Examining monotonicity in polytomous item response data

Brian W. Junker
Department of Statistics
232 Baker Hall
Carnegie Mellon University
Pittsburgh PA 15213
(412) 268-8873
brian@stat.cmu.edu

Paper presented at the Annual North American Meeting of the Psychometric Society, June 27–June 30, 1996, Banff, Alberta, CANADA. Research supported by the National Science Foundation, Grant DMS-94.04438.

1 Introduction: EDA for Monotonicity in IRT

- Often natural or desirable to assume that the responses are ordered

- Not always obvious what the correct order of responses is.

- Consider three scales:

  1. *transitive reasoning*

     (scored 0/1/2: wrong/judgement-only/judgement-plus-reasoning)

  2. *personality self-report*

     (5-point Likert items)

  3. *working memory (over-)load*

     (partially-ordered categories)

- The general questions motivating this study can be phrased as follows:

  1. can we confirm, nonparametrically, that the order of item responses in the *transitive reasoning* scale is compatible with a monotone unidimensional IRT model;

  2. can we select, nonparametrically, an order of item responses in the *personality* and *working memory* tests that make it possible to model the data with a monotone unidimensional IRT model?

- These are essentially questions about *exploratory data analysis* for montonocity in polytomous item response data.

2 Background: IRT and Monotonicity Assumptions

- X_1, X_2, ..., X_J: polytomous items with m_1, m_2, ..., m_J categories;

$$P_{js}(\theta) = P[X_j = s|\theta]$$

- Usual $d = 1$ nonparametric IRT model:

$$P[X_1 = x_1, \ldots, X_J = x_J] = \int P[X_1 = x_1, \ldots, X_J = x_J|\theta]\, dF(\theta)$$

assuming

Unidimensionality ($d = 1$): θ takes values on the real line;

Local Independence (LI):

$$P[X_1 = x_1, \ldots, X_J = x_J|\theta] = \prod_{j=1}^{J} P_{jx_j}(\theta)$$

and either

Monotonicity (M): The item step response function (ISRF)

$$P^*_{js}(\theta) = P[X_j > s|\theta] = \sum_{x=(s+1)}^{m_j} P_{jx}(\theta)$$

is nondecreasing in θ for all s.

as in Samejima's (1969) ***graded response model (GRM)***, or

Monotone Item Likelihood Ratio (MILR):

$$\frac{P_{j(s+1)}(\theta)}{P_{js}(\theta)} \text{ is nondecreasing in } \theta \text{ for all } s.$$

which follows from nondecreasingness of

$$P[X_j = s+1|\theta, \text{ and } X_j \in \{s, s+1\}] = \frac{P_{j(s+1)}(\theta)}{P_{js}(\theta) + P_{j(s+1)}(\theta)}$$

as in Masters' (1982) ***partial credit model (PCM)***.

- Hemker, Sijtsma, Molenaar and Junker (1995ab) call models satisfying M $\boxed{\textit{np-GRM's}}$
- Hemker et al. (1995ab) call models satisfying MILR $\boxed{\textit{np-PCM's}}$.

- np-GRM's have been studied under many names, mostly in efforts at characterization theorems.
  - Mokken (1971) and others: monotone homogeneity;
  - Holland and Rosenbaum (1986): monotone latent trait model;
  - Junker (1991, 1993, 1994): strictly unidimensional model;
  - Ellis and Junker (1994), Junker and Ellis (1995): monotone unidimensional latent variable representation.

- np-PCM's are new. See Hemker et al. (1995ab):
  - For each item, MILR implies M
  - Taxonomy of nonparametric polytomous IRT models similar to Thissen and Steinberg (1986).
  - Results like Grayson's (1988) MLR property for the total score do not generally extend to nonparametric polytomous models.

- Another set of nonparametric models, the ***doubly monotone*** models (Mokken, 1971; Sijtsma and Junker, 1996; Scheiblechner, 1995) will not be pursued here.

3 Theory and Methods

3.1 Stochastic ordering and monotone likelihood ratio properties

- Latent M and manifest M: Suppose the np-GRM holds...
  - Junker (1993): For binary items $P[X_j = 1|S = s]$ is monotone if S is the *rest score* but not if S is the *total score*
  - Hemker et al (1996ab): For polytomous items $P[X_j \geq r|S = s]$ is not monotone if S is total score *or rest score*
  - Junker (1991), Scheiblechner (1995): If polytomous np-GRM items are dichotomized *anywhere* the resulting binary items will be monotone unidimensional.

