A CHARACTERIZATION OF MONOTONE UNIDIMENSIONAL LATENT VARIABLE MODELS

BY BRIAN W. JUNKER[1] AND JULES L. ELLIS[2]

Carnegie Mellon University and University of Nijmegen

Recently, the problem of characterizing monotone unidimensional latent variable models for binary repeated measures was studied by Ellis and van den Wollenberg and by Junker. We generalize their work with a de Finetti–like characterization of the distribution of repeated measures $\mathbf{X} = (X_1, X_2, \ldots)$ that can be represented with mixtures of likelihoods of independent but not identically distributed random variables, where the data satisfy a stochastic ordering property with respect to the mixing variable. The random variables X_j may be arbitrary real-valued random variables. We show that the distribution of $\mathbf{X}$ can be given a monotone unidimensional latent variable representation that is *useful* in the sense of Junker if and only if this distribution satisfies *conditional association* (CA) and a *vanishing conditional dependence* (VCD) condition, which asserts that finite subsets of the variables in $\mathbf{X}$ become independent as we condition on a larger and larger segment of the remaining variables in $\mathbf{X}$. It is also interesting that the mixture representation is in a certain ordinal sense unique, when CA and VCD hold. The characterization theorem extends and simplifies the main result of Junker and generalizes methods of Ellis and van den Wollenberg to a much broader class of models.

Exchangeable sequences of binary random variables also satisfy both CA and VCD, as do exchangeable sequences arising as location mixtures. In the same way that de Finetti's theorem provides a path toward justifying standard i.i.d.-mixture components in hierarchical models on the basis of our intuitions about the exchangeability of observations, this theorem justifies one-dimensional latent variable components in hierarchical models, in terms of our intuitions about positive association and redundancy between observations. Because these conditions are on the joint distribution of the observable data $\mathbf{X}$, they may also be used to construct asymptotically power-1 tests for unidimensional latent variable models.

1. Introduction. Latent variable models for measurement are extremely common in psychometrics [e.g., Bartholomew (1987)], developmental and cognitive psychology [e.g., DiBello, Stout and Roussos (1995), Huguenard et al. (1996) and Sijtsma and Junker (1996)], medical diagnosis and psychiatric epi-

demiology [e.g., Eaton and Bohrnstedt (1989) and Junker and Pilkonis (1993)], multiple recapture methods for estimating population sizes [e.g., George and Robert (1992) and Darroch, Fienberg, Glonek and Junker (1993)], as well as educational testing, systems reliability, population genetics, geology, chemistry, archaeology and other areas as surveyed by Holland and Rosenbaum (1986) and Basilevsky (1994). Typically one is interested in measuring (making inferences on) a latent variable Θ, not directly observed, on the basis of repeated noisy "looks" at Θ via the repeated measures $\mathbf{X} = (X_1, X_2, \ldots)$. Unlike the usual development of hierarchical Bayes and mixture models, the X_j's are not assumed to be exchangeable. Most common models for this situation (factor analysis, item response models, ordered latent class models etc.) entail the following assumptions:

1. *conditional independence* (CI), $\coprod \mathbf{X}|\Theta$, the X_j are conditionally independent (but perhaps not identically distributed), given Θ;
2. *unidimensionality* (U), $\Theta \in \mathbb{R}$, the real line;
3. *monotonicity* (M), $P[X_j > t|\theta]$ is nondecreasing in θ, for all j and all t.

The notation $\coprod \mathbf{X}|\Theta$ for conditional independence follows Dawid's (1979) convention. The distribution for Θ need not have support on the whole real line; thus both latent variable models (in which Θ is continuous) and ordered latent class models (in which Θ is discrete) may be considered. The stochastic ordering property M incorporates the notion that the X_j's really are "measures" of Θ; for example, Θ may be a disease state and the X_j's may be symptoms, or Θ may be a level of achievement and the X_j's may code correct and incorrect answers to test questions. We will refer to the three assumptions CI, U and M together as the *monotone unidimensional representation*. [Junker (1993) called the same representation "strictly unidimensional," to distinguish it from the "essentially unidimensional" models of Stout (1990).]

From the point of view of model building, both in psychometrics and in general, it is important that these assumptions restrict the finite-dimensional distributions of $\mathbf{X}$ in some way. Therefore we note that, while these three assumptions may be weakened in various ways, none may be entirely omitted. For example, it is easy to see [e.g., Suppes and Zanotti (1981) and Billingsley (1986), page 276] that if condition M is fully relaxed, then any distribution for $\mathbf{X}$ can be represented as a unidimensional, conditionally independent mixture. A more complete discussion of these assumptions, from a similar point of view, is given by Junker (1993).

For continuous X_j's a familiar example of the monotone unidimensional representation is the one-dimensional factor analysis model,

$$X_j = a_j \Theta + \varepsilon_j, \qquad i = 1, 2, \ldots,$$

where the a_j are fixed nonnegative constants and the ε_j are distributed independently of each other and of Θ. If the $a_j \equiv 1$, this is also known as the errors-in-variables model. For discrete X_j, item response models provide a convenient example: letting $X_j \in \{0, 1\}$ and assuming each $P[X_j = 1|\theta] \equiv P_j(\theta)$

is nondecreasing in θ, these models state that, for each J,

$$P(X_1 = x_1, \ldots, X_J = x_J) \equiv p(x_1, \ldots, x_J)$$

$$(1) \qquad = \int \prod_{j=1}^{J} P_j(\theta)^{x_j}[1 - P_j(\theta)]^{1-x_j} \, dF(\theta).$$

The problem we take up in this paper is as follows. In practice we only get to see i.i.d. replications of the repeated measures vector $\mathbf{X}_J = (X_1, X_2, \ldots, X_J)$, and we must guess what model, or class of models, makes sense. Partly this is, and should be, done on substantive grounds, but it is also important to ask what features the joint distribution of $(X_1, X_2, \ldots, X_J)$ must satisfy in order for the monotone unidimensional representation to hold. Thinking about these features is helpful in deciding whether a unidimensional latent variable model is appropriate for the data.

Our main result is an asymptotic characterization of monotone unidimensional representations that satisfy a consistent estimation condition, in terms of two easy-to-state conditions on the joint distribution of an infinite sequence of measures $(X_1, X_2, \ldots)$ into which $(X_1, X_2, \ldots, X_J)$ has been embedded. For example, if $(X_1, X_2, \ldots, X_J)$ are questions on a math test, then $(X_{J+1}, X_{J+2}, \ldots)$ are just more math questions of a similar nature. Such an embedding, conceptually not much different from considering an infinite sequence of random variables in the law of large numbers or the central limit theorem, was introduced formally for examining latent structure by Stout (1987, 1990) as a way of addressing fundamental questions of identifiability and consistent estimation inherent in mixture representations. The characterization theorem we present extends and simplifies the main results of Junker (1993) and generalizes the methods of Ellis and van den Wollenberg (1993) to a much broader class of models.

