A view of nonparametric item response theory

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

Thanks to Ellis, Grayson, Holland, Levine, Molenaar, Sijtsma, Stout, Ramsay.

Nonparametric IRT

- Framework for J items.
  - * LI, M, d=1 (MH)
  - * Invariant item ordering (IO/DM)
  - * Polytomous generalizations
- Theoretical problems with finitely many items.
- The infinite item pool formulation.
- A combination of approaches.

Based primarily on:

– Junker (1993, *Ann Stat*)

– Sijtsma and Junker (1994a,b, submitted)

Framework for J items

- N examinees
- J items, $\underline{X} = (X_1, X_2, \ldots, X_J)$
- The locally independent "marginal" model:

$$p(\underline{x}) = p(x_1, \ldots, x_J) = \int_{-\infty}^{\infty} \prod_{j=1}^{J} P_j(\theta)^{x_j} (1 - P_j(\theta))^{1-x_j} dF(\theta).$$

- "Mokken" models:

Monotone Homogeneity (MH):
dim$(\theta) = 1$; $P_j(\theta)$ nondecreasing.
(LI, M, d=1).

Invariant item ordering (IO):
dim$(\theta) \geq 1$; $P_j(\theta)$'s do not cross.
Rosenbaum (1987 *Psy*)

Double Monotonicity (DM):
MH plus IO.

LI, M, d=1 (MH)

- Holland and Rosenbaum (1986. *Ann. Stat*)

For all $\underline{X} = (\underline{Y}, \underline{Z})$, all nondecr. $f(\underline{Y})$, $g(\underline{Y})$ and all $h(\underline{Z})$,

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

Conditional Association

- $f()$, $g()$: subtest scores
  - $\sum_j a_j Y_j$
  - $\min(Y_1, \ldots, Y_J)$
  - $\max(Y_1, \ldots, Y_J)$
  - Y_j

- $h()$ is "matching information" of any kind

- So much internal coherence that subtest scores must always "agree" no matter what other information *about examinees performance* you can get.

LI, M, d=1 (MH)

- Grayson (1988, *Psy*): Monotone Likelihood Ratio

$$s > t \Rightarrow \frac{P[X_+ = c|\theta = s]}{P[X_+ = c|\theta = t]} \text{ nondecr. in } c \qquad (MLR)$$

- Corollaries: Stochastic Ordering:

$$s > t \Rightarrow P[X_+ > c|\theta = s] \geq P[X_+ > c|\theta = t] \quad (SO-1)$$

and

$$s > t \Rightarrow P[\theta > c|X_+ = s] \geq P[\theta > c|X_+ = t] \quad (SO-2)$$

- SO-1 says there is "signal" in X_+
- SO-2 is directly useful for inference: $P[\text{mastery}|X_+]$
- Huynh-Huynh (1994, *Psy*);
- Ghurye and Wallace (1959, *Ann. Math. Stat*) —SO-2 problem posed by Loevinger in 1959!

LI, M, d=1 (MH)

• Manifest Monotonicity (Junker 1993, *Ann Stat*):

Let $S = X_+ - X_j$ (rest-score); then

$$t > s \Rightarrow P[X_j = 1|S = t] > P[X_j = 1|S = s] \qquad (MM)$$

* Nondecreasing item/rest-score regression.

* Holds under LI, M; converse under EI (later).

• Proof using SO-2:
Junker (1993) and Molenaar (priv. comm.)

• SO-2 can be true without MLR. But when? An open question in nonparametric IRT...

LI, M, d=1 (MH)

- Grayson (unpublished):
  - * $X_1, \ldots, X_J$ from LI, M, d=1 model
  - * $P_1(\theta), \ldots, P_J(\theta)$ *any* montone ICC's
  - * $\hat{\theta}(\underline{X})$ = the MLE *using the $P_j(\theta)$'s*

  Then
  - * $\hat{\theta}(\underline{X})$ is nondecreasing in $\underline{X}$
  - * $\hat{\theta}$ stochastically ordered by θ:

$$t > s \Rightarrow P[\hat{\theta} > c|\theta = t] \geq P[\hat{\theta} > c|\theta = s] \quad (1^*)$$

even if $\hat{\theta}$ from <u>wrong</u> ICC's!

- There is even "signal" in the wrong MLE's!

- $\hat{\theta}$ may or may not have MLR. If so, then also:

$$t > s \Rightarrow P[\theta > c|\hat{\theta} = t] \geq P[\theta > c|\hat{\theta} = s] \quad (2^*)$$

As $J \to \infty$, $\hat{\theta}$ should have approximate MLR; $\Rightarrow$ (2*), approximately.

