A characterization of monotone unidimensional latent variable models *

Brian W. Junker
Department of Statistics
Carnegie Mellon University
Pittsburgh PA, USA
brian@stat.cmu.edu

Jules L. Ellis
Department of Math. Psychology
University of Nijmegen
Nijmegen, The Netherlands
u212767@hnykun11.bitnet

*Presented June 24, 1994 at the Annual Meeting of the Psychometric Society, Champaign Illinois. This work was initiated during visits of Ellis to the Department of Statistics, Carnegie Mellon University and Junker to the University of Nijmegen. These visits were made possible by a grant from the Carnegie Mellon University Faculty Development program, and the generous hospitality of the Department of Mathematical Psychology, Catholic University of Nijmegen.

The Ideas

Motivation

Typical psychometrics / latent variable / errors-in-variables setting:

For repeated measures $\underline{X} = (X_1, X_2, \ldots)$, the "standard models" assume a *monotone unidimensional representation*,

Conditional Independence (CI): $\coprod \underline{X}|\Theta$ *(i. but not i.d.!)*.

Monotonicity (M): $P[X_i > t|\theta]$ nondecreasing in θ, for all t.

Unidimensionality ($d_L = 1$): $\Theta \in \mathbb{R}$.

Examples:

- *Errors in Variables, "True Score Theory", or One-Factor Factor Analysis:*

$$X_i = \alpha_i T + \epsilon_i;\ i = 1, 2, \ldots;\ \alpha_i \geq 0$$

- *"Item Response Models", e.g. used in Educational Testing:*

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

  where $(X_1, X_2, \ldots)$ are binary and each $P_j(\theta)$ is nondecr. in θ.

Example to think about

- The *generalized beta-binomial* model:

$$P(x_1, \ldots, x_J) = \int u^{x+} (1-u)^{J-x+} \, dG(u)$$

$$= \mathbf{G}(x_+)$$

- The *Rasch* model:

$$P(x_1, \ldots, x_J) = \int \prod_{j=1}^{J} \frac{e^{x_j(\theta-\beta_j)}}{1+e^{x_j(\theta-\beta_j)}} \, dF(\theta)$$

$$= e^{-\sum_j \beta_j x_j} \int u^{x+} \prod_{j=1}^{J} (1 + u\, e^{-\beta_j})^{-1} \, dH(u)$$

$$= e^{-\sum_j \beta_j x_j} \, \mathbf{H}(x_+)$$

What's the difference?

$(X_1, X_2, \ldots)$ are **exchangeable** iff for every J and every permutation

$$i(1),\ i(2),\ \ldots,\ i(J)$$

the distribution of

$$(X_1,\ X_2,\ \ldots,\ X_J) \quad \text{and} \quad (X_{i(1)}, X_{i(2)}, \ldots, X_{i(J)})$$

are the same.

Problem

We only get to see $\underline{X} = (X_1, X_2, \ldots)$, and must guess what model, or class of models, makes sense.

Goal:

- What features does a ***monotone unidimensional representation*** imply about the joint distribution of $\underline{X}$?
- What features of the distribution of $\underline{X}$ imply a ***monotone unidimensional representation***?

Analogue:

De Finetti, exchangeability, and the tail σ-field:

- $P(X_1, \ldots, X_J) = \int u^{x_+}(1-u)^{J-x_+} dG(u)$, $\forall J$, iff $(X_1, X_2, \ldots)$ [binary] are ***exchangeable***; in this case also $\sigma(\Theta) = \tau(\underline{X})$, the ***tail*** $\boldsymbol{\sigma}$***-field***, a.s. (e.g. Galambos, 1982).
- Aldous (1981): Any exchangeable sequence may be viewed as i.i.d. disturbances of a single underlying Θ.
- Diaconis and Freedman (1984): "Exchangeable-like" extreme-point representations by specifying sufficient statistics and conditional laws of the data given the sufficient statistics.

What property of the sequence $(X_1, X_2, \ldots)$ replaces ***exchangeability*** as a criterion for CI ***(but not i.d.)*** and M?

Observable constraints

What are the *constraints* on the the repeated measures $(X_1, X_2, \ldots)$ when the *monotone unidimensional representation* holds?