There is a natural analogue for this problem in de Finetti's characterization of *exchangeability*. For example, for binary data, de Finetti's theorem says that the finite-dimensional distributions of $\mathbf{X}$ are invariant under permutations of the X_j's (i.e., they are exchangeable) if and only if a representation of the form (1) holds, with each $P_j(\theta)$ equal to a common $P(\theta)$ [e.g., Galambos (1982)]. Olshen (1974) and Aldous (1981) present related characterizations for the distributions of more general exchangeable sequences $\mathbf{X}$. Many results in this direction essentially determine what structure the tail σ-field of $\mathbf{X}$ (defined in Section 3) must have in order to produce a representation like (1), and we will take this tack also. A rather different direction has been pursued by, for example, Diaconis and Freedman (1984) and Lauritzen (1988).

Note, however, that our situation is somewhat different from those in which exchangeability of the X_j might be assumed: in most applications in which the monotone unidimensional representation would be attractive, it is known that the measures X_j do not have the same marginal distributions (e.g., some test questions are hard and others are easy), but there are not usually reliable covariates upon which to condition to obtain a partially exchangeable structure. We use the information we have by *not* assuming identical marginal

distributions in the monotone unidimensional representation; thus we seek conditions that in some sense generalize exchangeability to representations in which the X_j are conditionally independent but not identically distributed.

In Section 2 we introduce the two constraints on the joint distribution of $(X_1, X_2, \ldots)$ used in our theorem, the conditional association constraint of Holland and Rosenbaum (1986) and a vanishing conditional dependence constraint that is related to certain constraints in the papers of Ellis and van den Wollenberg (1993) and Junker (1993). In Section 3, we present two fundamental lemmas which help to relate these two observable conditions on the repeated measures to the structure of the tail σ-field of $(X_1, X_2, \ldots)$. Section 4 gives the main theorem and its proof, and in Section 5 we explore our two constraints in some simple examples, including simple instances of the factor analysis and item response models mentioned above.

2. Observable constraints.
Holland and Rosenbaum (1986) studied, extended and unified various notions of positive dependence that must hold for **X** whenever **X** satisfies a monotone unidimensional representation. The most important of these notions was based on the idea of *associated* random variables due to Esary, Proschan and Walkup (1967). Holland and Rosenbaum show that the monotone unidimensional representation implies *conditional association* (CA): for all J, all partitions of $(X_1, \ldots, X_J)$ into disjoint subsets $(\mathbf{Y}, \mathbf{Z})$, all nondecreasing $f(\cdots)$ and $g(\cdots)$, and all $\mathbf{h}(\cdots)$,

$$\text{(CA)} \qquad\qquad \text{Cov}\big(f(\mathbf{Y}), g(\mathbf{Y})|\mathbf{h}(\mathbf{Z}) = \mathbf{c}\big) \geq 0.$$

Thus, if each X_j is driven monotonically by the same Θ, then $(X_1, X_2, \ldots)$ possesses so much internal coherence that all nondecreasing summaries of **Y** should have nonnegative correlation, conditional on any information at all on the complementary set of measures **Z**. The CA condition is quite strong; no examples are known of distributions for **X** which satisfy CA but do not admit a monotone unidimensional representation.

We introduce here a second condition, which we call *vanishing conditional dependence* (VCD), implied by any monotone unidimensional representation that is *useful* in the sense of Junker [(1993), Definition 2.1]. Suppose the monotone unidimensional representation holds, with Θ in the tail σ-field of **X**. Then by standard approximation arguments (see the proof of Theorem 4.1 below), for all partitions $(X_1, \ldots, X_J) = (\mathbf{Y}, \mathbf{Z})$ and all measurable $f(\cdots)$ and $g(\cdots)$,

$$\text{(VCD)} \qquad\qquad \lim_{m \to \infty} \text{Cov}\big(f(\mathbf{Y}), g(\mathbf{Z})|X_{J+1}, \ldots, X_{J+m}\big) = 0,$$

almost surely. Thus, repeated measures from a monotone unidimensional representation are strongly redundant: the information available from **Y** adds vanishingly little, as m grows, to that available from $(X_{J+1}, \ldots, X_{J+m})$ for predicting **Z**. VCD provides a simple condition, entirely in terms of the observable measures X_j, that ensures conditional independence. As we shall see, VCD also ensures the existence of consistent estimators of Θ.

Our main result, presented in Section 4, gives a characterization of the monotone unidimensional representation in terms of just CA and VCD that is applicable for arbitrary real-valued X_j's. Ellis and Junker (1996) consider this result from a psychometric point of view. Condition CA is pleasantly symmetric in the X_j's and can be checked, at least in principle, in whatever finite-dimensional distributions of **X** are available. In this respect, CA is very much like exchangeability. Condition VCD may also be formulated in a way that is symmetric in the X_j's (see Section 4.2), but it is fundamentally asymptotic in nature. This is less attractive from the point of view of thinking about whether the distribution of **X** will admit a monotone unidimensional representation. However, VCD seems to be a requirement, as our main theorem will show. It remains to be seen whether VCD is equivalent to some other, more finite-dimensional, condition on the distribution of **X**. Conditions CA and VCD are conditions on the joint distribution of observable measures $(X_1, X_2, \ldots, X_J)$ that become more constraining as J grows; hence they may be used to construct asymptotically power-1 tests of the monotone unidimensional representation.

3. Structure of the tail σ-field. Before presenting the main theorem, we present two interesting lemmas that elucidate the structure of the tail σ-field of **X**. The lemmas are not needed for understanding the statement and consequences of the theorem, and the reader may proceed directly to Section 4 after reviewing the definitions of Section 3.1.

3.1. *Some definitions.* Recall [Billingsley (1986), page 295] that the tail σ-field for the sequence $(X_1, X_2, \ldots)$ may be defined as

$$\tau(\mathbf{X}) = \bigcap_{n=1}^{\infty} \sigma\{X_j \colon j \geq n\}$$

where $\sigma(\cdots)$ is the Borel σ-field generated by "$\cdots$". It is useful to think of the tail σ-field as the set of "all" hypotheses and parameters for which there exist consistent inference procedures based on $X_1, X_2, \ldots$, even if we ignore some finite set of X_j's. In his discussion of latent variable models useful for measurement, Junker (1993) argues that if Θ is to be called a latent variable, it is sensible to require $\Theta \in \tau(\mathbf{X})$ [Θ measureable with respect to $\tau(\mathbf{X})$]—for then we can make arbitrarily precise inferences about Θ, but these inferences do not depend in any essential way on observing any particular X_j's. This corresponds to the notion of "trait validity," discussed, for example, by Messick (1989), in the construction of such models. Junker [(1993), Proposition 2.1] also shows that, for a certain class of monotone unidimensional latent variable models, $\sigma(\Theta) = \tau(\mathbf{X})$ holds almost surely (in a sense to be made precise following Lemma 3.1).