...So we can check for M by checking monotonicity of

$$P[X_j \geq r|S = s]$$

where S = rest-score *after dichotomizing all items.*

- Latent MILR and manifest MLR: Suppose the np-PCM holds...

  - Following Junker (1993, Lemma 4.1):

$$\frac{P[X_i = r+1|X_j = s]}{P[X_i = r|X_j = s]} \le \frac{P[X_i = r+1|X_j = s+1]}{P[X_i = r|X_j = s+1]}$$

or equivalently the table

$$\begin{array}{cc|cc} & & \multicolumn{2}{c}{X_j} \\ & & s & s+1 \\ \hline X_i & r & n_{rs} & n_{r(s+1)} \\ & r+1 & n_{(r+1)s} & n_{(r+1)(s+1)} \end{array}$$

is positively associated, for all r and s.

...Can check for positive association using:

- odds ratios or log odds ratios;
- determinants;
- signed chi-squared; etc.

and try to swap rows or columns to fix negative associations.

3.2 Nonparametric regression

- Ramsay (1991) and Abrahamowicz and Ramsay (1992).
  - Estimate $\hat{\theta} = g(S)$ where $g(\cdot)$ is chosen to force $\hat{\theta} \sim N(0,1)$. [Ordinal consistency: Stout, 1990; Junker, 1991; Douglas, 1996].
  - $P^*_{js}(\hat{\theta}) = P[X_j \geq s|\hat{\theta}]$ is estimated by kernel-based or spline-based logistic regression.
- We used S-PLUS's `gam` function to fit cubic B-splines as nonparametric regression estimators of $\log[P^*_{js}(\hat{\theta})/(1 - P^*_{js}(\hat{\theta}))]$ as functions of $\hat{\theta} = g(S)$ where S is the dichotomized rest score.
- Can be used to visually examine M for example.

3.3 Mokken's scaling coeffients

- E.g.: Mokken (1971); Mokken and Lewis (1982); Molenaar (1991); Sijtsma and Verweij (1992).

- $H_{ij} = \dfrac{\mathrm{Cov}\,(X_i, X_j)}{\mathrm{Cov}_{\,max}(X_i, X_j)}$

  Maximizing H_{ij} maximizes linear relationship between X_i and X_j with individual item response category frequencies fixed.

- $H_i = \dfrac{\sum_{j \neq i} \mathrm{Cov}\,(X_i, X_j)}{\sum_{j \neq i} \mathrm{Cov}_{\,max}(X_i, X_j)}$

  H_i is a scaled item-total covariance, used as a kind of item misfit statistic.

- $H = \dfrac{\sum\sum_{i \neq j} \mathrm{Cov}\,(X_i, X_j)}{\sum\sum_{i \neq j} \mathrm{Cov}_{\,max}(X_i, X_j)}$

  When H is large ($H > 0.3$ is a typical criterion), the items are thought to fit the np-GRM well. Maximizing H maximizes total score variance with individual item response category frequencies fixed.

3.4 Guttman scaling

- ***Goal:*** assign category scores to all items $X_1, \ldots, X_J$ to maximize the total-score variance (Basilevsky, 1994).

- The Guttman scores are the first nontrivial principal component $\underline{u}$ of $\underline{D}^{-1/2}\underline{Y}^T\underline{Y}\,\underline{D}^{-1/2}$, where

  * $\underline{Y}$ = raw data matrix with entries 0 or 1 for whether person i responded in category s to item j.
  * Usually, $\underline{D} = \text{diag}(\underline{Y}^T\underline{Y})$.

- Note that

$$\underline{Y}^T\underline{Y} = \begin{bmatrix} \underline{N}_{11} & \underline{N}_{12} & \cdots & \underline{N}_{1J} \\ \underline{N}_{12}^T & \underline{N}_{21} & \cdots & \underline{N}_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \underline{N}_{1J}^T & \underline{N}_{2J}^T & \cdots & \underline{N}_{JJ} \end{bmatrix}$$

  where $\underline{N}_{ij}$ is the two-way table of joint category response frequencies n_{rs} for $X_i = r$ and $X_j = s$.

