990) treats the class of *essentially unidimensional* models, in which the latent variable may be consistently estimated as the length of the response sequence grows using the proportion of positive responses. Of particular concern are strictly unidimensional representations that are minimally useful in the sense that (1) the latent variable can be consistently estimated from the responses; (2) the regression of proportion of positive responses on the latent variable is monotone; and (3) the latent variable is not constant in the population. We introduce two new conditions, a negative association condition and a natural monotonicity condition on the empirical response curves, that help link strict unidimensionality with the conditional association and essential unidimensionality approaches. These conditions are illustrated with a partial characterization of useful, strictly unidimensional representations.

1. Introduction. *Item response theory*, IRT, specializes latent variable models, for example as discussed by Holland and Rosenbaum (1986), to examinee responses to questions—items—on standardized achievement or aptitude tests. Widespread interest in binary (0/1) item response models was stimulated by Birnbaum (1968) and the more recent survey of Lord (1980). These and related models are also used in other applications such as medical diagnosis and psychiatric epidemiology (Eaton and Bohrnstedt, 1989), multiple recapture methods for estimating population sizes (Sanathanan, 1972; Chao, 1987; Darroch, Fienberg, Glonek and Junker, 1991), as well as systems reliability and population genetics as surveyed by Holland and Rosenbaum (1986). Although latent variable methods are also used to study polytomous and continuous response data (cf. e.g. Bartholomew, 1987), binary response models still predominate in applications (for example, item responses on standardized multiple-choice tests are often simply recoded as wrong/right).

Item response modeling and analysis is greatly facilitated by the assumption of *unidimensionality*, i.e. that the latent variable "driving" the item responses is a one-dimensional, typically real-valued, random variable. In this paper we are concerned with two recent approaches to characterizing (the distributions of) binary item response data for which traditional unidimensional IRT representations exist. We consider a vector of J response variables,

$$\boldsymbol{X}_J = (X_1, X_2, \ldots, X_J),$$

representing the responses (1 = positive response, 0 = negative response) of a randomly chosen subject to the J test items or other stimuli.

Let $\boldsymbol{x}_J = (x_1, x_2, \ldots, x_J)$ represent an arbitrary fixed outcome of $\boldsymbol{X}_J$; an IRT representation makes assumptions on the conditional distribution $P[\boldsymbol{X}_J = \boldsymbol{x}_J | \boldsymbol{\Theta} = \boldsymbol{\theta}]$ which impose restrictions on the marginal distribution $P[\boldsymbol{X}_J = \boldsymbol{x}_J]$ through the integral

$$P[\boldsymbol{X}_J = \boldsymbol{x}_J] = \int P[\boldsymbol{X}_J = \boldsymbol{x}_J | \boldsymbol{\Theta} = \boldsymbol{\theta}]\, dF(\boldsymbol{\theta}). \tag{1}$$

Here $F(\boldsymbol{\theta})$ is the sampling distribution of the latent variable or vector $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_d)$ in the examinee population under discussion. (A brief consideration of estimation in (1) follows Proposition 2.1 below.) Obviously, the representation (1) does not itself restrict the distribution of response variables $P[\boldsymbol{X}_J = \boldsymbol{x}_J]$ in any way. Standard practice involves the imposition of additional conditions that make (1) a restrictive, and hence meaningful, representation.

The traditional IRT assumptions are that *local independence* holds,

$$P[\boldsymbol{X}_J = \boldsymbol{x}_J | \boldsymbol{\Theta} = \boldsymbol{\theta}] = \prod_{j=1}^{J} P[X_j = x_j | \boldsymbol{\Theta} = \boldsymbol{\theta}] \tag{LI}$$

(this was called “latent conditional independence” by Holland and Rosenbaum, 1986), and that *monotonicity* holds for the response functions:

$$P_j(\boldsymbol{\theta}) \equiv P[X_j = 1 | \boldsymbol{\Theta} = \boldsymbol{\theta}] \ \textit{coordinatewise nondecreasing in } \boldsymbol{\theta},\ \forall\, j \tag{M}$$

in the sense that if $\theta_k^{(1)} \leq \theta_k^{(2)}$ for all $k = 1, 2, \ldots, d$ then $P_j(\boldsymbol{\theta}^{(1)}) \leq P_j(\boldsymbol{\theta}^{(2)})$. In view of the binary nature of the data, the LI condition may be rewritten as

$$P[\boldsymbol{X}_J = \boldsymbol{x}_J | \boldsymbol{\Theta} = \boldsymbol{\theta}] = \prod_{j=1}^{J} P_j(\boldsymbol{\theta})^{x_j} (1 - P_j(\boldsymbol{\theta}))^{1 - x_j}.$$

One additional assumption is needed to make (1) restrictive, namely that the *dimensionality* d of $\boldsymbol{\Theta}$ is much smaller than the test length J (see for example Holland and Rosenbaum, 1986), that is:

$$d \ll J. \tag{D}$$

(In the development that follows, this is formalized by requiring that d remain fixed as J grows.) The three assumptions, LI, M, and D form the foundation of item/test modeling in traditional IRT. Indeed, if any of these three assumptions is completely relaxed the resulting "model" will fit any distribution of binary data: Suppes and Zanotti (1981) show that M cannot be completely dropped; Holland and Rosenbaum (1986) show that D cannot be completely dropped; and it is easy to see, using $\Theta = \prod_1^J X_j$, that LI cannot be completely dropped. Holland (1990) provides a valuable account of the rationales for traditional IRT representations, of the features of various IRT-style models, and of fundamental estimation and inference issues.

The least d for which the representation (1) holds and satisfies LI and M (and smoothness of the response functions) we will denote d_L. We will refer to the case in which $d_L = 1$ as the *strictly unidimensional* case.

Throughout this paper we embed the observed response variables $\boldsymbol{X}_J$ in an infinite sequence of similar response variables

$$\boldsymbol{X} = (X_1, X_2, \ldots).$$

LI and other properties of $\boldsymbol{X}_J$ extend in a natural way to the infinite sequence $\boldsymbol{X}$ by requiring that they hold, in a fashion satisfying Kolmogorov's consistency conditions, in every finite set of responses $\boldsymbol{X}_J$ taken from $\boldsymbol{X}$.

Various unidimensionality assumptions have been investigated to see what properties they imply for the manifest distribution $P[\boldsymbol{X}_J = \boldsymbol{x}_J]$ through (1). Holland and Rosenbaum (Holland, 1981; Rosenbaum, 1984; Holland and Rosenbaum, 1986) have shown that when strict unidimensionality holds, the item responses $\boldsymbol{X}_J$ must be *conditionally associated* (CA); this shows that strict unidimensionality is a restrictive and hence meaningful set of conditions. Cressie and Holland (1983) (see also Tjur, 1982), have characterized the Rasch model in terms of a suitably restricted log-linear model for $P[\boldsymbol{X}_J = \boldsymbol{x}_J]$. And de Finetti's Theorem may be used to characterize an infinite sequence of binary response variables with identical response functions by the property that the response variables must be exchangeable. Except for the special cases of the Rasch model and de Finetti's theorem, no other characterizations of strictly unidimensional structure in terms of features of $P[\boldsymbol{X}_J = \boldsymbol{x}_J]$ seem to be known.

Stout (1987; 1990) capitalizes on the good θ-estimation properties of the proportion of positive responses $\overline{X}_J = \frac{1}{J}\sum_1^J X_j$, when J is large and unidimensionality holds, to produce a statistical test of latent variable unidimensionality. Stout's statistical test is tailored to his *essential unidimensionality* condition ($d_E = 1$) which, in contrast to strict unidimensionality, allows there to be some minor violations of the LI and M conditions.