Conditional Association (CA):

Holland and Rosenbaum (1986) show that the monotone unidimensional representation implies CA: for all J, all partitions $(X_1, \ldots, X_J) = (\underline{Y}, \underline{Z})$, all nondecreasing $f(\cdots)$ and $g(\cdots)$, and all $\underline{h}(\cdots)$,

$$\text{Cov}\,(f(\underline{Y}), g(\underline{Y})|\underline{h}(\underline{Z}) = \underline{c}) \geq 0$$

- If each X_i is driven monotonically by the same Θ, then $(X_1, X_2, \ldots)$ possesses so much internal coherence that all monotone summaries of $\underline{Y}$ should be correlated, in any reasonable data-defined subpopulation.

Vanishing Conditional Dependence (VCD):

Suppose the monotone unidimensional representation holds, with $\Theta \in \tau(\underline{X})$. Then by standard approximation arguments, for all partitions $(X_1, \ldots, X_J) = (\underline{Y}, \underline{Z})$, and all $f(\cdots)$ and $g(\cdots)$,

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

- If $\Theta \in \tau(\underline{X})$, then conditioning on $(X_{J+1}, \ldots, X_{J+m})$ approximates conditioning on Θ, for which independence holds.

Informal Definitions

- A *sigma field* $\sigma(X_1, \ldots, X_J)$ is the set of "all" inferences, parameters, and functions of $X_1, \ldots, X_J$ that can be known from observing $X_1, \ldots, X_J$.

- The *tail sigma field* of $\underline{X} = (X_1, X_2, \ldots)$,

$$\tau(\underline{X}) = \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, X_{n+2}, \ldots),$$

  is the set of "all" inferences, parameters and functions that can be consistently estimated from $X_1, X_2, \ldots$, *even if some (indeed, <u>any</u>!) finite set of X_j's are ignored.*

<u>*Junker (1993):*</u>

- Argues that $\Theta \in \tau(\underline{X})$ is a *requirement* for latent variable models that are "useful" for measurement.

- Proves that for binary data, the monotone unidimensional ($d_L = 1$) IRT model does satisfies the stronger condition $\sigma(\Theta) = \tau(\underline{X})$.

<u>The Theorem</u>

1. The following are equivalent:
   (a) There exists $\Theta \in \tau(\underline{X})$ such that the *monotone unidimensional representation* holds.
   (b) There exists a σ-field $\mathcal{F} \subseteq \tau(\underline{X})$ such that
      i. $\coprod \underline{X}|\mathcal{F}$; and
      ii. For all J, all nondecreasing $f(\cdots)$ and $g(\cdots)$, and all $A \in \mathcal{F}$ for which the covariance below is defined,
      $$\mathrm{Cov}\,(f(X_1,\ldots,X_J), g(X_1,\ldots,X_J)|A) \geq 0. \qquad (*)$$
   (c) CA and VCD hold for $\underline{X} = (X_1, X_2, \ldots)$.
2. When any (hence all) of the above conditions hold, then $\sigma(\Theta) = \mathcal{F} = \tau(\underline{X})$, a.s.
3. If condition (a) holds for both Θ_1 and Θ_2 then these Θ's are *comonotone*: There exists an almost-sure set C on which
$$\Theta_1(v) > \Theta_1(w) \;\Rightarrow\; \Theta_2(v) \geq \Theta_2(w).$$

<u>*What does this mean?*</u>

- This gives a de Finetti-style characterization of the *monotone unidimensional representation.*
- Part (c) [CA and VCD] gives "observable" criteria for including monotone unidimensional representations in a statistical model.

Main ideas of the proof

$(a) \Rightarrow (c)$: Essentially done on the page introducing CA and VCD!

$(c) \Rightarrow (b)$: Take $\mathcal{F} = \tau(\underline{X})$ and use standard approximation arguments.

$(b) \Rightarrow (a)$: *The hard part!* We construct a $\Theta \in \mathbb{R}$ and show that CI and M hold for this Θ...

Central ideas for $(b) \Rightarrow (a)$:

Suppose there is an $\mathcal{F}$ satisfying the conditions of part (b).

- Define a set of *true scores* $\underline{T} = (T_{iq} : i \in \mathbb{N}, q \in \mathcal{Q})$, where $T_{iq} = P[X_i > q|\mathcal{F}]$, and let
$$\tau(\underline{T}) = \cap_{n=1}^{\infty} \sigma(T_{iq} : i > n, q \in \mathcal{Q}).$$

- $\tau(\underline{T})$ provides a bridge between $\mathcal{F}$ and $\tau(\underline{X})$.

- The true scores T_{iq} are all comonotone with one another; and hence can be used to construct an appropriate Θ.