In this section, we consider the effects of conditioning on a general σ-field $\mathscr{F}$ that is contained in $\tau(\mathbf{X})$. To provide a bridge between the observable variables $X_1, X_2, \ldots$ and the σ-field $\mathscr{F}$, we shall define a set of *true scores* $\mathbf{T} = \{T_{iq} \colon i \in$

$\mathbb{N}$, $q \in \mathbb{Q}\}$, where $T_{iq} = P[X_i > q | \mathscr{F}]$, $\mathbb{N}$ is the set of natural numbers and $\mathbb{Q}$ is the set of rational numbers. The dichotomized random variables $Y_{iq} \equiv 1_{\{X_i > q\}}$, defined to be 1 when $X_i > q$ and 0 when $X_i \leq q$, are often used in the analysis of psychometric models [e.g., Samejima (1972) and Bartholomew (1987), Chapters 5 and 7]. Clearly $T_{iq} = E[Y_{iq} | \mathscr{F}]$, and if conditioning on the abstract σ-field $\mathscr{F}$ were replaced with conditioning on the latent variable Θ, $T_{iq} \equiv E[Y_{iq} | \Theta]$ would be recognized as a kind of dose–response function for responding above threshhold q given a "dose" Θ of the latent trait. The response functions T_{iq} are often called true scores in psychometrics.

We also define a kind of tail σ-field for the true scores,

$$\tau(\mathbf{T}) = \bigcap_{n=1}^{\infty} \sigma\{T_{iq} \colon i > n, \ q \in \mathbb{Q}\}.$$

3.2. *Two lemmas.* In Lemma 3.1, which generalizes Proposition 2.1 of Junker (1993), we show that everything that can be known about the conditional behavior of the X_i's given $\mathscr{F}$ can be learned from $\tau(\mathbf{T})$ alone, and moreover this tail σ-field is essentially identical to $\tau(\mathbf{X})$. In Lemma 3.2, which generalizes an important comonotonicity result of Ellis and van den Wollenberg (1993), we show that when CA holds the joint variation of the T_{iq} is greatly constrained.

LEMMA 3.1. *Suppose $\mathscr{F} \subseteq \tau(\mathbf{X})$ and $\coprod \mathbf{X} | \mathscr{F}$. Then $\mathscr{F} = \sigma(\mathbf{T}) = \tau(\mathbf{T}) = \tau(\mathbf{X})$, a.s.*

REMARKS. We use "$\mathscr{F} \subseteq \mathscr{G}$ a.s." (almost surely) to mean that for any set $F \in \mathscr{F}$ there is a set $G \in \mathscr{G}$ such that $P(F \Delta G) = 0$; and "$\mathscr{F} = \mathscr{G}$ a.s." means that the inclusion goes both ways. The equation $\mathscr{F} = \tau(\mathbf{X})$ in Lemma 3.1 states that if $\mathscr{F}$ is rich enough to induce CI, then $\mathscr{F}$ must "fill out" the entire tail σ-field. On the other hand, the equation $\sigma(\mathbf{T}) = \tau(\mathbf{T})$ tells us that the true scores T_{iq} are quite redundant, in the sense that, for any n, all T_{iq} for $i \leq n$ are completely determined by the T_{iq} with $i > n$.

PROOF OF LEMMA 3.1. By standard approximations, $\sigma(\mathbf{T})$ contains all conditional probabilities $T_{ir} = P[X_i > r | \mathscr{F}]$, $r \in \mathbb{R} \cup \{-\infty, \infty\}$. Now consider each equality of Lemma 3.1 in turn.

$\mathscr{F} = \sigma(\mathbf{T})$ a.s. From the definition of T_{iq}, $\sigma(\mathbf{T}) \subseteq \mathscr{F}$. For the reverse inclusion, it suffices to show that $P[A | \mathscr{F}] \in \sigma(\mathbf{T})$ for each $A \in \sigma(\mathbf{X})$; for then, if $A \in \mathscr{F}$, $1_A = P[A | \mathscr{F}] \in \sigma(\mathbf{T})$ a.s. Now if A is an interval of the form $\{a < X_j \leq b\}$, then $P[A | \mathscr{F}] = T_{ja} - T_{jb} \in \sigma(\mathbf{T})$; then monotone convergence and monotone class arguments show that $P[A | \mathscr{F}] \in \sigma(\mathbf{T})$ for all $A \in \sigma(X_j)$ as well. Next, if A is a cylinder set $\bigcap_1^J A_j$, $A_j \in \sigma(X_j)$, conditional independence implies $P[A | \mathscr{F}] = \prod_1^J P[A_j | \mathscr{F}] \in \sigma(\mathbf{T})$. Finally, we extend to $A \in \sigma(X_1, \ldots, X_J)$ and then to $A \in \sigma(X_1, X_2, \ldots)$ by considering the field of finite disjoint unions of cylinder sets and applying further monotone class arguments.

$\sigma(\mathbf{T}) = \tau(\mathbf{T})$ *a.s.* It is enough to show $\tau(\mathbf{T}) = \mathscr{F}$ a.s., and again $\tau(\mathbf{T}) \subseteq \mathscr{F}$ by definition of the T_{iq}'s. For the reverse inclusion, consider $A \in \mathscr{F}$ and define $P_n(A) \equiv P[A|\sigma\{T_{iq}: i \geq n, \ q \in \mathbb{Q}\}]$, which is a bounded martingale converging to $P[A|\tau(\mathbf{T})]$ as $n \to \infty$ by reverse martingale convergence. Now, since $(X_n, X_{n+1}, \ldots)$ has the same tail σ-field as $\mathbf{X}$, it follows from the previous paragraph that $\mathscr{F} = \sigma\{T_{iq}: i \geq n, \ q \in \mathbb{Q}\}$ a.s., for each n. So $1_A = P_n(A)$ a.s. for all n, and hence $1_A = P[A|\tau(\mathbf{T})]$ a.s. This shows $\mathscr{F} \subseteq \tau(\mathbf{T})$ a.s., as required.

$\tau(\mathbf{T}) = \tau(\mathbf{X})$ *a.s.* Obviously $\tau(\mathbf{T}) \subseteq \tau(\mathbf{X})$. For the reverse inclusion, the classical 0–1 law for independent random variables (conditional on $\mathscr{F}$) implies that for any $A \in \tau(\mathbf{X})$, $P[A|\mathscr{F}] = 1_{A'}$ for some $\mathscr{F}$-measurable set A', and it is easy to show that $P[A \Delta A'] = 0$. $\square$

For the next lemma, we define two random variables S and T to be *comonotone* [cf. Schmeidler (1989) or Wakker (1989)] if there is an almost-sure set $C \in \sigma(S, T)$ such that

$$\forall \ v, w \in C, \qquad S(v) > S(w) \quad \Rightarrow \quad T(v) \geq T(w).$$

It is easy to show that comonotonicity is a symmetric relationship in S and T.