In this paper we examine the intersection of the Holland/Rosenbaum and Stout approaches to unidimensionality. In Section 2 we introduce the notion of *useful* latent variable representations for binary response data. In Section 3 we review the conditional association and essential unidimensionality conditions, and relate them to the existence of a useful strictly unidimensional representation for $\boldsymbol{X}$. Section 4 introduces two new conditions, negative association and monotonicity of the empirical response curves $P[X_i = 1|\overline{X}_J - X_i/J]$. Section 5 gives a partial characterization of useful $d_L = 1$ representations in terms of CA, $d_E = 1$, and the new conditions. A consequence of our work here is a better understanding of how "far" each of the conditional association and essential unidimensionality approaches is from the strict unidimensionality assumptions.

2. Useful unidimensional representations. In seeking an understanding of the traditional $d_L = 1$ model, we will be considering latent variable representations that are somewhat more general. Hence it is worthwhile to ask what constitutes a "useful" unidimensional latent variable representation. We avoid specific parametric assumptions about $P[\boldsymbol{X}_J = \boldsymbol{x}_J|\Theta = \theta]$ and $F(\theta)$ in (1), and instead require that a representation (1) for $(\Theta, X_1, X_2, \ldots)$ satisfy the following somewhat informal definition. Note that what we mean by "useful" here relates primarily to connecting observations of $\boldsymbol{X}_J$ with inferences about Θ.

Definition 2.1 *An IRT representation, in which LI may or may not hold, will be called* useful *if and only if the following three principles are satisfied:*

U1 Θ can be estimated from the observed values of $X_1, X_2, \ldots, X_J$. At minimum we require that there are functions $t_J(x_1, \ldots, x_J)$ that consistently estimate Θ in the sense that

$$t_J(X_1, \ldots, X_J) - \Theta \to 0,$$

in probability, as the test length J grows. Moreover, consistent estimation should still be possible even though any fixed group of response variables $(Y_1, \ldots, Y_{J_0})$ in $\boldsymbol{X}$ is absent.

U2 Subjects with higher Θ values tend to produce more positive responses. In particular we require that the average response curve be increasing in θ:

$$\overline{P}_J(\theta) = E[\overline{X}_J | \Theta = \theta] \textit{ is increasing in } \theta.$$

U3 Θ is useful for categorizing subjects. In particular Θ should be able to take on at the very least two distinct values, each with positive probability.

These principles are implicit in traditional item response models, and are easily justified on practical grounds:

First, Θ has little objective value as a index of ability, achievement, aptitude, or other trait, if it cannot be estimated; hence U1. There is no hope that Θ can be estimated with high precision unless $J \to \infty$ (e.g. the survey by Fienberg, 1986; see also Levine, 1992, for a related discussion), so U1 represents, in some sense, a minimal estimation condition. The further requirement that estimation of Θ should not depend strongly on the presence of particular response variables is central to what is meant by "latent trait."

Principle U2 reflects the interpretation of Θ as a quantity of the latent trait, and of $\boldsymbol{X}_J$ as an instrument for measuring that quantity. U2 also has the effect of bounding the Fisher information for estimating θ away from zero (Junker, 1991), and in general U2 makes it easier for the representation to satisfy U1. Finally, we will only require U2 to hold for J larger than some $J_0 > 0$; see Definition 3.2 below. In this sense U2 does not imply that the individual response curves $P_j(\theta)$ are monotone, for any j.

Principle U3 simply reflects the practical desire to use the $\boldsymbol{X}_J$ to diagnose, assess or otherwise categorize subjects and populations. If there is no variation in Θ then there is no sensible way to use $\boldsymbol{X}_J$ in this way. Indeed, if U3 fails in a $d_L = 1$ representation, the components of $\boldsymbol{X}_J$ become independent, nonidentical Bernoulli's under $P[\boldsymbol{X}_J = \boldsymbol{x}_J]$ itself. In terms of (1), U3 asserts that the (prior) Θ distribution does not concentrate at a single θ value.

We may illustrate Definition 2.1 with the following proposition.

PROPOSITION 2.1 *Suppose that there is a $d_L = 1$ representation for $\boldsymbol{X}$ and Θ, for which principle* U2 *holds, in the sense of Definition 3.2 below. Let $\mathcal{T}_{\boldsymbol{X}}$ be the tail σ-field of $\boldsymbol{X}$, and $\sigma(\Theta)$ the sigma field generated by Θ. Then*

(a) $\mathcal{T}_{\boldsymbol{X}} = \sigma(\Theta)$, almost surely; and

(b) If Θ has a nontrivial distribution, then the representation is useful *in the sense of Definition 2.1.*

PROOF. It can be deduced from Theorem 3.2 below that U1 holds, and this is enough to obtain part (b). For part (a), observe that U1 implies that $\sigma(\Theta) \subset \mathcal{T}_{\boldsymbol{X}}$; conversely, the 0–1 law for independent random variables shows that $\mathcal{T}_{\boldsymbol{X}} \subset \sigma(\Theta)$ almost surely. □

The distribution of $\boldsymbol{X}_J$ can be estimated to arbitrary accuracy by increasing number of cases of $\boldsymbol{X}_J$ observed, that is, by increasing the number of subjects and leaving J fixed. On the other hand, by Proposition 2.1, knowing the distribution of Θ is equivalent to knowing the tail behavior of the response variables. Hence, to estimate $P[\boldsymbol{X}_J = \boldsymbol{x}_J|\Theta = \theta]$ or $F(\theta)$ from data, it is necessary in general to increase both the number of cases of $\boldsymbol{X}_J$ observed and the length J of each response vector—unless specific parametric assumptions are made, e.g. the Rasch model (Cressie and Holland, 1983; Tjur, 1982)—even when strict unidimensionality holds. Following psychometric traditions, we will call the marginal distributions $P[\boldsymbol{X}_J = \boldsymbol{x}_J]$ the *manifest structure*, and the marginal distribution $F(\theta)$ and conditional distributions $P[\boldsymbol{X}_J = \boldsymbol{x}_J|\Theta = \theta]$ the *latent structure*, of the sequence $\boldsymbol{X}$ of response variables.

3. Conditional association and essential independence.

3.1. Conditional association. Holland and Rosenbaum have sought covariance conditions, or equivalently probability inequalities, in the distribution of $\boldsymbol{X}_J$ that must be satisfied if any $d_L = 1$ model applies. The starting place for their investigations may be taken to be coordinatewise nondecreasing functions $f(\boldsymbol{y})$ of finite response vectors $\boldsymbol{Y} = (Y_1, \ldots, Y_{J_0})$ taken from $\boldsymbol{X}$: if $y_j^{(1)} \leq y_j^{(2)}$ $\forall j$, then $f(\boldsymbol{y}^{(1)}) \leq f(\boldsymbol{y}^{(2)})$. These are exactly the functions that assign higher summary scores to comparable response sequences with more positive responses. Under LI, the response variables must be *associated* (cf. Esary, Proschan and Walkup, 1967), conditional on the latent variable Θ:

$$\mathrm{Cov}\,(f(\boldsymbol{Y}), g(\boldsymbol{Y})|\Theta = \theta) \geq 0, \tag{2}$$

for each possible θ and each pair of coordinatewise nondecreasing functions f and g. Although LI is not by itself a restrictive condition, if M is also assumed this "local association" condition can be converted from a condition on the latent structure into to a condition on the manifest structure.