The true scores

$$T_{iq} = P[X_i > q \mid \mathcal{F}]$$

Consider

$$Y_{iq} = \begin{cases} 1, & \text{if } X_i > q; \\ 0, & \text{if } X_i \leq q. \end{cases}$$

so that

$$T_{iq} = P[X_i > q \mid \mathcal{F}] = E[Y_{iq} \mid \mathcal{F}]$$

- Y_{iq} are often used to analyze polytomous response data (the "item steps"). We use them to analyze arbitrary (even continuous) response data.

- If $\mathcal{F}$ is replaced with a latent trait Θ, then

  $$T_{iq} = P[X_i > q \mid \theta] = E[Y_{iq} \mid \theta]$$

  are the "ISRF's" in a nonparametric polytomous analysis.

- *Everything* that can be known about the "latent behavior" of the X_i's (conditional on $\mathcal{F}$) *can be learned from the true scores* T_{iq}.

The Proof

<u>Lemma 1</u>

Suppose $\mathcal{F} \subseteq \tau(\underline{X})$ and $\amalg \underline{X}|\mathcal{F}$. Then

$$\mathcal{F} = \sigma(\underline{T}) = \tau(\underline{T}) = \tau(\underline{X}), a.s.$$

<u>*Interpretation*</u>

- If $\mathcal{F}$ is rich enough to induce CI, then $\mathcal{F} = \tau(\underline{X})$.
- The true scores T_{iq} are quite redundant, in the sense that $\sigma(\underline{T}) = \tau(\underline{T})$.

<u>Lemma 2</u>

Suppose $\mathcal{F} \subseteq \tau(\underline{X})$; $\amalg \underline{X}|\mathcal{F}$; and the covariance condition $(*)$ of part (b) of the theorem holds. Then each pair (T_{iq}, T_{jr}) is comonotone.

<u>*Interpretation*</u>

- Condition $(*)$, related to CA, severely constrains the variation of the T_{iq}'s.
- Since every pair (T_{iq}, T_{jr}) is comonotone, we go hunting for a common variable Θ such that $T_{iq} = f_{iq}(\Theta)$. Our quarry is

$$\Theta = \sum_{i=1}^{\infty} \sum_{q \in \mathcal{Q}} a_{iq} T_{iq}$$

for any summable sequence $a_{iq} > 0$ (and a fixed ordering of $\mathcal{Q}$).

Lemma 1

Suppose $\mathcal{F} \subseteq \tau(\underline{X})$ and $\coprod \underline{X} | \mathcal{F}$. Then

$$\mathcal{F} = \sigma(\underline{T}) = \tau(\underline{T}) = \tau(\underline{X}), a.s.$$

Interpretation

- If $\mathcal{F}$ is rich enough to induce CI, then $\mathcal{F} = \tau(\underline{X})$.

- The true scores T_{iq} are quite redundant, in the sense that $\sigma(\underline{T}) = \tau(\underline{T})$.

Proof

For the first "=", "$\supseteq$" follows from the definition of T_{iq}, and "$\subseteq$" is a kind of 0/1 law that can be built up from cylinder sets in $\sigma(\underline{X})$.

The second "=" follows, since the first is still true if you drop a few X's [because $\mathcal{F} \subseteq \tau(\underline{X})$].

For the third "=", "$\subseteq$" is again from the definition of T_{iq}, and "$\supseteq$" is the elementary 0/1 law for independent r.v.'s, given $\mathcal{F}$. $\square$

Lemma 2

Suppose $\mathcal{F} \subseteq \tau(\underline{X})$; $\amalg\, \underline{X}|\mathcal{F}$; and the covariance condition $(*)$ of part (b) of the theorem holds. Then each pair (T_{iq}, T_{jr}) is comonotone.

Interpretation

- Condition $(*)$, related to CA, severely constrains the variation of the T_{iq}'s.

- Since every pair (T_{iq}, T_{jr}) is comonotone, this suggests that we look for a common variable Θ such that $T_{iq} = f_{iq}(\Theta)$.

Proof

$i \neq j$: By conditioning on $\mathcal{F}$ it follows that $\mathrm{Cov}\,(T_{iq}, T_{jr}|A) \geq 0$ for all $A \in \mathcal{F}$ for which the covariance is defined. Consider the almost-sure set $C = \{(s,t) : P[(T_{iq}, T_{jr}) \in B_\delta(s,t)] > 0, \forall \delta > 0\}$: by geometric arguments one can show that C cannot contain (s_1, t_1), (s_2, t_2) with $s_1 < s_2$ and $t_1 > t_2$.