LEMMA 3.2. *Suppose* $\coprod \mathbf{X}|\mathscr{F}$, $\mathscr{F} \subseteq \tau(\mathbf{X})$, *and suppose, for all J, all nondecreasing $f(\cdots)$ and $g(\cdots)$ and all $A \in \mathscr{F}$ for which the covariance below is defined,*

$$(*) \qquad \mathrm{Cov}\left(f(X_1, \ldots, X_J), g(X_1, \ldots, X_J)|A\right) \geq 0.$$

Then every pair (T_{iq}, T_{jr}) is comonotone.

REMARKS. If the distribution of $\mathbf{X}$ satisfies CA, and $\mathscr{F} \subseteq \tau(\mathbf{X})$, then automatically condition $(*)$ is satisfied, since any set in $\mathscr{F}$ can be approximated using conditions of the form $\mathbf{h}(\mathbf{Z}) = \mathbf{c}$ on the right in CA. [Indeed, for any $m > J$, $\mathscr{F} \subseteq \tau(\mathbf{X}) \subseteq \sigma(X_m, X_{m+1}, \ldots) = \bigcup_n \sigma(X_m, X_{m+1}, \ldots, X_n)$, so that any set $A \in \mathscr{F}$ can be approximated arbitrarily well by some $A_n \in \sigma(X_m, X_{m+1}, \ldots, X_n)$, in the sense that $\lim_{n \to \infty} P(A \Delta A_n) = 0$, and therefore

$$\lim_{n \to \infty} \left| \mathrm{Cov}\left(f(\mathbf{X}_J), g(\mathbf{X}_J)|A\right) - \mathrm{Cov}\left(f(\mathbf{X}_J), g(\mathbf{X}_J)|A_n\right) \right| = 0;$$

see the appendix of Ellis and Junker (1996) for further details.]. Accordingly, when CA holds, the variation of the true scores T_{iq} is severely restricted. It is easy to see that two random variables are comonotone if and only if each is a monotone function of a third random variable. Since every pair (T_{iq}, T_{jr}) is comonotone, this suggests that we look for a common variable Θ and monotone functions f_{iq} such that $T_{iq} = f_{iq}(\Theta)$.

PROOF OF LEMMA 3.2. $i \neq j$. By also conditioning on $\mathscr{F}$ and using the fact that $\coprod \mathbf{X}|\mathscr{F}$, it follows that

$$(**) \qquad \mathrm{Cov}\left(T_{iq}, T_{jr}|A\right) \geq 0$$

for all $A \in \mathscr{F}$ for which the covariance is defined. Let $B_\delta(s, t)$ be the ball of radius δ about $(s, t) \in \mathbb{R}^2$, and consider the almost-sure set

$$C' = \{(s, t): P[(T_{iq}, T_{jr}) \in B_\delta(s, t)] > 0, \ \forall \ \delta > 0\}$$

for a fixed pair (T_{iq}, T_{jr}). [The set C' is sometimes called the closed support of the distribution; see Billingsley (1986), page 181.] We will show that C' cannot contain (s_1, t_1) and (s_2, t_2) with $s_1 < s_2$ and $t_1 > t_2$; using this fact, it follows immediately that (T_{iq}, T_{jr}) is comonotone on the almost-sure set $C = \{w: (T_{iq}, T_{jr})(w) \in C'\}$. The following geometric argument is adapted from Ellis and van den Wollenberg (1993); since it is short we repeat it here for clarity.

Suppose, by way of contradiction, that the set C' contains two points (s_1, t_1) and (s_2, t_2) with $s_1 < s_2$ and $t_1 > t_2$. Let $A = B_{\delta_1}(s_1, t_1) \cup B_{\delta_2}(s_2, t_2)$, where for sufficiently small δ_1 and δ_2 the union is disjoint, let $(X, Y) = (T_{iq}, T_{jr})$ and let $Z = 1$ or 2 according as (X, Y) is in $B_{\delta_1}(s_1, t_1)$ or $B_{\delta_2}(s_2, t_2)$. [Let $Z = 0$ otherwise, but this will not be important.] If we condition on the event $\{w: (X, Y)(w) \in A\}$, but drop the conditioning from the notation for simplicity, then from $(\ast\ast)$ we have

$$0 \le \mathrm{Cov}(X, Y) = E[\mathrm{Cov}(X, Y | Z)] + \mathrm{Cov}\big(E[X | Z], E[Y | Z]\big) \equiv \mathrm{I} + \mathrm{II}.$$

From the Cauchy–Schwarz inequality, we know $\mathrm{I} \le p 4 \delta_1^2 + (1 - p) 4 \delta_2^2$, where $p = P[Z = 1]$; and calculation shows that

$$\mathrm{II} = p(1 - p)\big(E[X | Z = 1] - E[X | Z = 2]\big)\big(E[Y | Z = 1] - E[Y | Z = 2]\big)$$

$$\le -p(1 - p)(s_2 - s_1 + \delta_1 + \delta_2)(t_1 - t_2 + \delta_1 + \delta_2).$$

Hence

$$\mathrm{Cov}(X, Y) \le 4(\delta_1^2 + \delta_2^2) - p(1 - p)(s_2 - s_1 + \delta_1 + \delta_2)(t_1 - t_2 + \delta_1 + \delta_2).$$

If we now let δ_1 and δ_2 tend to zero in such a way that $p(1 - p)$ is bounded below by some $\varepsilon > 0$, we will clearly have $\mathrm{Cov}(X, Y) < 0$, contradicting $(\ast\ast)$.

$i = j$. By Lemma 3.1, we know that $T_{iq}, T_{ir} \in \sigma(T_{js}: j > i, \ s \in \mathbb{Q})$ a.s., and since $\sigma(T_{iq}, T_{ir})$ is countably generated we can construct an almost-sure set A such that $\sigma(T_{iq}, T_{ir}) \cap A \subseteq \sigma(T_{js}: j > i, \ s \in \mathbb{Q}) \cap A$. Thus, T_{iq} and T_{ir} are really functions of the T_{js} on A. Let C'' be an almost-sure set on which each of T_{iq} and T_{ir} is comonotone with all T_{js}, $j > i$, $s \in \mathbb{Q}$ (available by countable applications of the case $i \ne j$), and consider $w, v \in C \equiv C'' \cap A$. If $T_{iq}(v) > T_{iq}(w)$, then for some $j > i$, $T_{js}(v) \ne T_{js}(w)$, which by the case $i \ne j$ forces $T_{js}(v) > T_{js}(w)$. Therefore $T_{ir}(v) \ge T_{ir}(w)$, again by the case $i \ne j$. $\square$

4. The monotone unidimensional representation.

4.1. *The main result.* Theorem 4.1 is a characterization of distributions on $\mathbf{X}$ for which the monotone unidimensional representation holds, with respect to some $\Theta \in \tau(\mathbf{X})$. On the other hand, it is easy to construct models for $\mathbf{X}$ in which the monotone unidimensional representation holds, but $\Theta \notin \tau(\mathbf{X})$;

see Example 5.4. As observed in Section 3, however, $\Theta \in \tau(\mathbf{X})$ is a natural condition to impose on latent variable models.