- We do not want the Guttman scores directly, but rather sort the item categories according to the Guttman scores.

- Similar use of optimal scores in archaeological seriation problems:
  - Now $\underline{N}_{ij}$ is replaced by an adjacency matrix $\underline{A}_{ij}$ with entries $a_{rs} = 1$ or 0 according to whether artifact r was found at site s.
  - Sorting on optimal scores tends to bring 1's to main diagonal of $\underline{A}_{ij}$ and 0's away from main diagonal of $\underline{A}$ (maximizing correlation).
  - Seriation by optimal scoring cannot deal with more than two variables at a time, and we must deal with J variables at a time.

- An analogue to seriation that is helpful for our EDA problem:

  Modify $\underline{Y}^T\underline{Y}$ as follows:

  – Replace each $\underline{N}_{ij}$ with the ***residual adjacency matrix*** $\underline{A}_{ij}$:

$$a_{rs} = \begin{cases} 1, & \text{if the standardized residual } (n_{rs} - n_{r\cdot}n_{\cdot s}/n)/\sqrt{n_{r\cdot}n_{\cdot s}/n} \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

    * tends to select configurations with positive, perhaps not linear, association

  – Replace each $\underline{N}_{ii}$ with all zeros.

    * tends to force the Guttman method to select category scores that maximize between-item covariance rather than within-item variance.

  – Call the resulting matrix $\underline{A}$ instead of $\underline{Y}^T\underline{Y}$.

- Both Guttman analysis of $\underline{Y}^T\underline{Y}$ and Guttman analysis of $\underline{A}$ will

  be considered.

3.5 Sorting to achieve MLR

- Consider a two-way table of counts cross-classifying responses from a three-level item X_1 with those of a three-level item X_2:

		X2		
		0	1	2
	0	11	0	3
X1	1	238	11	57
	2	83	1	21

- As an illustration, the matrix of 2×2 associations for the table above, using determinants to measure association, is

	0,1	0,2	1,2
0,1	121	-87	-33
0,2	11	-18	-3
1,2	-675	267	174

- If columns 0 and 1 of the original two-way table were swapped, the resulting matrix of associations would be

	1,0	1,2	0,2
0,1	-121	-33	-87
0,2	-11	-3	-18
1,2	675	174	267

- *Goal:* try to swap rows and columns so that all entries in the table of 2×2 associations are nonnegative (a bit like Rubik's cube!)

- Since only the signs of the measures of 2×2 association matter, can avoid ***capitalization on chance*** by using

$$\delta_c = \begin{cases} 1, & \text{if } \xi > c; \\ 0, & \text{if } -c \leq \xi \leq c; \\ -1, & \text{if } \xi < -c, \end{cases}$$

  as the measure of 2×2 association (ξ = signed chi-squared).

- For example, $c = 3.84$ corresponds to an approximate level 0.05 test of association in each 2×2 table. The larger c is, the more conservative (less re-labelling) we will be.

- We have found in preliminary work that $c = 1.00$ works somewhat better than $c = 3.84$ in recovering the correct labels from randomly mislabelled simulation data.

3.6 RC(1) Models and 2×2 Odds Ratios

- RC(1) model:

$$\log p_{ij} = \alpha + \beta_i + \gamma_j + \mu_i \nu_j$$

  Log-bilinear!

- Goodman's MLE Hack: Alternatively fit with μ free and ν fixed, and vice-versa.

- For 2×2 subtable with rows i, i' and columns j, j',

$$\log(\textit{odds ratio})_{ii'jj'} = (\mu_i - \mu_{i'})(\nu_j - \nu_{j'})$$

  so μ's and ν's become category scores that can be sorted analogously to Guttman scaling.