THEOREM 3.1 *(Rosenbaum, 1984; Holland and Rosenbaum, 1986). If* $\boldsymbol{X}$ *satisfies* $d_L = 1$, *then* $\boldsymbol{X}$ *is* conditionally associated (CA): *For every pair of disjoint, finite response vectors* $\boldsymbol{Y}$ *and* $\boldsymbol{Z}$ *in* $\boldsymbol{X}$, *every pair of coordinatewise nondecreasing functions* $f(\boldsymbol{Y})$ *and* $g(\boldsymbol{Y})$, *and every function* $h(\boldsymbol{Z})$,

$$\mathrm{Cov}\left(f(\boldsymbol{Y}), g(\boldsymbol{Y}) | h(\boldsymbol{Z}) = c\right) \geq 0 \ \forall\ c \in \mathrm{range}(h). \tag{CA}$$

Consider a test with item responses $\boldsymbol{X}_J$, which are rearranged and partitioned into the "subtests" $\boldsymbol{Y}$ and $\boldsymbol{Z}$. Intuitively, if $\boldsymbol{X}_J$ satisfies $d_L = 1$, the test possesses so much internal coherence (the item responses are driven monotonically by the single latent variable Θ) that all "reasonable" subtest scores on $\boldsymbol{Y}$ must be correlated, in any subpopulation of examinees selected by *any* criterion $h(\boldsymbol{Z})$ relating to another part of the test.

Our statement of Theorem 3.1 is an easy extension of Holland and Rosenbaum's result for finite length response vectors to infinite sequences. Note also that for finitely many discrete random variables $Z_1, Z_2, \ldots, Z_M$, conditioning on a scalar-valued function $h(\boldsymbol{Z})$ is equivalent to Holland and Rosenbaum's practice of conditioning on vector-valued $\boldsymbol{h}(\boldsymbol{Z})$. Seminal special cases of (2) and CA were developed by Holland (1981). Applications to studying the internal coherence of a set of standardized test items may be found in Rosenbaum (1984) or Holland and Rosenbaum (1986). Related work appears in Rosenbaum (1985; 1987; 1988). An application of CA to assessing the dimensionality of standardized tests for the National Assessment of Educational Progress is described by Zwick (1987).

3.2. Essential independence. A successful approach to identifying unidimensional latent structure outside the strict $d_L = 1$ framework has been pursued by Stout (1987, 1990), and extended by Junker (1991). The main idea, which borrows from both the large sample tradition in mathematical statistics and the factor analysis tradition in psychometrics, is that of *essential independence*. Actually, we use Stout's *strong essential independence*, with some changes in terminology to match Junker (1991). For any (infinite) sequence of binary response variables $\boldsymbol{X} = (X_1, X_2, X_3, \ldots)$, consider uniformly bounded item scores $A_j(X_j)$, such that for some $M < \infty$, $|A_j(\cdot)| \leq M$ for all j, and denote $\overline{A}_J = \frac{1}{J}\sum_{j=1}^{J} A_j(X_j)$.

DEFINITION 3.1 $\boldsymbol{X}$ *is* essentially independent *(EI)* with respect to Θ *if*

$$\lim_{J\to\infty} \mathrm{Var}\,(\overline{A}_J|\Theta = \theta) = 0 \tag{EI}$$

for every set of uniformly bounded item scores $\{A_j(\cdot) : j = 1, 2, \ldots\}$.
In particular, for a sequence of binary response variables $\boldsymbol{X}$, EI implies that the proportion of positive responses $\overline{X}_J$ consistently estimates values of the average response curve $\overline{P}_J(\theta)$, as $J \to \infty$.

Let us call the uniformly bounded item scores $\{A_j(\cdot)\}$ *ordered* if $A_j(0) \leq A_j(1)$; and call the ordered item scores *asymptotically discriminating* if $\frac{1}{J}\sum_{j=1}^{J}\{A_j(1) - A_j(0)\}$ is positive and bounded away from 0 as $J \to \infty$. Also, denote $\overline{A}_J(\theta) = E[\overline{A}_J|\theta]$.

DEFINITION 3.2 $\boldsymbol{X}$ *is* locally asymptotically discriminating *(LAD), if for every set of asymptotically discriminating item scores, to every* θ *there corresponds an interval* N_θ *containing* θ *and an* $\epsilon_\theta > 0$ *such that*

$$\liminf_{J\to\infty} \frac{\overline{A}_J(t) - \overline{A}_J(\theta)}{t - \theta} \geq \epsilon_\theta, \;\; \forall\, t \in N_\theta, \;\; t \neq \theta. \tag{LAD}$$

DEFINITION 3.3 $\boldsymbol{X}$ *is* essentially unidimensional, *if there exists* Θ *such that* $\boldsymbol{X}$ *is EI and LAD with respect to* Θ.

If $\boldsymbol{X}$ is essentially unidimensional, we will write $d_E = 1$. If no such unidimensional Θ exists, we write $d_E > 1$. A statistical procedure for testing the hypothesis that $\boldsymbol{X}_J$ comes from a $d_E = 1$ sequence has been developed by Stout (1987). When $d_E = 1$, $\overline{A}_J(\theta)$ may be inverted to produce estimates of θ directly.

THEOREM 3.2 *(Stout, 1990). If the sequence $\boldsymbol{X}$ satisfies $d_E = 1$ with respect to Θ then for any set of asymptotically discriminating item scores,*

$$\forall\, \epsilon > 0, \lim_{J\to\infty} P\left[|\overline{A}_J^{-1}(\overline{A}_J) - \theta| \leq \epsilon \middle| \Theta = \theta\right] = 1, \tag{3}$$

where $\overline{A}_J^{-1}(u)$ is the inverse function for $\overline{A}_J(\theta)$.

Indeed, under the conditions of Theorem 3.2 and some mild smoothness conditions, the maximum likelihood estimate of θ calculated as though LI were true is also consistent for θ (Junker, 1991). It is valuable to think of EI as the greatest possible weakening of LI under which LI-based trait estimation/prediction schemes might be expected to work. In this sense, the study of EI is the study of robustness of latent trait estimators to variations from an LI latent structure. Clarke and Junker (1991) pursue this matter in a more general setting.

Moreover the latent variable Θ with respect to which $d_E = 1$ holds is unique, up to a monotone transformation. The following theorem can be deduced from Theorem 3.3 of Stout (1990).

THEOREM 3.3*(Stout, 1990). Suppose $\boldsymbol{X}$ is essentially unidimensional with respect to both Θ and τ. Then there exists an invertible function $h(\cdot)$ such that $\Theta = h(\tau)$, almost surely.*

Because LAD formalizes principle U2 and Theorem 3.2 implies principle U1, any $d_E = 1$ model in which Θ has a nontrivial distribution is useful in the sense of Definition 2.1. However, the existence of a useful $d_E = 1$ representation does not imply the existence of a useful $d_L = 1$ model. Stout (1990, Example 2.3) gives a model for a sequence of "paragraph comprehension" questions that is a useful $d_E = 1$ model: Suppose that $P[X_j = 1|\theta] \equiv \theta$; and the items are arranged in successive groups of g_o items as

$$\begin{aligned}&X_1, X_2, \ldots, X_{g_o};\\&X_{g_o+1}, X_{g_o+2}, \ldots, X_{2g_o};\\&etc.,\end{aligned}$$

such that different groups of g_o items are independent of one another, given θ, and items within a single group are positively correlated, given θ, and with

$$\mathrm{Corr}\,(X_i, X_j|\theta) = \begin{cases} c \text{ if } X_i \text{ and } X_j \text{ are in the same group,} \\ 0 \text{ if not,} \end{cases}$$

for some fixed $c \in (0, 1]$.

No useful $d_L = 1$ model can be formulated for $\boldsymbol{X}$: For suppose $kg_o < i < j \leq (k+1)g_o$, so that $\mathrm{Cov}\,(X_i, X_j|\Theta = \theta_o) > 0$. If there were a unidimensional latent variable τ with respect to which LI and LAD held, we could use $h(\cdot)$ from Theorem 3.3 to obtain

$$\begin{aligned} 0 &= \mathrm{Cov}\,(X_i, X_j|\tau = t_o) \\ &= \mathrm{Cov}\,(X_i, X_j|h(\Theta) = h(\theta_o)) \\ &= \mathrm{Cov}\,(X_i, X_j|\Theta = \theta_o) \\ &\neq 0. \end{aligned}$$

This contradiction shows that no such τ can exist, i.e. no useful $d_L = 1$ model exists for the sequence of paragraph comprehension items.