THEOREM 4.1. *Let* $\mathbf{X} = (X_1, X_2, \ldots)$ *be any sequence of real-valued random variables.*

Part 1. *The following three conditions are equivalent*:

(a) *There exists* $\Theta \in \tau(\mathbf{X})$ *such that the monotone unidimensional representation holds.*

(b) *There exists a* σ-*field* $\mathscr{F} \subseteq \tau(\mathbf{X})$ *such that* (i) $\coprod \mathbf{X}|\mathscr{F}$ *and* (ii) *condition* (∗) *of Lemma* 3.2 *holds.*

(c) *Conditions CA and VCD hold for* $\mathbf{X} = (X_1, X_2, \ldots)$.

Part 2. *When any (hence all) of the above conditions hold, then* $\sigma(\Theta) = \mathscr{F} = \tau(\mathbf{X})$, *a.s.*

Part 3. *If condition* (a) *holds for both* Θ_1 *and* Θ_2, *then these* Θ*'s are strictly increasing functions of one another, a.s.*

REMARKS. This theorem gives a de Finetti–style characterization of the monotone unidimensional representation. In particular, Part 1(c) of the theorem gives "observable" criteria, CA and VCD, for including monotone unidimensional latent variable components in a statistical model, in much the same way that exchangeability is an "observable" condition for including conditionally i.i.d. components in a statistical model. Part 2 says that if the monotone unidimensional representation holds with respect to $\Theta \in \tau(\mathbf{X})$, then $\sigma(\Theta)$ must fill out the whole tail σ-field of $\mathbf{X}$; this is a consequence of Lemma 3.1. Part 3 gives a uniqueness result that is important from a model-building perspective: if the monotone unidimensional representation holds, it holds with respect to an essentially unique Θ. Part 3 also expresses formally the notion that in general the monotone unidimensional representation leads to an essentially ordinal level of measurement for the latent trait (i.e., Θ is identified only up to an arbitrary strictly increasing transformation).

PROOF OF THEOREM 4.1. (a) $\Rightarrow$ (c). Holland and Rosenbaum (1986) show that the monotone unidimensional representation implies CA. To obtain VCD from the monotone unidimensional representation, we observe that, by Lemma 3.1, $\sigma(\Theta) = \tau(\mathbf{X})$ a.s.; hence $\coprod \mathbf{X}|\Theta$ implies $\coprod \mathbf{X}|\tau(\mathbf{X})$, and in particular $\coprod (X_1, \ldots, X_J)|\{\sigma(X_{J+1}, \ldots, X_{J+m}), \tau(\mathbf{X})\}$. Now let $\mathbf{Y}$ and $\mathbf{Z}$ be disjoint sets of variables from $(X_1, \ldots, X_J)$; we can use standard martingale convergence arguments [e.g., Billingsley (1986), Theorems 35.5 and 35.7] to show that

$$\lim_{m \to \infty} \mathrm{Cov}\left(f(\mathbf{Y}), g(\mathbf{Z})|\sigma(X_{J+1}, \ldots, X_{j+m}), \tau(\mathbf{X})\right)$$

$$= \mathrm{Cov}\left(f(\mathbf{Y}), g(\mathbf{Z})|\sigma(X_{J+1}, X_{J+2}, \ldots)\right)$$

$$= \lim_{m \to \infty} \mathrm{Cov}\left(f(\mathbf{Y}), g(\mathbf{Z})|\sigma(X_{J+1}, \ldots, X_{J+m})\right),$$

for any (measurable) functions $f(\cdots)$ and $g(\cdots)$. From this we can deduce VCD.

(c) $\Rightarrow$ (b). Take $\mathscr{F} = \tau(\mathbf{X})$. Then for any (measurable) f and g, and any $n > 0$,

$$\lim_{m \to \infty} \mathrm{Cov}\left(f(X_1, \ldots, X_J), g(X_1, \ldots, X_J)|\sigma(X_{J+n}, \ldots, X_{J+m})\right)$$

$$= \mathrm{Cov}\left(f(X_1, \ldots, X_J), g(X_1, \ldots, X_J)|\sigma(X_{J+n}, \ldots)\right)$$

$$\to \mathrm{Cov}\left(f(X_1, \ldots, X_J), g(X_1, \ldots, X_J)|\tau(\mathbf{X})\right)$$

as n tends to ∞, by (reverse) martingale convergence. If we require f and g to depend on disjoint subsets of $X_1, \ldots, X_J$, we obtain $\coprod \mathbf{X}|\tau(\mathbf{X})$ from VCD. If we merely require f and g to be nondecreasing, we obtain $(*)$, with $\mathscr{F} = \tau(\mathbf{X})$, from CA, as in the remark following Lemma 3.2.

(b) $\Rightarrow$ (a). As suggested in the remarks following Lemma 3.2, we can directly construct a $\Theta \in \mathbb{R}$ and show that CI and M hold for this Θ. Indeed, for any particular fixed ordering of the rationals $q \in \mathbb{Q}$, choose $a_{iq} > 0$ such that $\sum_{i=1}^{\infty} \sum_q a_{iq} < \infty$, and define

$$\Theta = \sum_{i=1}^{\infty} \sum_{q \in \mathbb{Q}} a_{iq} T_{iq} \in \mathscr{F} \subseteq \tau(\mathbf{X}).$$

Let C be a common almost-sure set on which all pairs (T_{iq}, T_{jr}) are comonotone (available by countable applications of Lemma 3.2). We observe the following:

(i) On C, $\Theta(v) > \Theta(w)$ implies that there must be some $T_{iq}(v) > T_{iq}(w)$; hence, by Lemma 3.2, $T_{jr}(v) \geq T_{jr}(w)$ for all j and r. It follows that *each T_{iq} is a monotone function of Θ, a.s.*

(ii) By observation (i), $T_{iq} \in \sigma(\Theta)$ a.s., so from Lemma 3.1 we may deduce

$$\mathscr{F} = \sigma(\mathbf{T}) = \tau(\mathbf{T}) = \tau(\mathbf{X}) \subseteq \sigma(\Theta) \subseteq \mathscr{F} \quad \text{a.s.}$$

Hence, almost surely, $T_{iq} = P[X_i > q|\mathscr{F}] = P[X_i > q|\Theta]$ is nondecreasing in Θ, which is condition M of the monotone unidimensional representation, and $\coprod \mathbf{X}|\mathscr{F}$ implies $\coprod \mathbf{X}|\Theta$, which is condition CI of the monotone unidimensional representation.

This proves (b) $\Rightarrow$ (a), as well as Part 2 of the theorem.