- Useful references: Goodman (1979), Becker and Clogg (1989), Agresti (1990), *Anderson and Hsieh (1996, AERA).*

4 Exploring Monotonicity in the Example Data Sets

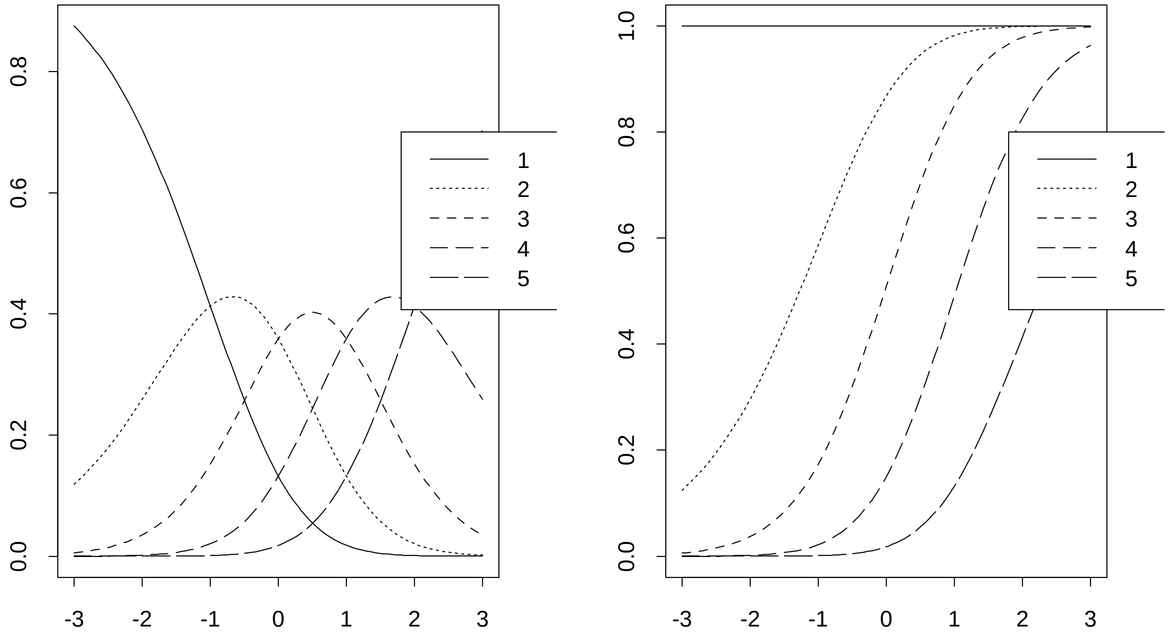

Fig 1: ICRF's (left plot) and ISRF's (right plot) of a Masters' partial credit item with five response categories.

- Ideal polytomous items look something like this
  - M: ISRF's nondecreasing
  - MILR: difficult to see in these plots
  - By definition: ISRF's never overlap, lowest category always has ISRF $\equiv 1$.

4.1 M (np-GRM) and transitive reasoning

- $J = 9$ items from a 12-item scale (7, 9, 12 omitted), $n = 425$ pupils.
- Measure the ability to reason about transitive relations with respect to length, weight and size.
- $$X_j = \begin{array}{l} 0 \text{ if judgement of transitive relation incorrect;} \\ 1 \text{ if judgement of transitive relation correct;} \\ 2 \text{ if judgement correct, and pupil gives a correct ver-} \\ \text{bal justification for the judgement.} \end{array}$$
- Reason to believe np-GRM and np-PCM already hold, up to random error.
- Sijtsma and Verweij (1992); Sijtsma and Junker (1996).