3.3. Combining CA and $d_E = 1$. The following lemma tells us that under $d_E = 1$, for any finite response vector $\boldsymbol{Y}$ taken from $\boldsymbol{X}_J$, we may approximate expected values of the form $E[f(\boldsymbol{Y})|\Theta]$ with expected values of the form $E[f(\boldsymbol{Y})|\alpha_J \leq \overline{X}_J \leq \beta_J]$, as $J \to \infty$. We shall assume that

$$E[f(\boldsymbol{Y})|\Theta = \theta] \textit{ is continuous in } \theta \tag{4}$$

for any function $f(\boldsymbol{Y})$ of finitely many response variables.

LEMMA 3.1 *Suppose* $\boldsymbol{X}$ *satisfies* EI *and* LAD *with respect to some unidimensional* Θ*, and assume* (4). *If* $f(\boldsymbol{Y})$ *is a function that depends on only finitely many response variables* $\boldsymbol{Y} = (Y_1, \ldots, Y_{J_0})$ *(J_0 fixed) from* $\boldsymbol{X}$*, then for every set of asymptotically discriminating item scores* $\{A_j(\cdot) : j = 1, 2, \ldots\}$ *and for each* θ *there exist* $\epsilon_J \to 0$ *for which*

$$\lim_{J\to\infty} E\left[f(\boldsymbol{Y}) \middle| |\overline{A}_J^{-1}(\overline{A}_J) - \theta| < \epsilon_J\right] = E[f(\boldsymbol{Y})\,|\Theta = \theta].$$

REMARKS. Using Lemma 3.1 we could approximate a uniformly bounded function $f(\boldsymbol{Y})$ defined on all of $\boldsymbol{Y}$ in a similar manner. For example $f_k(\boldsymbol{Y}) = E[f(\boldsymbol{Y})|Y_1, \ldots, Y_k]$ converges to $f(\boldsymbol{Y})$ in L^1 and *a.s.* as k tends to infinity, by a standard martingale limit theorem; so that for large J and k, $E\left[f_k(\boldsymbol{Y}) \middle| |\overline{A}_J^{-1}(\overline{A}_J) - \theta| < \epsilon_J\right] \approx E[f(\boldsymbol{Y})\,|\Theta = \theta]$. However we will not pursue this extension, since it will not be needed here.

PROOF. For any event C, let 1_C take the value 1 if C is true and 0 if C is false, and let $E[f(\boldsymbol{Y}); C] = E[f(\boldsymbol{Y})1_C]$. We may decompose the expectation on the left above as

$$\begin{aligned} &E\left[f(\boldsymbol{Y}) \middle| |\overline{A}_J^{-1}(\overline{A}_J) - \theta| < \epsilon\right] \\ &= \frac{E[f(\boldsymbol{Y}); |\Theta - \theta| < \epsilon]}{P[|\Theta - \theta| < \epsilon]} \cdot \frac{E[f(\boldsymbol{Y}); |\overline{A}_J^{-1}(\overline{A}_J) - \theta| < \epsilon]}{E[f(\boldsymbol{Y}); |\Theta - \theta| < \epsilon]} \cdot \frac{P[|\Theta - \theta| < \epsilon]}{P[|\overline{A}_J^{-1}(\overline{A}_J) - \theta| < \epsilon]} \\ &= \mathrm{I}(\epsilon) \;\cdot\; \mathrm{II}(\epsilon) \;\cdot\; \mathrm{III}(\epsilon). \end{aligned}$$

Note that for any rate $\epsilon = \epsilon_J \to 0$, $\mathrm{I}(\epsilon_J) \to E[f(\boldsymbol{Y})|\Theta = \theta]$, as $J \to \infty$, using the continuity condition (4) and the integral mean value theorem. The idea now is to choose $\epsilon = \epsilon_J \to 0$ so that $\mathrm{II} \to 1$ and $\mathrm{III} \to 1$ as $J \to \infty$. We will look at II explicitly; note that III is a special case of II. We have

$$\mathrm{II}(\epsilon) = 1 - \frac{E\{f(\boldsymbol{Y})[1_{\{|\overline{A}_J^{-1}(\overline{A}_J) - \theta| < \epsilon\}} - 1_{\{|\Theta - \theta| < \epsilon\}}]\}}{E\{f(\boldsymbol{Y})1_{\{|\Theta - \theta| < \epsilon\}}\}};$$

one can apply Theorem 3.2 to show that the numerator on the right tends to zero for each fixed $\epsilon > 0$ as $J \to \infty$; a simple diagonalization argument now yields a rate $\epsilon_J \to 0$ for which II $\to 1$. A similar argument works for III, and a further diagonalization completes the proof. $\square$

We can use Lemma 3.1 to gain information about the latent structure of $\boldsymbol{X}$ from the manifest condition CA. Proposition 3.1 shows that CA and $d_E = 1$ together give the same local association condition (2) as $d_L = 1$ alone.

PROPOSITION 3.1 *Suppose $\boldsymbol{X}$ satisfies* CA *and $d_E = 1$, and suppose that* (4) *holds. Then* (2) *holds.*

REMARKS. By modifying the proof of Lemma 3.1 we could also conclude conditional association given $\Theta = \theta$, i.e., if $\boldsymbol{Z}$ were another finite response vector from $\boldsymbol{X}$ disjoint from $\boldsymbol{Y}$ and $h(\boldsymbol{Z})$ were any function, then

$$\mathrm{Cov}\,(f(\boldsymbol{Y}), g(\boldsymbol{Y})|h(\boldsymbol{Z}), \Theta = \theta) \geq 0.$$

PROOF. Let $\boldsymbol{Y}$ be a response vector from $\boldsymbol{X}_{J_0}$, for fixed J_0, let $\boldsymbol{W} = (X_{J_0+1}, X_{J_0+2}, \ldots)$ and let $\overline{A}_J(\theta) = E[\frac{1}{J}\sum_1^J W_j|\theta]$. Using CA and a sequence ϵ_J obtained with Lemma 3.1,

$$\begin{aligned} 0 &\leq \mathrm{Cov}\left[f(\boldsymbol{Y}), g(\boldsymbol{Y})\,\middle|\,|\overline{A}_J^{-1}(\overline{W}_J) - \theta| < \epsilon_J\right] \\ &\to \mathrm{Cov}\,(f(\boldsymbol{Y}), g(\boldsymbol{Y})|\Theta = \theta) \end{aligned}$$

as $J \to \infty$. $\square$

Let us digress briefly to indicate another way in which CA and $d_E = 1$ interact well. Two alternative definitions of EI have been proposed by Stout (1990), one involving the full sequence $\boldsymbol{X}$ but taking absolute values of covariances,

$$\lim_{J\to\infty} \binom{J}{2}^{-1} \sum\sum_{1\leq i<j\leq J} |\mathrm{Cov}\,(X_i, X_j|\theta)| = 0, \tag{5}$$

and another involving "nonsparse" subsequences of responses which, in the present context, is equivalent to considering only those asymptotically discriminating item scores for which $A_j(0) \equiv 0$ and $A_j(1) \in \{0,1\}$, and requiring

$$\lim_{J\to\infty} \binom{J}{2}^{-1} \sum\sum_{1\le i<j\le J} \mathrm{Cov}\,(A_i(X_i), A_j(X_j)|\theta) = 0. \tag{6}$$

It is not known in general whether these three definitions are equivalent. However, under CA and LAD they are:

COROLLARY 3.1 *If* CA *and* LAD *hold for* $\boldsymbol{X}$, *then all three definitions of* EI *are equivalent.*

PROOF. Condition (5) implies EI as defined in Definition 3.1, which in turn implies (6). For the converse directions, observe that if LAD holds (for the restricted case of $A_j(X_j) \in \{0,1\}\ \forall j$), then by Proposition 3.1, $\mathrm{Cov}\,(X_i, X_j|\theta) \ge 0$, $\forall i,j$. In this case, (5) follows from (6). □

Returning to our main development, the next result complements Proposition 3.1 by characterizing LI in terms of quantities that can be approximated by manifest quantities $E[f(\boldsymbol{Y})|\alpha_J \le \overline{X}_J \le \beta_J]$ as in Lemma 3.1 under EI and LAD.