For Part 3, let Θ_1 and Θ_2 be two random variables satisfying Part 1(a) of the theorem, and use Lemma 3.1 with $\mathscr{F}_k = \sigma(\Theta_k)$, for each $k = 1, 2$, to show that $\sigma(\Theta_1) = \tau(\mathbf{X}) = \sigma(\Theta_2)$, a.s. Therefore $\Theta_2 = f(\Theta_1)$ a.s. for some invertible measurable function $f(\cdot)$; and moreover the true scores $T_{iq} = P[X_i > q|\mathscr{F}_k] = P[X_i > q|\Theta_k]$, $k = 1, 2$, are equal. Now using the monotonicity assumption M for each Θ_i, all pairs (Θ_1, T_{iq}) and (Θ_2, T_{iq}) must be comonotone on some common almost-sure set C. Therefore, if $\Theta_1(v) > \Theta_1(w)$ for $v, w \in C$, then there must be some $T_{iq}(v) > T_{iq}(w)$, and hence $\Theta_2(v) \geq \Theta_2(w)$. Since Θ_1 and Θ_2 are therefore comonotone, it follows that $f(\cdot)$ may be taken to be strictly increasing. $\square$

4.2. *A symmetric VCD condition.* In Section 2 we observed that CA is a nonasymptotic symmetric condition on the sequence $(X_1, X_2, \ldots)$ but VCD appears to be both asymptotic and asymmetric, depending on the order in which the X_j are encountered. However,

$$\tau(\mathbf{X}) = \bigcap_n \sigma(X_{n+1}, X_{n+2}, \ldots)$$

$$= \bigcap_{n \in \mathbb{N}} \bigcap_{\mathscr{K} \subseteq \{1,\ldots,n\}} \sigma(X_j \colon j \in \mathbb{N} \setminus \mathscr{K})$$

$$= \bigcap_{\{\mathscr{K} \subseteq \mathbb{N}, \; |\mathscr{K}| < \infty\}} \sigma(X_j \colon j \in \mathbb{N} \setminus \mathscr{K}),$$

where $\mathscr{K}$ extends over all finite subsets of the natural numbers $\mathbb{N}$, and the same argument works if the numbers $1, 2, \ldots$ are replaced with any permutation (i.e., any 1–1 function from $\mathbb{N}$ onto $\mathbb{N}$). Thus the tail σ-field $\tau(\mathbf{X}) = \bigcap_n \sigma(X_{n+1}, X_{n+2}, \ldots)$ does not depend on the ordering of the X_j.

Since the role of VCD in the proof of Theorem 4.1 was to be sensitive to conditional independence given $\tau(\mathbf{X})$, it follows immediately that VCD may be replaced in the statement of the theorem with an apparently more restrictive symmetric condition. Namely, we may assume that, for any finite subsets $\mathbf{Y}$ and $\mathbf{Z}$ from $(X_1, X_2, \ldots)$, and any permutation $(W_1, W_2, \ldots)$ of the remaining items in $\mathbf{X} \setminus \mathbf{Y} \cup \mathbf{Z}$,

$$\lim_{m \to \infty} \operatorname{Cov}\left(f(\mathbf{Y}), g(\mathbf{Z}) | W_1, W_2, \ldots, W_m\right) = 0,$$

for all f and g for which the covariance is defined. However, this is still a fundamentally asymptotic condition on the distribution of $(X_1, X_2, \ldots)$.

4.3. *True scores and multidimensional representations.* The technical arguments above were greatly facilitated by the use of the true scores $T_{iq} = P[X_i > q | \tau(\mathbf{X})]$ to provide a bridge between the X_j's and the tail σ-field. Holland (1990) discusses two standard formulations of the latent variable model in psychometrics and educational measurement; it is worth noting that our results apply within either of these formulations, since the definition of the T_{iq} does not depend on an a priori specification of a latent variable for the model. This point is explored further by Ellis and Junker (1996).

The fact that the true scores T_{iq} can be defined in a way that does not depend on an a priori specification of the latent variable(s) is important both for technical manipulations and for interpretation of the results. One can think of the T_{iq} as filling out a manifold in some possibly infinite dimensional space. The dimensionality of the latent space can then be understood as the dimensionality of the manifold "spanned" by the true scores (Ramsay, 1996). By exploiting the infinite item pool framework of Stout (1990), we have shown that CA and VCD hold in the distribution of $(X_1, X_2, \ldots)$ if and only if T_{iq} in fact trace out a one-dimensional curve in this space; and the latent variable may be thought of as naturally parametrizing this curve. In the proof, the role of VCD is to ensure conditional independence in the representation,

while CA guarantees unidimensionality via comonotonicity arguments. The central question in generalizing our results to characterizations of monotone d-dimensional representations $(d > 1)$ is to discover what replaces CA when the manifold of true scores is d-dimensional.

5. Examples. Example 5.1 gives some connections between Theorem 4.1 and characterizations of exchangeable sequences. In Example 5.2 we interpret the CA and VCD conditions of Theorem 4.1 in terms of the partial correlations of the observable variables $(X_1, X_2, \ldots)$ in a factor analysis model. In Example 5.3 we observe that for the Rasch item response model CA is equivalent to a well-known condition on the parameters of an equivalent log-linear model. In both examples, we show that mild conditions that guarantee VCD also ensure that $\Theta \in \tau(\mathbf{X})$; in general we expect that VCD would always be closely tied to this measurability property of Θ. We also show that Theorem 4.1 can be used to distinguish between one- and two-dimensional monotone representations. Finally, Example 5.4 displays a case in which the monotone unidimensional representation holds, but not with respect to a $\tau(\mathbf{X})$-measurable Θ. This suggests that Θ is "too rich"—there exist features of Θ that cannot be measured with $\mathbf{X}$ alone—and a simpler Θ can be found for which a (different) monotone unidimensional representation holds.

EXAMPLE 5.1 (Connections with exchangeable sequences). The VCD condition is always true for an exchangeable sequence, using an argument like that of Theorem 4.1, (a) $\Rightarrow$ (c). However, CA may fail for an exchangeable sequence, so—as one readily conjectures—a monotone unidimensional representation is not possible for arbitrary exchangeable sequences:

Let $\Theta = 1$ with probability p and $\Theta = 0$ with probability $1 - p$. Suppose $X_{01}, X_{02}, \ldots$ and $X_{11}, X_{12}, \ldots$ are two i.i.d. sequences, and consider the exchangeable sequence $X_j = \Theta X_{1j} + (1 - \Theta)X_{0j}$. For $x < y$ and the indicator random variables $1_{\{X_i > x\}}$ and $1_{\{X_j > y\}}$, it follows that

$$\text{Cov}\,(1_{\{X_i > x\}}, 1_{\{X_j > y\}})$$
$$= p(1 - p)(P[X_{1i} > x] - P[X_{0i} > x])(P[X_{1j} > y] - P[X_{0j} > y]).$$

This can fail to be nonnegative, despite the fact that $\text{Cov}\,(X_i, X_j)$ must be nonnegative for any exchangeable sequence; for example, consider the scale mixture with X_{0j} i.i.d. $N(0, 4)$, X_{1j} i.i.d. $N(0, 1)$ and $x = -y$. Thus $\mathbf{X}$ can be exchangeable, yet fail CA.