LI, M, d=1 (MH)

- Mokken/Loevinger's H coefficient(s) for covariances between *Items*

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

$$H_j = \frac{\sum_{i \neq j} \mathrm{Cov}\,(X_i, X_j)}{\sum_{i \neq j} \mathrm{Cov}_{max}(X_i, X_j)}$$

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

- $\mathrm{Cov}_{max}(X_i, X_j)$ maximized over fixed marginal distributions of X_i and X_j.
- H=0 $\Leftrightarrow$ Pairwise Independence of *Items*
  * Flat item response functions
- H=1 $\Leftrightarrow$ Guttman Scale

- $H \in (0, 1)$?

LI, M, d=1 (MH)

- Fréchet bounds:

$$F_i(x) = P[X_i \le x];$$
$$F_{ij}(x,y) = P[(X_i, X_j) \le (x,y)]$$

 * $F_{ij}(x,y) \ge F_{ij}^L(x,y) = \max\{0, F_i(x) + F_j(y) - 1\}$

 * $F_{ij}(x,y)$ vs. $F_{ij}^I(x,y) = F_i(x)F_j(y)$?

 * $F_{ij}(x,y) \le F_{ij}^U(x,y) = \min\{F_1(x), F_2(y)\}$

- H: Cov(X_i, X_j) under $F_{ij}(x,y)$
  vs. Cov(X_i, X_j) under $F_{ij}^U(x,y)$.

- Joe (1989, *JASA*; 1990, *J. Mul. Anal.*; etc):
  * Modeling with fixed $F_i(x)$, $F_j(x)$;
  * measures of *multivariate concordance*.

- Kimeldorf and Sampson (1978. *Ann. Stat*):
  * $\max\{\rho(f(X), g(Y)) \; : \; f(), g()$ nondecreasing$\}$
  (see also: Kimeldorf, May, Sampson, 1982)

Invariant Item Order (IO)

- X_i uniformly harder than X_j (Rosenbaum, 1987, *Psy*):

$$P_i(\theta) \leq P_j(\theta) \;\forall\; \theta \qquad (IO)$$

(M, d=1: not required!)

- Invariant item ordering: $i < j \Rightarrow X_i$ unif. harder than j:

$$\text{(hard)} \quad P_1(\theta) \leq P_2(\theta) \leq \cdots \leq P_J(\theta) \quad \text{(easy)}$$

- Would like to have stochastic ordering:

$$i < j \Rightarrow P[\theta > c | X_i = 1] \geq P[\theta > c | X_j = 1]$$

Post and Snijders (1993, *M'ka*): This works iff

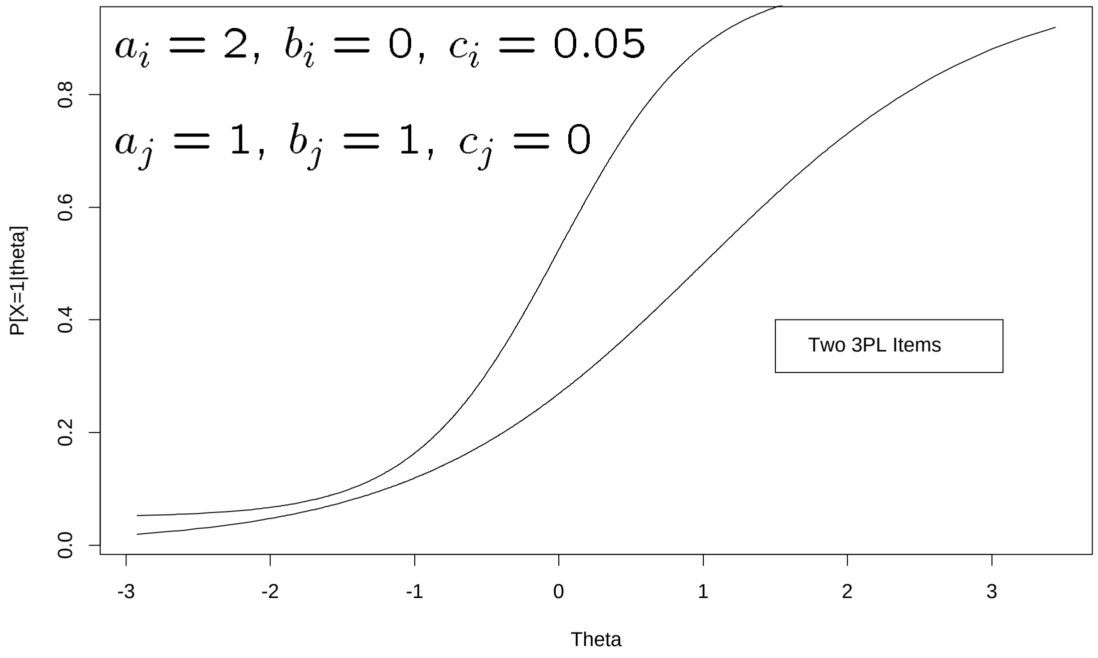
$$\frac{P[X_i = 1|\theta]}{P[X_j = 1|\theta]} \text{ is nondecreasing in } \theta \qquad (MTR)$$