PROPOSITION 3.2 $\boldsymbol{X}$ *satisfies* LI *with respect to* Θ *if and only if the following two conditions hold: For all* θ, *all coordinatewise nondecreasing* f *and* g, *and all finite response vectors* $\boldsymbol{Y}$ *taken from* $\boldsymbol{X}$,

$$\mathrm{Cov}\,(f(\boldsymbol{Y}), g(\boldsymbol{Y})|\Theta = \theta) \ge 0; \tag{7}$$

and for all θ, i, *and* j,

$$\mathrm{Cov}\,(X_i, X_j|\Theta = \theta) \le 0. \tag{8}$$

REMARKS. Note that (7) is the same local association condition as introduced in (2).

PROOF. That LI implies (7) follows from Esary, Proschan and Walkup (1967); (8) is trivially satisfied under LI with $\mathrm{Cov}\,(X_i, X_j|\theta) \equiv 0$. For a proof of the converse, in unconditional form, see Newman and Wright (1981) or Joag-Dev (1983). □

Despite the strength of the CA and $d_E = 1$ conditions, it seems unlikely that they together guarantee that a useful $d_L = 1$ model exists. To see why, suppose $\boldsymbol{X}$ satisfies $d_E = 1$ with respect to some unidimensional Θ, and suppose that the Θ in this $d_E = 1$ representation is the first coordinate of a latent vector $\boldsymbol{\Theta} = (\Theta, \Theta_2, \Theta_3, \ldots, \Theta_d)$ needed for a $d_L = d$ representation:

$$P[\boldsymbol{X}_J = \boldsymbol{x}_J | \Theta = \theta] = \int \prod_{j=1}^{J} P_j(\boldsymbol{\theta})^{x_j}(1 - P_j(\boldsymbol{\theta}))^{1-x_j} dF(\boldsymbol{\theta}|\Theta = \theta)$$

for each J. In general, although $d_E = 1$ implies that

$$\lim_{J\to\infty} \left(\begin{array}{c} J \\ 2 \end{array} \right)^{-1} \sum\sum_{1 \le i < j \le J} \mathrm{Cov}\,(X_i, X_j | \Theta = \theta) = 0,$$

the individual covariances

$$\mathrm{Cov}\,(X_i, X_j | \Theta = \theta) = \mathrm{Cov}\,(P_i(\boldsymbol{\Theta}), P_j(\boldsymbol{\Theta}) | \Theta = \theta)$$

may be positive or negative, depending on the conditional distribution of $(\Theta_2, \ldots, \Theta_d)$ given Θ.

Now suppose $\boldsymbol{X}$ satisfies CA also. Then by Proposition 3.1,

$$\mathrm{Cov}\,(X_i, X_j | \Theta = \theta) \ge 0, \forall\ i \ne j; \tag{9}$$

in fact the stronger condition (2) holds. Thus, not only does Θ represent the dominant trait for $\boldsymbol{X}$ (in the sense that EI holds), but the "minor traits" needed for LI to hold are concordant with Θ, in the sense that they interact with Θ so as to keep the local inter-item covariances nonnegative.

Under CA and $d_E = 1$, therefore, there is enough "coherence" that covariances between responses, given $\Theta = \theta$, are nonnegative. Indeed it is quite plausible that under these conditions, for some sequence of response variables $\boldsymbol{X}$, some of the inequalities in (9) will be strict (despite its plausibility we have not been able to construct an example in which this may rigorously be shown). But if any of the inequalities (9) are strict for a latent variable Θ with respect to which $d_E = 1$, then there cannot exist another latent variable τ with respect to which a useful $d_L = 1$ model exists; this follows by appealing to Theorem 3.3 again. Thus some condition in addition to CA and $d_E = 1$ seems to be needed to get a useful $d_L = 1$ representation.

4. Two helpful conditions. In this section we introduce general conditions on the manifest distribution $P[\boldsymbol{X}_J = \boldsymbol{x}_J]$ which allow us to promote the "local association" conditions (2) and (7) to LI, and to ensure that individual response functions $P_j(\theta)$ are monotone.

4.1. Negative association. As indicated at the end of Section 3, there may be some situations in which EI and CA hold but the implied coherence among items is so tight that LI cannot also hold. The following theorem provides an additional *negative association* condition enjoyed by many sequences of independent random variables.

THEOREM 4.1 *(Joag-Dev and Proschan, 1982). Suppose the random variables $\boldsymbol{X}_J = (X_1, X_2, \ldots, X_J)$ are independent with (possibly nonidentical) log-concave densities. Then for any partition $(\boldsymbol{Y}, \boldsymbol{Z})$ of $\boldsymbol{X}_J$ and any coordinatewise nondecreasing functions $f(\boldsymbol{y})$ and $g(\boldsymbol{z})$,*

$$\mathrm{Cov}\,(f(\boldsymbol{Y}), g(\boldsymbol{Z})|\overline{X}_J) \leq 0.$$

Taking f and g to be functions that select single items from $\boldsymbol{X}_J$, and noting that the Bernoulli density is log-concave, we see that when LI with respect to Θ holds, then

$$\mathrm{Cov}\,(X_i, X_j|\overline{X}_J, \Theta) \leq 0 \qquad \text{(LCSN)}$$

for all $i < j \leq J$. This says that, under LI, the item responses are "not too tightly bound together" even though (2) holds: each X_i and X_j are sufficiently free of one another among examinees at the same latent trait level that when X_i increases from one examinee to the next, X_j is free to decrease so that the summary score $\overline{X}_J$ may be kept constant. The abbreviation LCSN stands for *locally, covariances given summary score are negative.*

However, LCSN is a condition on the latent, not the manifest, structure. To obtain a natural manifest structure analogue to LCSN it is useful to consider the special case of the locally independent Rasch model, for which by definition logit $P[X_j = 1|\theta] = \theta - b_j$. Here $\overline{X}_J$ is sufficient for Θ: $(X_1, X_2, \ldots, X_J)$ are independent of Θ given $\overline{X}_J$. Consequently,

$$\mathrm{Cov}\,(X_i, X_j|\overline{X}_J) \leq 0 \qquad \text{(CSN)}$$

for all $i < j \leq J$. CSN should be read as *covariances given summary score are negative.*

Because $\overline{X}_J$ is not sufficient for Θ outside the Rasch model, CSN is an imperfect substitute for LCSN. However even outside the Rasch model, LCSN and CSN are closely related. For consider, under $d_L = 1$, the decomposition

$$\begin{aligned} \mathrm{Cov}\,(X_i, X_j|\overline{X}_J) &= E\left[\mathrm{Cov}\,(X_i, X_j|\overline{X}_J, \Theta)\Big|\overline{X}_J\right] \\ &\quad + \mathrm{Cov}\,\left[E[X_i|\overline{X}_J, \Theta], E[X_j|\overline{X}_J, \Theta]\Big|\overline{X}_J\right]. \end{aligned}$$

The first term on the right is nonpositive under LI, by LCSN. The second term may be negative or positive, but should be small since under $d_L = 1$ the (posterior) distribution of Θ given $\overline{X}_J$ should have very low variance, as J grows. A small simulation study (Junker, 1990) suggests that this holds in more realistic logistic response models for even moderate values of J ($J \geq 40$), and Theorems 2.1 and 3.2 of Clarke and Ghosh (1991) indicate that this holds more generally as $J \to \infty$. In particular when $\mathrm{Cov}\,(X_i, X_j|\overline{X}_J)$ fails to be negative for a LI model, we at least expect it to be near zero.