$i = j$: By **Lemma 1**, $\exists\, k > i$ such that $T_{iq}, T_{ir} \in \sigma(T_{js} : j > k, s \in \mathcal{Q})$. Then $T_{iq}(v) > T_{ir}(w) \Rightarrow \exists\, T_{js}(v) \neq T_{js}(w) \Rightarrow T_{js}(v) > T_{js}(w)$ $\Rightarrow T_{ir}(v) \geq T_{ir}(w)$ [by the case $i \neq j$]. $\Box$

Now construct Θ

For any particular ordering of the rationals q, let $a_{iq} > 0$ such that $\Sigma_{i=1}^{\infty} \Sigma_q a_{iq} < \infty$, and define

$$\Theta = \sum_{i=1}^{\infty} \sum_q a_{iq} T_{iq}.$$

Let C be a common almost-sure set on which all pairs (T_{iq}, T_{jr}) are comonotone (available by countable applications of **Lemma 2**).

Loose ends:

(1) On C,

$$\begin{aligned} \Theta(v) > \Theta(w) &\Rightarrow \exists\, T_{iq}(v) > T_{iq}(w) \\ &\Rightarrow \forall j, r,\ T_{iq}(v) \geq T_{iq}(w). \end{aligned}$$

It follows that *each T_{iq} is a monotone function of Θ, a.s.*

(2) By (1), $T_{iq} \in \sigma(\Theta)$ a.s., so from **Lemma 1**,

$$\mathcal{F} = \sigma(\underline{T}) = \tau(\underline{T}) = \tau(\underline{X}) \subseteq \sigma(\Theta) \subseteq \mathcal{F},\ a.s.$$

Hence

(M): $T_{iq} = P[X_i > q|\mathcal{F}] = P[X_i > q|\Theta]$ is nondecreasing in Θ;
(CI): $\coprod \underline{X}|\mathcal{F} \Rightarrow \coprod \underline{X}|\Theta$.

(3) If Θ_1 and Θ_2 are two random variables satisfying part (a) of the theorem, we can use **Lemma 1** with $\mathcal{F} = \sigma(\Theta_i)$, for each $i = 1, 2$, to show that $\sigma(\Theta_1) = \sigma(\Theta_2)$, a.s., and hence the true scores T_{iq} are equal. Now using the monotonicity assumption in (a), all pairs (Θ_1, T_{iq}) and (Θ_2, T_{iq}) must be comonotone; hence

$$\begin{aligned} \Theta_1(v) > \Theta_1(w) &\Rightarrow \exists\, T_{iq}(v) > T_{iq}(w) \\ &\Rightarrow \Theta_2(v) \geq \Theta_2(w). \end{aligned}$$

The Examples

Example: Linear Factor Analysis

d = 1: $X_j = a_j T + \epsilon_j$ $\qquad$ (T, $\epsilon_j \sim$ i.i.d. $N(0,1)$; $a_j \geq 0$)

$$\not\Sigma_{(X_1,\ldots,X_J)\cdot(X_{J+1},\ldots,X_{J+m})} = I_{J\times J} + \frac{1}{1+\sum_{j=J+1}^{J+m} a_j^2} \begin{bmatrix} a_1^2 & a_1a_2 & \cdots & a_1a_J \\ a_2a_1 & a_2^2 & \cdots & a_2a_J \\ \vdots & \vdots & \ddots & \vdots \\ a_Ja_1 & a_Ja_2 & \cdots & a_J^2 \end{bmatrix}.$$

- CA $\Rightarrow$ Cov$\,(X_i, X_j|$any subset of X_k's$) \geq 0$
- Positive partial correlations $\equiv$ MTP_2.
  - MTP_2 $\Rightarrow$ association .... *but not conversely.*
  - For normals, CA $\Rightarrow$ MTP_2 .... *Converse?*
- VCD $\Rightarrow$ Cov$\,(X_i, X_j|X_{J+1}, \ldots, X_{J+m}) \to 0$; $\Rightarrow \Sigma_{j=1}^J a_j^2 \to \infty$. Then T can be estimated by $\Sigma_{j=1}^J a_j X_j / \Sigma_{j=1}^J a_j^2$, which ensures $T \in \tau(\underline{X})$.

d = 2: $X_j = a_j T_1 + b_j T_2 + \epsilon_j$ (T_i, $\epsilon_j \sim$ i.i.d. $N(0,1)$; a_j, $b_j \geq 0$)