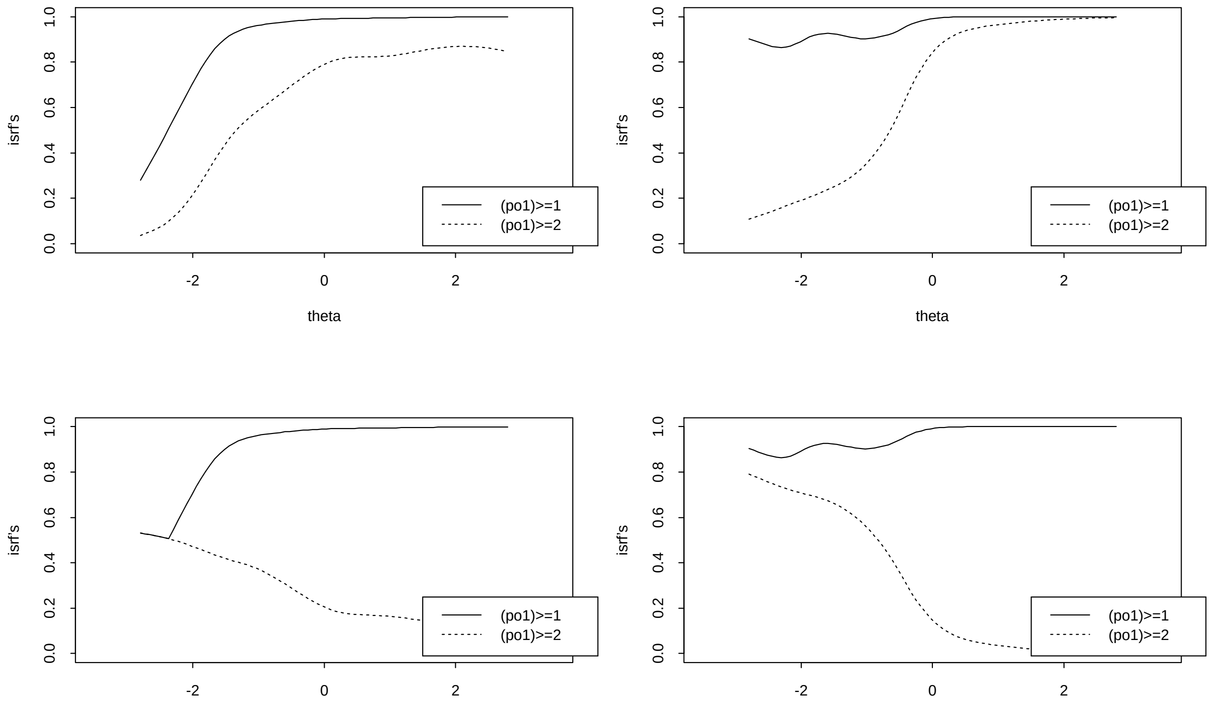

Fig 2: Effect of mislabelling categories on item 1 of the Transitive Reasoning scale (nonparametric regression ISRF's).

- In figure 2:
  - *Top row:* correct categories 0,1,2, using rest score with items dichotomized at level 1 (left) and rest score with items dichotomized at level 2 (right).
  - *Bottom row:* categories 1 and 2 swapped, using rest score with items dichotomized at level 1 (left) and rest score with items dichotomized at level 2 (right).

- Mislabelling: nonmonotonicity and/or reduced discrimination.

- For the original Transitive Reasoning data, $H = 0.4094$.

- Re-sorting categories using Guttman scaling with $\underline{Y}^T\underline{Y}$ produced no relabelling of categories; $H = 0.4094$.

- Re-sorting categories using Guttman scaling with $\underline{A}$: items 1, 2 and 3 have categories 0 and 1 reversed, and item 10 has categories 1 and 2 reversed; $H = 0.3754$.