When $\text{Cov}\,(1_{\{X_i > x\}}, 1_{\{X_j > y\}}) \geq 0$ for all x and y, then X_i and X_j are said to be *positive quadrant dependent* [PQD; Lehmann (1966)]. For exchangeable sequences [although not in general; see Holland and Rosenbaum (1986)], PQD for all i and j implies CA: one observes that $\mathbf{X}$ will still be exchangeable given any $A \in \tau(\mathbf{X})$; from this and PQD, $(**)$ in the proof of Lemma 3.2 follows; and then arguing as in Theorem 4.1 one obtains CA as well as a monotone unidimensional representation for $\mathbf{X}$. Location mixtures, and indeed any i.i.d.

mixtures in which the X_j are stochastically ordered by Θ, provide examples of this.

EXAMPLE 5.2 (Errors-in-variables models and linear factor analysis). Consider first a sequence $(X_1, X_2, \ldots)$ satisfying the one-dimensional model

$$(2) \qquad X_j = a_j T + \varepsilon_j,$$

where T, $\varepsilon_j \sim$ i.i.d. $N(0, 1)$ and a_j are nonnegative constants. For the first $J + m$ variables in the sequence, we may directly compute the conditional covariance matrix of $(X_1, \ldots, X_J)$ given $(X_{J+1}, \ldots, X_{J+m})$ as

$$(3) \qquad \begin{aligned} &\Sigma_{(X_1,\ldots,X_J)\cdot(X_{J+1},\ldots,X_{J+m})} \\ &\quad = I_{J\times J} + \frac{1}{1 + \sum_{j=J+1}^{J+m} a_j^2} \begin{bmatrix} a_1^2 & a_1 a_2 & \cdots & a_1 a_J \\ a_2 a_1 & a_2^2 & \cdots & a_2 a_J \\ \vdots & \vdots & \ddots & \vdots \\ a_J a_1 & a_J a_2 & \cdots & a_J^2 \end{bmatrix}. \end{aligned}$$

It is easy to deduce from (3) that the partial covariance of any pair (X_i, X_j) conditional on any subset of the other X_k's must be nonnegative, which is consistent with CA.

[Regardless of whether representation (2) holds, Karlin and Rinott (1983), Theorems 2 and 3, show that nonnegativity of all possible partial covariances of pairs (X_i, X_j) is equivalent to *multivariate total positivity of order* 2 (MTP$_2$) for multivariate normal distributions. Combining this fact with Pitt's (1982) result that multivariate normals are associated, in the sense of Esary, Proschan and Walkup (1967), if and only if all pairwise unconditional covariances are nonnegative, we may deduce that CA implies MTP$_2$, for multivariate normals. It is an open question whether the converse implication also holds, for multivariate normals.]

Now let us consider the asymptotic condition VCD. If VCD is to hold, the conditional covariances in (3) must vanish as m grows; hence, VCD implies that

$$(4) \qquad \lim_{J \to \infty} \sum_{j=1}^{J} a_j^2 = \infty.$$

This is precisely the condition needed to ensure, for example, that

$$\lim_{J \to \infty} \frac{\sum_{j=1}^{J} a_j X_j}{\sum_{j=1}^{J} a_j^2} = T,$$

in L^2 and hence a.s.; it follows from this that $T \in \tau(\mathbf{X})$, a.s., as claimed by Theorem 4.1.

Finally, consider a two-dimensional model for the sequence $(X_1, X_2, \ldots)$:

$$(5) \qquad X_j = a_{1j} T_1 + a_{2j} T_2 + \varepsilon_j,$$

where T_i, $\varepsilon_j \sim$ i.i.d. $N(0, 1)$ and a_{ij} are nonnegative constants. It is easy to construct sequences of a_{ij}'s for which CA fails. For example, if $a_{11} = 1$, $a_{21} = 0$, $a_{12} = 0$, $a_{22} = 1$ and $a_{1j} = a_{2j} = 1$ for all $j > 2$, then

$$\mathrm{Cov}\,(X_1, X_2 | X_3 + \cdots + X_J = c) < 0,$$

violating CA. In this case, there *cannot* exist a $\Theta \in \mathbb{R}$ for which representation (5) can be converted into a monotone unidimensional representation for $(X_1, X_2, \ldots)$ using Θ, even if we abandon the linear factor model and normal distribution assumptions.

EXAMPLE 5.3 (The Rasch model). If there exist β_j such that logit $P_j(\theta) = \theta + \beta_j$ in the integral representation (1), then that representation is known as the random effects Rasch model. It is well-known [see, e.g., Cressie and Holland (1983) and Lindsay, Clogg and Grego (1991)] that in this case the integral representation may be converted to a log-linear representation,

$$(6) \qquad \log p(x_1, \ldots, x_J) = \alpha + \sum_{j=1}^{J} \beta_j x_j + \gamma(x_+),$$

where $x_+ = \sum_1^J x_j$, displaying an "i. but not i.d. part" $\sum_{j=1}^{J} \beta_j x_j$, and an "exchangeable part" $\gamma(x_+)$. Conversely, it is known that the log-linear representation (6) can be converted back to the integral form (1)—and hence satisfies CA—if and only if $\gamma(k)$ behaves like the log-moments of a nonnegative random variable.

The condition VCD is again closely related to the condition that $\Theta \in \tau(\mathbf{X})$. If $(X_1, \ldots, X_{J+m})$ satisfies the Rasch model, VCD requires that

$$p(x_1, \ldots, x_J | x_{J+1}, \ldots, x_{J+m}) \approx \prod_{j=1}^{J} p_i(x_i;\ x_{J+1}, \ldots, x_{J+m}),$$

an independence distribution for $(X_1, \ldots, X_J)$, as m grows. Intuitively this should be easy to achieve, since, using CI (conditional independence given θ),

$$p(x_1, \ldots, x_J | x_{J+1}, \ldots, x_{J+m})$$
$$= \int \prod_{j=1}^{J} P_j(\theta)^{x_j} [1 - P_j(\theta)]^{1-x_j}\, dF(\theta | x_{J+1}, \ldots, x_{J+m}),$$

and the posterior distribution $dF(\theta | x_{J+1}, \ldots, x_{J+m})$ must tend to a point mass, under suitable regularity conditions. The regularity conditions are available in many places: item response models are considered directly, for example, by Chang and Stout (1993). These conditions also ensure that the MLE is consistent for θ, which forces $\Theta \in \tau(\mathbf{X})$, just as in Example 5.2.

Here too it is easy to create examples for which the CA condition does not hold. Indeed, one can begin with a model in which

$$(7) \qquad \mathrm{logit}\, P_j(\theta_1, \theta_2) \equiv \log \frac{P_j(\theta_1, \theta_2)}{1 - P_j(\theta_1, \theta_2)} = a_{1j}\theta_1 + a_{2j}\theta_2 + \beta_j$$

and proceed exactly as in Example 5.2.