4.2. Manifest monotonicity. Let $\overline{X}_{iJ} = \overline{X}_J - X_i/J$; we will say *manifest monotonicity*, MM, holds if

$$E[X_i|\overline{X}_{iJ}] \text{ is nondecreasing in } \overline{X}_{iJ} \tag{MM}$$

for all $i \leq J$ (and all J). MM is intimately related to $d_L = 1$ latent structure, as Proposition 4.1 and Corollary 4.1 below show.

LEMMA 4.1 *If* LI *and* M *hold for* $\boldsymbol{X}$ *with respect to* Θ, *then* Θ *is stochastically increasing in* $S = \overline{X}_{iJ}$*:*

$$\forall\ a \leq b\ \forall\ c : P[\theta > c|S = a] \leq P[\theta > c|S = b] \tag{10}$$

whenever the conditional probabilities are defined.

PROOF. We may apply Theorem 2 of Grayson (1988) to the response vector (X_1, ..., X_{i-1}, X_{i+1}, ..., X_J) to see that the score $S = \overline{X}_{iJ}$ has the monotone likelihood ratio property

$$R_{ab}(\theta) \equiv \frac{P[S = b|\theta]}{P[S = a|\theta]} \text{ nondecreasing in } \theta,\ \forall\ a \leq b, \tag{11}$$

whenever the conditional probabilities are defined. To establish (10) (for any score S satisfying (11)), we may write its left hand side as

$$P[\theta > c|S = a] = \frac{\int_c^{+\infty} P[S = a|\theta]dF(\theta)}{\int_{-\infty}^{+\infty} P[S = a|\theta]dF(\theta)} = P[T > c]$$

where T is a random variable with density proportional to $P[S = a|t] \cdot dF(t)$. (I.e., $T = [\Theta|S = a]$.) On the other hand, the right hand side of (10) may be written as

$$P[\theta > c|S = b] = \frac{\int_c^{+\infty} P[S = a|\theta]R_{ab}(\theta)dF(\theta)}{\int_{-\infty}^{+\infty} P[S = a|\theta]R_{ab}(\theta)dF(\theta)} = \frac{E[R_{ab}(T)1_{\{T>c\}}]}{E[R_{ab}(T)]}$$

for the same random variable T. Hence (10) is equivalent to the assertion that

$$P[T > c] \cdot E[R_{ab}(T)] \le E[R_{ab}(T) 1_{\{T>c\}}]$$

which follows from property (P3) of Esary, Proschan and Walkup (1967), since $g(T) = 1_{\{T>c\}}$ and $h(T) = R_{ab}(T)$ are both nondecreasing functions of T. $\Box$

PROPOSITION 4.1

(a) LI, M $\Rightarrow$ MM

(b) EI, LAD, MM *and* (4) $\Rightarrow$ M

PROOF. (a) Note that

$$E[X_i|\overline{X}_{iJ}] = E\left[E[X_i|\overline{X}_{iJ}, \Theta]\Big|\overline{X}_{iJ}\right] = E[P_i(\Theta)|\overline{X}_{iJ}] \tag{12}$$

by LI. This last expectation is nondecreasing in $\overline{X}_{iJ}$ by M and Lemma 4.1, using a result of Lehmann (1955).

(b) Let $\theta^{(1)} < \theta^{(2)}$; then there exist sequences $\alpha_J^{(1)} \le \beta_J^{(1)}$ and $\alpha_J^{(2)} \le \beta_J^{(2)}$ with $\beta_J^{(1)} < \alpha_J^{(2)}$ for all large J, such that $\{\alpha_J^{(i)} \le \overline{X}_{iJ} \le \beta_J^{(i)}\} = \{|\overline{P}_J^{-1}(\overline{X}_{iJ}) - \theta^{(i)}| < \epsilon_J^{(i)}\}$, from Lemma 3.1. Then

$$\begin{aligned}
E[X_i|\Theta = \theta^{(1)}] &= \lim_{J\to\infty} E[X_i|\alpha_J^{(1)} \le \overline{X}_{iJ} \le \beta_J^{(1)}] \\
&= \lim_{J\to\infty} \frac{\sum_{\alpha_J^{(1)} \le c \le \beta_J^{(1)}} E[X_i|\overline{X}_{iJ} = c] P[\overline{X}_{iJ} = c]}{\sum_{\alpha_J^{(1)} \le c \le \beta_J^{(1)}} P[\overline{X}_{iJ} = c]} \\
&\le \lim_{J\to\infty} \frac{\sum_{\alpha_J^{(2)} \le c \le \beta_J^{(2)}} E[X_i|\overline{X}_{iJ} = c] P[\overline{X}_{iJ} = c]}{\sum_{\alpha_J^{(2)} \le c \le \beta_J^{(2)}} P[\overline{X}_{iJ} = c]} \\
&= \lim_{J\to\infty} E[X_i|\alpha_J^{(2)} \le \overline{X}_{iJ} \le \beta_J^{(2)}] \\
&= E[X_i|\Theta = \theta^{(2)}]
\end{aligned}$$

(under MM, the second ratio of sums above is a weighted average of larger conditional expectations than the first one). $\Box$

COROLLARY 4.1 *Under* LI, LAD *and* (4) *we have*

$$\text{MM} \Leftrightarrow \text{M}$$

Molenaar (personal communication, July 1990) has independently discovered Proposition 4.1(a). Example 4.1 below is related to an example of his, and shows that the method of proof of Lemma 4.1 cannot be extended to the polytomous case, even if the cumulative response curves are continuous and strictly increasing. Example 4.2 below, also reported by Molenaar, shows that in Proposition 4.1(a) we cannot replace the "deleted average" $\overline{X}_{iJ}$ with the more natural average over all items $\overline{X}_J$.

EXAMPLE 4.1 *The monotone likelihood ratio property (11) does not extend to polytomous item response variables.* Let $0 \leq \Theta \leq 1$ and consider a single graded-response variable X taking the three values 0, 1 or 2, with

$$P[X \geq 1|\theta] = \begin{cases} 3\theta, & 0 < \theta \leq 1/4 \\ \frac{2}{3} + \frac{1}{3}\theta, & 1/4 < \theta \leq 1 \end{cases}$$

$$P[X \geq 2|\theta] = \begin{cases} 2\theta, & 0 \leq \theta \leq 1/4 \\ \theta + \frac{1}{4}, & 1/4 < \theta \leq 1/2 \\ \frac{1}{2} + \frac{1}{2}\theta, & 1/2 < \theta \leq 1. \end{cases}$$

For $\theta_0 = 1/4$ and $\theta_1 = 1/2$, calculation shows that the likelihood ratio $P[X = x|\theta_0]/P[X = x|\theta_1]$ is not monotone in x. $\square$

EXAMPLE 4.2 *(T. Snijder). $P[X_j = 1|\overline{X}_J = s]$ need not increase with s, and hence we may not replace $\overline{X}_{iJ}$ with $\overline{X}_J$ in Proposition 4.1(a).* Consider three binary response variables, and a two-point distribution for Θ, $P(\Theta = \theta_0) = P(\Theta = \theta_1) = \frac{1}{2}$. Let

$$P_j(\theta_0) \equiv \epsilon, \; j = 1, 2, 3; \; P_1(\theta_1) = \frac{1}{2}; \text{ and } P_2(\theta_1) = P_3(\theta_1) = 1 - \epsilon.$$

It follows that, as $\epsilon \to 0$, $P[X_1 = 1|\overline{X}_J = 1/3] \to \frac{1}{4}$; and $P[X_1 = 1|\overline{X}_J = 2/3] \to 0$. $\square$

5. A partial characterization of $d_L = 1$. The major results of the previous two sections may be summarized in the following theorem.