- If $a_1 = 1$, $a_2 = 0$, $b_1 = 0$, $b_2 = 1$, and $a_j = b_j = 1$ for all $j > 2$, then Cov$\,(X_1, X_2|X_3 + \ldots + X_J = c) \leq 0$, violating CA.
- While $\exists$ U such that $\coprod \underline{X}|U$ (Suppes and Zanotti, 1981), $\nexists$ Θ such that CI and M hold, *even if we use a non-linear-regression likelihood and a nonnormal* Θ *distribution!*

Example: The Rasch Model

For $x_j = 0$ or 1, and logit $P_j(\theta) = \theta + \beta_j$,

$$P(x_1, \ldots, x_J) = \int \prod_{j=1}^{J} P_j(\theta)^{x_j} [1 - P_j(\theta)]^{1-x_j} \, dF(\theta).$$

• Log-linear representation:

$$\log P(x_1, \ldots, x_J) = \alpha + \sum_{j=1}^{J} \beta_j x_j + \sum_{k=0}^{J} \gamma_k 1_{\{x_+ = k\}},$$

CA $\Rightarrow$ conditions on the γ_k's (worked out, from a different p.o.v., by Cressie and Holland, 1983).

• VCD $\Rightarrow$

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

In the Rasch (or any IRT) model this is easy to see with a Bayesian argument:

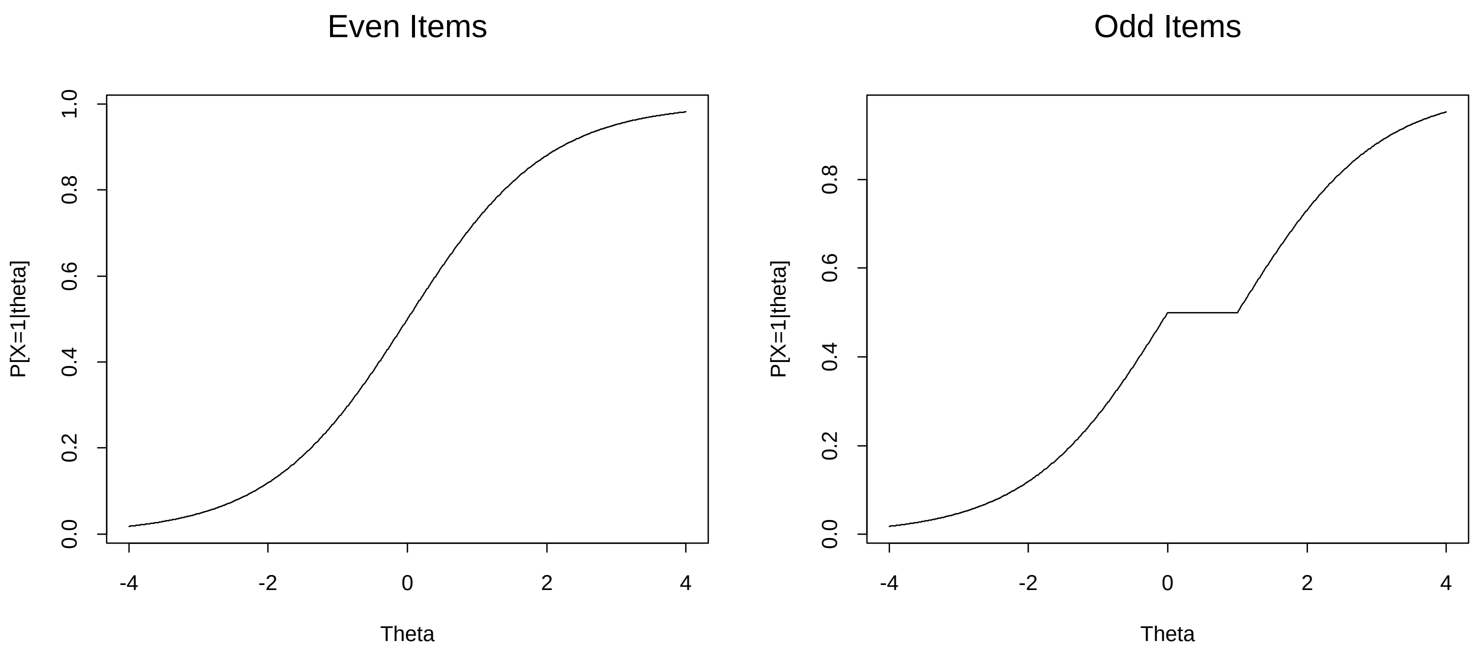
$$p(x_1, \ldots, x_J | x_{J+1}, \ldots, x_{J+m}) = \int \prod_{j=1}^{J} P_j(\theta)^{x_j} [1 - P_j(\theta)]^{1-x_j} \, dF(\theta | x_{J+1}, \ldots, x_{J+m})$$

and $dF(\theta | x_{J+1}, \ldots, x_{J+m}) \longrightarrow 1_{\{\theta \ = \ \text{examinee's } \theta\}}$.