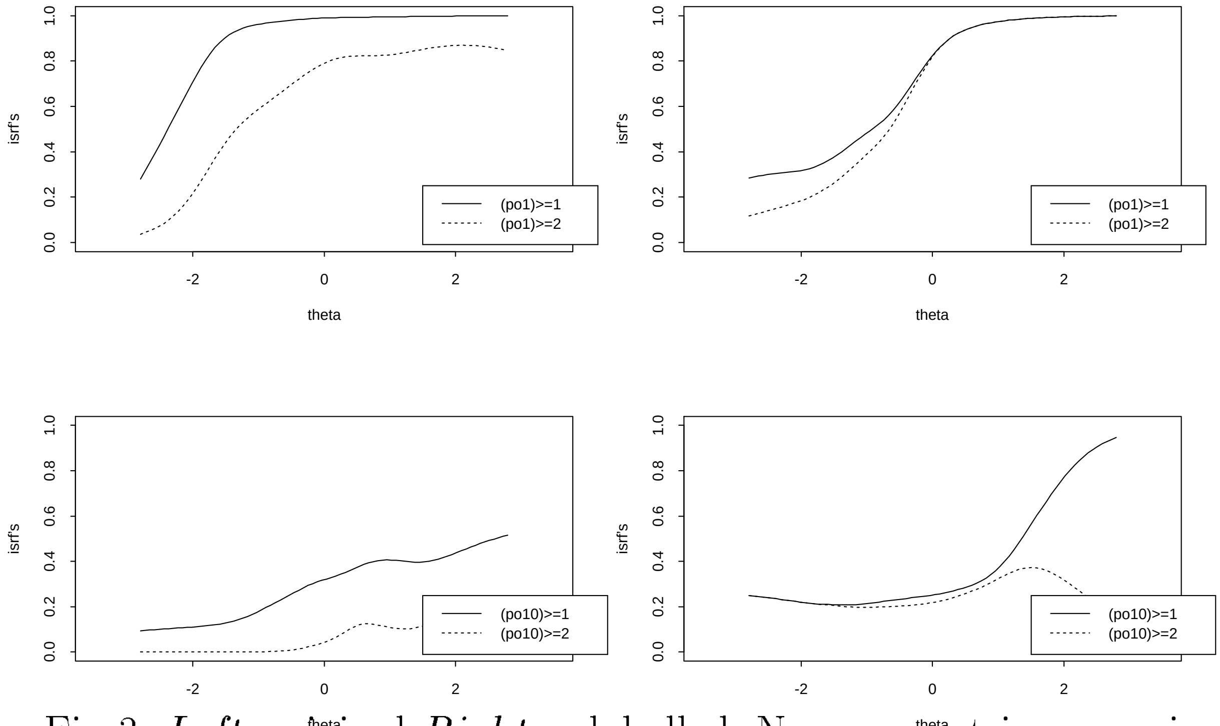

Fig 3: ***Left:*** original ***Right:*** relabelled. Nonparametric-regression ISRF's; rest scores of items dichotomized at level 2.

4.2 M (np-GRM) and personality self-report

- The Inventory of Interpersonal Problems (IIP)
- $J = 127$ 5-point (0–4) Likert items; $n = 97$ subjects.
- Horowitz (designer) identifies six subscales:

  Hard to be Assertive 21 items
  Hard to be Sociable 18 items
  Hard to be Submissive 10 items
  Hard to be Intimate 12 items
  Too Responsible 12 items
  Too Controlling *10 items*

- Do succesive points on the Likert scale indicate succesively more evidence of various personality disorders?
- With $n = 97$ and $J = 127$ items, cannot analyze all items at once.
- We concentrate on "Too Controlling"...

- The "too controlling" subscale consists of the items:

4. It is hard to keep things private from other people. 0 1 2 3 4
79. I fight with other people too much. 0 1 2 3 4
90. *I am too agressive toward other people.* 0 1 2 3 4
94. I clown around too much. 0 1 2 3 4
96. I criticize other people too much. 0 1 2 3 4
102. *I try to change other people too much.* 0 1 2 3 4
104. I am overly generous to other people. 0 1 2 3 4
109. I compliment other people too much. 0 1 2 3 4
111. I manipulate other people too much to get what I want. 0 1 2 3 4
113. I tell personal things to others too much. 0 1 2 3 4

- Calculations for the "too controlling" subscale, especially items 90 *(agressive)* and 102 *(change others)*:
  - Original items: $H = 0.34$
  - After re-sorting with Guttman ($\underline{Y}^T\underline{Y}$): $H = 0.36$
  - After re-sorting with Guttman ($\underline{A}$): $H = 0.31$.

item:	item90:	item102:
Orig. order	0 1 2 3 4	0 1 2 3 4
Guttman (YTY)	0 2 3 1 4	0 1 3 2 4
Guttman (A)	4 0 1 3 2	4 3 0 1 2

Fig 6: Effects of sorting original items (left) using Guttman method with YTY (center) and A (right).

- Over the whole “too controlling” subscale, both Guttman sorting methods improved M.
- Guttman sorting using $\underline{Y}^T\underline{Y}$ tends to make smaller changes than sorting using the residual covariance matrix $\underline{A}$.
- If we are happy with the re-ordered item categories, we would probably want to allow subjects to continue to mark the items using the present Likert labels, and then re-score the items afterwards using the re-labellings discovered here.
- Happiness depends on subject-matter experts!
- Maybe more useful to re-write the items so that the Likert category labels do lead to monotone ISRF’s.

4.3 M (np-GRM) and working memory

- $n = 87$ undergraduates performed $J = 27$ tasks with a tree-structured audio menu for course registration and campus information, with varying working memory load.

- $X_j = $ 5 if task completed efficiently with no reminders;
  4 if task completed inefficiently;
  3 if task completion required a reminder;
  2 if both 3 and 4 occurred;
  1 if task never completed.