THEOREM 5.1 *If $\boldsymbol{X}$ satisfies $d_L = 1$ and* LAD *with respect to a unidimensional Θ, then each of the conditions* CA, $d_E = 1$, LCSN *and* MM *hold.*

We show in this section that the converse is also true: the four conditions CA, $d_E = 1$, LCSN and MM guarantee a useful $d_L = 1$ representation. Moreover this converse implication is still true if LCSN is replaced with its manifest structure analogue CSN.

To obtain these two converses of Theorem 5.1 we must connect conditioning on $\overline{X}_J$ alone, as in CSN and LCSN, with conditioning on intervals $\alpha_J \leq \overline{X}_J \leq \beta_J$, as in Lemma 3.1. In the proof of the next lemma, we assume that for each J and $i \leq J$ there exist differentiable "interpolating functions" g_{iJ} such that

$$\left.\begin{array}{l} E[X_i|\overline{X}_J] \equiv g_{iJ}(\overline{X}_J) \\ \sup_{i,J,u} |g'_{iJ}(u)| \leq M < \infty \end{array}\right\} \tag{13}$$

and that for each $J, i \leq J$, and θ there exist differentiable "interpolating functions" $g_{iJ\theta}$ such that

$$\left.\begin{array}{l} E[X_i|\overline{X}_J, \Theta = \theta] \equiv g_{iJ\theta}(\overline{X}_J) \\ \sup_{i,J,u} |g'_{iJ\theta}(u)| \leq M_\theta < \infty \end{array}\right\} \tag{14}$$

The conditions (13) and (14) are maximum rate-of-change conditions on the regressions of X_i onto $\overline{X}_J$; most likely they would be acceptable in practice. In particular, (13) and (14) do not by themselves imply monotonicity of the response curves $P_j(\theta)$.

LEMMA 5.1

(a) Suppose CSN *holds, and suppose* (13) *also holds. Then for any constants $\alpha_J \leq \beta_J$ for which the covariances are defined, and for which $\beta_J - \alpha_J \to 0$,*

$$\limsup_{J\to\infty} \mathrm{Cov}\,(X_i, X_j|\alpha_J \leq \overline{X}_J \leq \beta_J) \leq 0. \tag{15}$$

(b) Suppose LCSN *holds, and suppose* (14) *also holds. Then for any constants* $\alpha_J \le \beta_J$ *for which the covariances are defined, and for which* $\beta_J - \alpha_J \to 0$,

$$\limsup_{J\to\infty} \mathrm{Cov}\,(X_i, X_j | \alpha_J \le \overline{X}_J \le \beta_J, \theta) \le 0. \tag{16}$$

PROOF. We will do part (a) only; part (b) is virtually identical. We have

$$\begin{aligned}\mathrm{Cov}\,(X_i, X_j | \alpha_J \le \overline{X}_J \le \beta_J) &= E\left[\mathrm{Cov}\,(X_i, X_j|\overline{X}_J)\Big|\, \alpha_J \le \overline{X}_J \le \beta_J\right] \\ &+ \mathrm{Cov}\,\left(E[X_i|\overline{X}_J], E[X_j|\overline{X}_J]\Big|\, \alpha_J \le \overline{X}_J \le \beta_J\right).\end{aligned} \tag{17}$$

The first term on the right is evidently nonpositive, by CSN. Dropping the conditioning on $\alpha_J \le \overline{X}_J \le \beta_J$ from the notation for brevity, the second term in (17) is

$$\begin{aligned}\mathrm{Cov}\,(E[X_i|\overline{X}_J], E[X_j|\overline{X}_J]) &= \mathrm{Cov}\,\left(g_{iJ}(\overline{X}_J), g_{iJ}(\overline{X}_J)\right) \\ &\le \{\mathrm{Var}\, g_{iJ}(\overline{X}_J) \cdot \mathrm{Var}\, g_{jJ}(\overline{X}_J)\}^{\frac{1}{2}},\end{aligned}$$

by the Cauchy-Schwarz inequality. Now applying Taylor's theorem,

$$\mathrm{Var}\, g_{iJ}(\overline{X}_J) \quad\le\quad [\sup_{u\in(0,1)} |g'_{iJ}(u)|]^2 \cdot \mathrm{Var}\,\overline{X}_J$$

so that conditioning on $\alpha_J \le \overline{X}_J \le \beta_J$, which forces $\mathrm{Var}\,\overline{X}_J \to 0$ and hence $\mathrm{Var}\, g_{iJ}(\overline{X}_J) \to 0$ as $J \to \infty$, also forces the second term in (17) to go to zero, completing the proof. $\square$

Now we are ready to state and prove the two converses to Theorem 5.1.

THEOREM 5.2 *Suppose* $\boldsymbol{X}$ *is sequence of binary response variables and* Θ *is a unidimensional variable, and suppose* (4), (13) *and* (14) *hold. Then*

(a) CA, $d_E = 1$, LCSN, MM $\Leftrightarrow$ $d_L = 1$, LAD;

(b) CA, $d_E = 1$, CSN, MM $\Rightarrow$ $d_L = 1$, LAD.

REMARKS. In the implications "$\Rightarrow$" in (a) and (b), Θ is the latent variable with respect to which $d_E = 1$ holds, and the theorem asserts that in fact $d_L = 1$ holds with respect to this Θ. In the implication "$\Leftarrow$" in (a), Θ is the latent variable with respect to which $d_L = 1$ holds. In both cases, Θ is unique up to monotone transformation, by Theorem 3.3.

PROOF. It is more convenient to prove (b) first.

Part (b), "$\Rightarrow$": There are three conditions to check on the right: LI, M, and LAD. LAD follows from $d_E = 1$ by definition. M follows from MM and $d_E = 1$ via Proposition 4.1(b). LI follows from CA, $d_E = 1$ and LCSN, using Proposition 3.1, Proposition 3.2, and Lemma 5.1(a), since (15) implies that under CSN and $d_E = 1$ we have $\mathrm{Cov}\,(X_i, X_j|\theta) \leq 0$ for all i, j and θ.

Part (a), "$\Rightarrow$": Again we must check LI, M and LAD. LAD and M follow as before. LI follows again, using Proposition 3.1, Proposition 3.2, and Lemma 5.1(b), since now (16) implies that under LCSN $\mathrm{Cov}\,(X_i, X_j|\theta) \leq 0$ for all i, j, and θ (a conditional [given $\Theta = \theta$] form of Lemma 3.1 is needed to show this, but this is straightforward).

Part (a), "$\Leftarrow$": This is Theorem 5.1, but we state the proof for completeness. We must check MM, CA, EI, LAD and LCSN. MM and CA follow from $d_L = 1$ by Proposition 4.1(a) and Theorem 3.1, respectively. EI follows from LI trivially, LAD is assumed on the right, and LCSN follows from LI via a conditional form of Theorem 4.1. □

6. Concluding Remarks. Considerable attention has been paid to the development of nonparametric conditions on $P[\boldsymbol{X}_J = \boldsymbol{x}_J]$ that characterize a $d_L = 1$ (locally independent, monotone, unidimensional) latent variable representation for the binary items $\boldsymbol{X}_J = (X_1, \ldots, X_J)$. In this paper we have examined the relationships between conditional association (CA), essential unidimensionality ($d_E = 1$), and useful, strictly unidimensional ($d_L = 1$), latent variable representations for binary item response data.