EXAMPLE 5.4 (A monotone unidimensional representation outside the scope of Theorem 4.1).   Consider binary $(X_1, X_2, \ldots)$ satisfying (1) with

$$\text{logit } P_{2j}(\theta) = \theta$$

$$\text{logit } P_{2j-1}(\theta) = \begin{cases} \theta, & \text{if } \theta < 0, \\ 0, & \text{if } 0 \leq \theta < 1, \\ \theta - 1, & \text{if } 1 \leq \theta. \end{cases}$$

Following the arguments of Example 5.3 it is easy to see that there is a monotone unidimensional representation for $X_1, X_2, \ldots$, with respect to a latent variable Θ that can be consistently estimated with $\hat{\theta}_n = \text{logit}(2/n) \sum_1^n X_{2j}$, and both CA and VCD hold for $\mathbf{X}$.

Now consider the subsequence $\mathbf{Y} = (Y_1, Y_2, \ldots) = (X_1, X_3, \ldots)$ of X's with odd index. Since the monotone unidimensional representation holds with respect to Θ for the entire sequence, it still holds for the subsequence $\mathbf{Y}$. However, $\Theta \notin \tau(\mathbf{Y})$, since in particular it is not possible consistently to estimate Θ from $\mathbf{Y}$ when $0 \leq \Theta < 1$. [It is still true that CA and VCD hold for the subsequence, so there must be another latent variable $\Psi \in \tau(\mathbf{Y})$ with respect to which a monotone unidimensional representation for $\mathbf{Y}$ is possible; indeed, $\Psi = P_1(\Theta)$ will do the trick.]

The monotone unidimensional representation for $\mathbf{Y}$ in terms of Θ described in Example 5.4 is outside the scope of Theorem 4.1 since CA and VCD hold, but $\Theta \notin \tau(\mathbf{Y})$ and hence is not consistently estimable from the Y_j's. It is also possible to construct examples in which CA holds and VCD fails, but a monotone unidimensional representation is still possible: for example, consider a sequence $\mathbf{X}$ consisting of five items satisfying the Rasch model of Example 5.3, followed by an infinite sequence of i.i.d. coin flips. Once again, $\Theta \notin \tau(\mathbf{X})$ and hence is not consistently estimable from the X_j's. In the former case since CA and VCD do hold, another monotone unidimensional representation can be found, in terms of a trait Ψ that is consistently estimable from the Y_j's. In the latter case no such Ψ or alternative representation exists, since VCD does not hold. Thus, while CA guarantees comonotonicity properties (see the Remarks following Lemma 3.2), VCD is a condition on the observable measures $\mathbf{X}$ that guarantees the existence of a consistently estimable latent trait or mixing parameter in the monotone unidimensional representation.

two referees for interesting and thoughtful commentary that improved our presentation.

REFERENCES

ALDOUS, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11** 581–598.

BASILEVSKY, A. (1994). *Statistical Factor Analysis and Related Methods: Theory and Applications*. Wiley, New York.

BARTHOLOMEW, D. J. (1987). *Latent Variable Models and Factor Analysis*. Oxford Univ. Press.

BILLINGSLEY, P. (1986). *Probability and Measure*. Wiley, New York.

CHANG, H. H. and STOUT, W. F. (1993). The asymptotic posterior normality of the latent trait in an IRT model. *Psychometrika* **58** 37–52.

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. (1993). A three-sample multiple-recapture approach to census population estimation with heterogeneous catchability. *J. Amer. Statist. Assoc.* **88** 1137–1148.

DAWID, A. P. (1979). Conditional independence in statistical theory. *J. Roy. Statist. Soc. Ser. B* **41** 1–31.

DIACONIS, P. and FREEDMAN, D. (1984). Partial exchangeability and sufficiency. In *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on "Statistics: Applications and New Directions"* 205–236. Indian Statist. Inst., Calcutta.

DiBELLO, L., STOUT, W. F. and ROUSSOS, L. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In *Cognitively Diagnostic Assessment* (P. Nichols, S. Chipman and R. Brennen, eds.) 361–389. Earlbaum, Hillsdale, NJ.

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

ELLIS, J. L. and JUNKER, B. W. (1996). Tail-measurability in monotone latent variable models. *Psychometrika*. To appear.

ELLIS, J. L. and VAN DEN WOLLENBERG, A. L. (1993). Local homogeneity in latent trait models. A characterization of the homogeneous monotone IRT model. *Psychometrika* **58** 417–429.

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

GALAMBOS, J. (1982). Exchangeability. In *Encyclopedia of Statistical Sciences* **2** 573–577. Wiley, New York.

GEORGE, E. I. and ROBERT, C. P. (1992). Capture–recapture estimation via Gibbs sampling. *Biometrika* **79** 677–683.

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. *Ann. Statist.* **14** 1523–1543.

HUGUENARD, B., LERCH, F. J., JUNKER, B. W., PATZ, R. J. and KASS, R. E. (1996). Working memory failure in phone based interaction. *ACM Transactions on Computer–Human Interaction*. To appear.

JUNKER, B. W. (1993). Conditional association, essential independence and monotone unidimensional item response models. *Ann. Statist.* **21** 1359–1378.

JUNKER, B. W. and PILKONIS, P. A. (1993). Personality and depression: modeling and measurement issues, with commentaries by H. C. Kraemer and R. D. Gibbons. In *Personality and Depression: A Current View* (M. H. Klein, D. J. Kupfer and M. T. Shea, eds.). Guilford, New York.

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

LAURITZEN, S. (1988). *Extremal Families and Systems of Sufficient Statistics*. Springer, New York.

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

LINDSAY, B., CLOGG, C. C. and GREGO, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *J. Amer. Statist. Assoc.* **86** 96–107.

MESSICK, S. (1989). Validity. In *Educational Measurement* (R. L. Linn, ed.), 3rd ed. Macmillan, New York.

OLSHEN, R. (1974). A note on exchangeable sequences. *Z. Wahrsch. Verw. Gebiete* **28** 317–321.

PITT, L. D. (1982). Positively correlated normal variables are associated. *Ann. Probab.* **10** 496–499.

RAMSAY, J. O. (1996). A geometrical approach to item response theory. *Behaviormetrika* **23** 3–16.

SAMEJIMA, F. (1972). A general model for free-response data. *Psychometrika Monograph No. 18, 37* (4, Part 2).

SCHMEIDLER, D. (1989). Subjective probability and expected utility with additivity. *Econometrika* **57** 571–587.

SIJTSMA, K. and JUNKER, B. W. (1996). A survey of theory and methods of invariant item ordering. *British J. Math. Statist. Psych.* **49** 79–105.

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.

WAKKER, P. P. (1989). *Additive Representations of Preferences. A New Foundation of Decision Analysis*. Kluwer, Dordrecht.

DEPARTMENT OF STATISTICS
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213-3890
E-MAIL: brian@stat.cmu.edu

NIJMEGEN INSTITUTE FOR COGNITION
 AND INFORMATION
UNIVERSITY OF NIJMEGEN
NIJMEGEN
THE NETHERLANDS
E-MAIL: ellis@nici.kun.nl