- Some response categories (1, 2, 5) have apparently clear order relations and others (3, 4) do not.

- Patz, Junker, Lerch and Huguenard (1996).

- All 27 items were analyzed like the groups of 10 IIP items.
  - Original items: $H = 0.1994$
  - Re-sorting using Guttman ($\underline{Y}^T\underline{Y}$): $H = 0.2319$
  - Re-sorting using Guttman ($\underline{A}$): $H = 0.1937$
- Categories 1 and 2 tend to remain "low" after re-ordering (by either method) and category 5 tends to remain "high". Moreover neither method of ordering categories tends to reverse the order of categories as much as seemed to be the case in the IIP data.
- Neither method helped much with M; tends to confirm that M is restrictive, even under re-labelling.

4.4 MILR (np-PCM) and Transitive Reasoning

- An algorithm to sort categories in pairs of 3-category items, to achieve MILR:

  Repeatedly sweep through the matrix of associations, and swap rows or columns in the original two-way table if either of the following conditions holds:

  Fully reversed labels If any row or column of the matrix of 2×2 associations has all-negative associations, swap the corresponding pair of rows or columns of the two-way table;

  Look-ahead If any swap of adjacent rows or columns of the two-way table leads to fully-reversed labels on the next sweep, perform that swap.

- Illustration: Consider items X_1 and X_2 from the TR scale

```
             X2
         0    1  2
      0  3   11  0
X1    1 21   83  1
      2 57  238 11
```

- We apply the algorithm after randomly permuting the labels:

```
Two-way table:
    1  2  0
0  11  0  3
2 238 11 57
1  83  1 21

2x2 associations:
     1,2  1,0  2,0
0,2  121  -87  -33
0,1   11  -18   -3
2,1 -675  267  174

Swapping labels 1 and 2 of X1...
```

```
Two-way table:
    1  2  0
0  11  0  3
1  83  1 21
2 238 11 57

2x2 associations:
    1,2  1,0  2,0
0,1  11  -18   -3
0,2 121  -87  -33
1,2 675 -267 -174

Swapping labels 0 and 1 of X2...

Two-way table:
   0  2   1
0  3  0  11
1 21  1  83
2 57 11 238

2x2 associations:
    0,2 0,1  2,1
0,2   3  18  -11
0,1  33  87 -121
2,1 174 267 -675

Swapping labels 0 and 2 of X2...
```

```
Two-way table:
   0   1  2
0  3  11  0
1 21  83  1
2 57 238 11

2x2 associations:
    0,1 0,2 1,2
0,1  18   3  11
0,2  87  33 121
1,2 267 174 675

All sorted!
```

- Algorithm halts if
  - Success;
  - No more swaps possible;
  - A maximum number of swaps have taken place.

- Algorithm can be applied with raw association measures (log-odds or determinants) or with the decision function δ_c to reduce capitalization on chance.

Summary of results for the original Transitive Reasoning Data:

- *Using 2×2 determinants:*
  - 22 violations of MLR in the ${}_9C_2 = 36$ two-way tables.
  - All categories preserved by sorting, except:
    - item 2 in the "item 2 × item 5" table (swap 0 and 1)
    - item 2 in the "item 2 × item 8" table (swap 0 and 1)
- *Using δ_c with $c = 1.00$*
  - 8 detectable violations of MLR in the 36 two-way tables.
  - All categories preserved by sorting except:
    - item 2 in the "item 2 × item 10" table (swap 0 and 1)
    - item 11 in the "item 4 × item 11" table (swap 0 and 1)
- *Using δ_c with $c = 3.84$:*
  - Only 2 detectable violations.
  - All categories preserved by sorting except:
    - item 11 in the "item 4 × item 11" table (swap 0 and 1)
- These results are generally consistent with the results of Guttman sorting; M and the np-GRM are supported.
- Also suggest that a np-PCM might be appropriate for the transitive reasoning data.