Both CA and $d_E = 1$ follow from a $d_L = 1$ representation which is *useful* in the sense of Definition 2.1. Conversely, when both CA and $d_E = 1$ hold, Proposition 3.1 provides a unidimensional Θ such that the conditional distribution of $\boldsymbol{X}$ given Θ is associated. But $d_L = 1$ requires the stronger local independence (LI) and monotonicity (M) conditions of Section 1. If, in addition to CA and $d_E = 1$, the negative association condition CSN, $\mathrm{Cov}\,(X_i, X_j|\overline{X}_J) \leq 0$, is also satisfied, LI results. Proposition 4.1 shows that monotonicity of the empirical response curves $P[X_j = 1|\overline{X}_J - X_j/J]$ is intimately related to M: this "manifest monotonicity" (MM) must hold if $d_L = 1$ holds; and conversely it can be used to verify M when $d_E = 1$ holds.

Theorem 5.2 summarizes these relationships: Part (a) characterizes $d_L = 1$ representations among "smooth" representations satisfying the mild monotonicity condition LAD—this is essentially the class of useful $d_L = 1$ representations, assuming that the distribution of Θ is not concentrated at one point—in terms of CA, $d_E = 1$, MM, and a "local" version of CSN. Part (b) gives reasonably general conditions—CA, $d_E = 1$, MM and CSN itself—on the manifest structure of $\boldsymbol{X}$ that are sufficient to guarantee useful $d_L = 1$ latent structure. These results suggest that conditions like CSN and MM will be needed to produce a more general characterization of $d_L = 1$ latent structure.

The approach taken in this paper is somewhat novel in the context of latent variable modeling, in that it explicitly embeds the observable responses $\boldsymbol{X}_J$ for each subject in an infinite response sequence $\boldsymbol{X}$ of responses for the same subject. This embedding seems absolutely vital to clarify estimation and model-identification issues. Haberman (1977) treats joint maximum likelihood estimation of θ and the $P_j(\theta)$ in this fashion for some important exponential family cases; Levine (1992) details some of the limitations of our ability to know Θ from $\boldsymbol{X}_J$ for finite J; and Stout (1990) makes a determined case for interpreting IRT applications in terms of this embedding.

Overly restrictive parametric assumptions, such as detailed knowledge of the forms of the response curves or of the distribution of Θ, are not needed in our approach. However, we must explicitly employ some form of monotonicity or response function smoothness to avoid meaningless models. Our preferred "nonparametric" condition has been Stout's local asymptotic discrimination (LAD) condition. In settings in which the response curves $P_j(\theta)$ are themselves parametrized (e.g. Jannarone, 1986; Thissen and Steinberg, 1986), a general monotonicity condition such as LAD might be dropped in the face of other smoothness available from the parametric form of the model. However LAD is often plausible, even if the individual response curves are not monotone, and greatly enhances the interpretability of the model.

Acknowledgements. A preliminary version of this material appeared as Chapter 4 of the author's Ph. D. dissertation. This work was greatly facilitated by enjoyable and stimulating discussions with Bill Stout. Comments by Bertrand Clarke, Paul Holland, and an anonymous referee were very much appreciated. In addition, the author is grateful to Ivo Molenaar for the examples appearing in Section 4.2.

REFERENCES

Bartholomew, D. J. (1987). *Latent variable models and factor analysis.* Oxford University Press. New York.

Birnbaum, A. (1968). *Some latent trait models and their use in inferring an examinee's ability*, in Lord, F. M. and Novick, M. R. (1968). *Statistical Theory of Mental Test Scores.* Addison-Wesley. Reading, Massachusetts.

Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics, 43,* 783–791.

Clarke, B. S. and Ghosh, J. K. (1991). Posterior convergence given the mean, with applications. Technical Report #91-68. Department of Statistics, Purdue University. West Lafayette, IN. Submitted for publication.

Clarke, B. S. and Junker, B. W. (1991). Inference from the product of marginals of a dependent likelihood. Technical Report #508. Department of Statistics, Carnegie Mellon University. Pittsburgh, PA. Submitted for publication.

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

Darroch, J. N, Fienberg, S. E., Glonek, G. F. V. and Junker, B. W. (1991). A three-sample multiple-perceptive approach with heterogeneous catchability for census population estimation. Carnegie Mellon University Statistics Department Technical Report #543. Submitted for publication.

Eaton, W. W. and Bohrnstedt G., eds. (1989). Latent variable models for dichotomous outcomes: analysis of of data from the epidemiological catchement area program. Special issue of *Sociological Methods and Research, 18,* 3–182.

Esary, J. D., Proschan, F. and Walkup, D. W. (1967). Association of random variables, with applications. *Annals of Mathematical Statistics, 38,* 1466–1474.

Fienberg, S. E. (1986). The Rasch Model, in Kotz, S., Johnson, N. L. and Read, C. B., eds., *Encyclopedia of Statistical Sciences, 7,* 627–632. John Wiley and Sons, Inc. New York.

Grayson, D. A. (1988). Two-group classification in latent trait theory: scores with monotone likelihood ratio. *Psychometrika, 53,* 383–392.

Haberman, S. (1977). Maximum likelihood estimates in exponential response models. *The Annals of Statistics, 5,* 815–841.

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

Holland, P. W. (1990). On the sampling theory foundations of item response theory models. *Psychometrika, 55,* 577–601.

Holland, P. W. and Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent trait models. *Annals of Statistics, 14,* 1523–1543.

Jannarone, R. J. (1986). Conjunctive item response theory kernels. *Psychometrika, 51,* 357–373.

Joag-Dev, K. (1983). Independence via uncorrelatedness under certain dependence structures. *Annals of Probability, 11,* 1037–1041.

Joag-Dev, K. and Proschan, F. (1982). Negative association of random variables, with applications. *Annals of Statistics, 10,* 286–295.

Junker, B. W. (1990). Progress in characterizing strictly unidimensional IRT representations. Office of Naval Research, Research Report 90-1. Department of Statistics, University of Illinois at Urbana-Champaign.

Junker, B. W. (1991). Essential independence and likelihood-based ability estimation for polytomous items. *Psychometrika, 56,* 255–278.

Lehmann, E. L. (1955). Ordered families of distributions. *Annals of Mathematical Statistics, 26,* 399–419.

Levine, M. V. (1992). Orthogonal functions and the finiteness of item response theories. Model-based Measurement Laboratory, Department of Educational Psychology, University of Illinois at Urbana-Champaign. Submitted for publication.

Lord, F. M. (1980). *Application of Item Response Theory to Practical Testing Problems.* Lawrence Erlbaum Associates, Inc. Hillsdale, NJ.

Molenaar, I. (July 1990). Personal communication.

Newman, C. M. and Wright, A. L. (1981). An invariance principle for certain dependent sequences. *Annals of Probability, 9,* 671–675.

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 Journal of Mathematical and Statistical Psychology, 38,* 206–215.

Rosenbaum, P. R. (1987). Probability inequalities for latent scales. *British Journal of Mathematical and Statistical Psychology, 40,* 157–168.

Rosenbaum, P. R. (1988). Item bundles. *Psychometrika, 53,* 349–359.

Sanathanan, L. (1972). Model and estimation methods in visual scanning experiments. *Technometrics, 14,* 813–829.

Stout, W. F. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika, 52,* 589–617.

Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika, 55,* 293–325.

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

Thissen, D. and Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika, 51,* 567-577.

Zwick, R. (1987) Assessing the dimensionality of NAEP reading data. *Journal of Educational Measurement, 24,* 293–308.