• Chang and Stout (1993 ***Psy***): this works if the β_j's are bounded on average.

An example where uniqueness "fails"

Consider a $d_L = 1$ representation for $(X_1, X_2, \ldots)$ with ICC's $P_j(\theta)$:



- Whole sequence: Unidimensional with trait Θ.
- Even items: Unidimensional with the same Θ.
- Odd items: Unidimensional with trait Θ'; "compresses out" [0,1].
- $\sigma(\Theta') \in \sigma(\Theta) = \tau(X_1, X_2, X_3, \ldots)$, but $\sigma(\Theta) \notin \sigma(\Theta') = \tau(X_1, X_3, X_5, \ldots)$.
- Possible fix: Require $\Theta \in$ "subtail", the set of all inferences, functions, etc., estimatable from *every* infinite subsequence of X_j's.

(A very strong form of "specific objectivity")

Summary and Conclusions

Interpretations

Summary

- "CA and VCD hold, if and only if there is a $d_L = 1$ (monotone, unidimensional, LI) representation for $\underline{X} = (X_1, X_2, \ldots)$, using a Θ equivalent to $\tau(\underline{X})$."
- Under these conditions, Θ is uniquely determined by the items (ordinal uniqueness). But if you change the items you can change Θ.
- Other $d_L = 1$ representations are possible; CA will hold and VCD will fail. In these representations, Θ *will contain more information than* $\tau(\underline{X})$: $\tau(\underline{X}) \subseteq \sigma(\Theta)$.

A Geometric interpretation; dimensionality of tests

- Think of the true scores T_{iq} as "filling out" a manifold in some possibly infinite dimensional space.
- The dimensionality of the manifold is the dimensionality of the test.

  Similar ideas: Ramsay's (1993) TESTDIST program/analysis.
- CA and VCD together hold, iff the manifold is in fact 1-dimensional, i.e., iff there is a θ such that $T_{iq} \equiv T_{iq}(\theta)$.

Generalizations

- CA: Implied by LI, M, $d = 1$.

- VCD says that

  $$p(x_1, \ldots, x_J \mid x_{J+1}, \ldots, x_{J+m}) \approx \prod_1^J p_j(x_j;\ x_{J+1}, \ldots, x_{J+m}).$$

  VCD should work whenever LI, $d = k$ hold, and the k-dimensional $\theta \in \tau(\underline{X})$. *[See Rasch example.]*

- In a LI, M, $d = k$ VCD will still hold. But CA will fail. *What replaces CA when $d > 1$?*

- Difficulty of nonparametric definitions of $d > 1$.

Connections with other work

- *Holland and Rosenbaum (1986, Ann Stat).*
  - Connections with literature on probability inequalities.
  - CA as a very strong necessary condition for $d_L = 1$.
- *Stout (1990, Psy).*
  - Embed $(X_1, \ldots, X_J)$ into an *infinite item sequence* to fully understand latent structure.
  - Ordinal uniqueness.
- *Junker (1993, Ann Stat).*
  - Partial characterization using $d_E = 1$ and CA.
  - Natural to think about $\tau(\underline{X})$, the consistently estimatable quantities.
- *Ellis and van den Wollenberg (1993, Psy).*
  - Complete characterization of a much stronger model (monotone/local homogeneity; see also Meredith and Millsap (1992)).
  - Connections with the notion of co-monotonicity.
- *Ramsay (1993, TESTDIST).*
  - Nonparametric estimation of the "test manifold" using similarity measures between items.
  - Does not depend on fixed-dimensional latent variable representation.

The References

References

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

Diaconis, P. and Freedman, D. (1984). Partial exchangeability and sufficiency. *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on "Statistics: Applications and New Directions,"* Calcutta, 16 December–19 December, 1981, pp. 205–236.

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.

Galambos, J. (1982). Exchangeability. *Encyclopedia of Statistical Sciences, 2,* 573–577.

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

Junker, B. W. (1993). Conditional association, essential independence and monotone unidimensional item response models. *Annals of Statistics, 21,* 1359–1378.