Comparing two variations of MILR sorting with RC(1) sorting:

- Simulated PCM data ($n = 400$, $J = 10$, $m = 3$)

- *Yes* means original orders were recovered

- *Reversed* means original orders recovered in reverse order (both variables)

- *No* means original orders not recovered

Item pair	Recovered using Determinants?	Recovered using δ_c, $c = 1$?	Recovered using RC(1) model?	Remarks
1,2	Reversed	Yes	Yes	
1,3	Yes	Yes	Yes	
1,4	Yes	Yes	Yes	
1,5	Yes	Yes	Yes	
1,6	Yes	Yes	Yes	
1,7	Yes	Yes	Yes	
1,8	Yes	Yes	Yes	Table insoluble
1,9	No	Yes	Yes	
1,10	Yes	Yes	Yes	
2,3	No	Reversed	Reversed	Table insoluble
2,4	Reversed	Reversed	Reversed	
2,5	No	No	Reversed	Table insoluble
2,6	Reversed	Reversed	Reversed	
2,7	Reversed	Reversed	Reversed	
2,8	Reversed	Reversed	Reversed	
2,9	Reversed	Reversed	Reversed	
2,10	Reversed	Reversed	Reversed	
3,4	No	Yes	Yes	Table insoluble
3,5	Yes	No	Yes	
3,6	Yes	Yes	Yes	
3,7	—	—	—	(No permutation)
3,8	—	—	—	(No permutation)
3,9	No	No	Yes	Table insoluble
3,10	Yes	Yes	Yes	
4,5	Reversed	Reversed	Reversed	
4,6	Reversed	Reversed	Reversed	
4,7	No	No	Reversed	Table insoluble
4,8	Yes	Yes	Reversed	
4,9	No	Yes	Reversed	Table insoluble
4,10	Reversed	Reversed	Reversed	
5,6	No	No	Reversed	Table insoluble
5,7	No	Yes	Reversed	Table insoluble
5,8	No	No	Reversed	Table insoluble
5,9	Reversed	Reversed	Reversed	Table insoluble
5,10	Reversed	Reversed	Reversed	
6,7	Yes	Yes	Yes	Table insoluble
6,8	No	Yes	Yes	
6,9	Yes	No	Yes	
6,10	Yes	Yes	Yes	
7,8	—	—	—	(No permutation)
7,9	Yes	Yes	Yes	
7,10	No	Yes	Yes	Table insoluble
8,9	No	Yes	Yes	Table insoluble
8,10	Yes	Yes	Yes	
9,10	No	Reversed	Yes	Table insoluble

Applying RC(1) to transitive reasoning data:

- Nine violations of original order after RC(1) sorting:

Item	Number of Violations	Type of Violation
1	1	Swap 0 and 1
2	3	Swap 0 and 1
4	1	Swap 0 and 1
5	1	Swap 0 and 1
8	1	Swap 0 and 1
10	1	Swap 0 and 1
11	1	Swap 0 and 1

4.5 MILR for more than three categories

- Exhaustive search would be time-consuming:
  - for a 5×5 table, the total number of orderings is $5!^2 = 14400$;
  - at 0.5 seconds of computation to examine each table, that's 2 hours of computation per table.
  - For IIP, there are $\binom{127}{2} = 8001$ tables, leading to over 16,000 hours of computation in the worst case.
- Two approaches to generalizing to higher-order tables will be considered in the future:
  - Directly generalizing the algorithm to more than three categories; *(No luck yet)*
  - Adapting principal components analysis to the problem of maximizing all 2×2 subtable associations in a cross-classification of pairs of items. *(RC models fit the bill nicely)*

5 Discussion

- Goal: exploratory data analysis (EDA) techniques for examining the order of item responses for unidimensional IRT.
- M (np-GRM):
  - $P[X_j > r|\theta]$ is nondecreasing in θ;
  - manifest monotonicity $P[X_j \geq r|S = s]$ should be nondecreasing in s
  - Variations on Guttman scaling can be used to sort categories and improve monotonicity properties
- MILR (np-PCM):
  - $P_{j(s+1)}(\theta)/P_{js}(\theta)$ is nondecreasing in θ.
  - manifest MILR: all 2×2 subtables the table cross-classifying any two items X_i and X_j should be positively associated.
  - Combinatorial algorithms are possible but difficult to derive and subject to capitalization on chance.
  - RC models give scores that can be used analogously to the scores in Guttman scaling for sorting categories; currently limited to pairs of items only.

- It is important to have a fully-developed EDA methodology, to confirm existing category orders as well as to search for new category orders.

- Whether the new category orders should be adopted remains, as always, a subject-matter question that is best answered in conjunction with the item writers and the users of the scale, test or item set being considered.