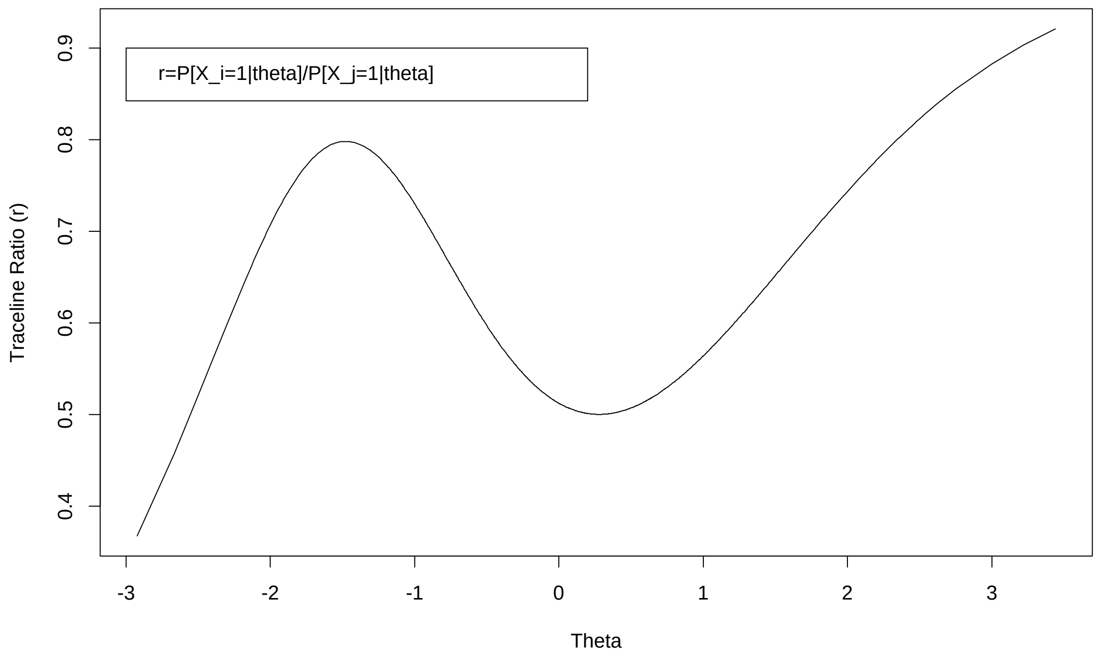
a kind of MLR for items.

- What kinds of items have both MTR and IO?

Invariant Item Order (IO)

- IO doesn't always imply MTR





Invariant Item Order (IO)

- X_i vs. X_j, given $\underline{Y}$ = (other items):

X_i uniformly harder than X_j $\Rightarrow$

$$P[X_i = 1|h(\underline{Y}) = c] \le P[X_j = 1|h(\underline{Y}) = c]$$

(<u>*Rosenbaum's inequality*</u>, 1987 *Psy*)

- $h(\underline{Y}) = X_k$, $k \neq i, j$.

  * Mokken's (1971) $P(++)$ and $P(--)$:

$$\begin{array}{rcl} P[X_i = 1, X_k = 1] & \le & P[X_j = 1, X_k = 1] \\ P[X_i = 0, X_k = 0] & \ge & P[X_j = 0, X_k = 0] \end{array}$$

  * Proposal of Molenaar (1991, ms.):

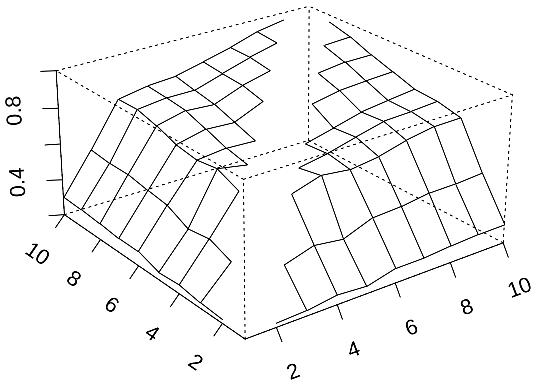
$$\begin{array}{rcl} P[X_i = 1|X_k = 1] & \le & P[X_j = 1|X_k = 1] \\ P[X_i = 0|X_k = 1] & \le & P[X_j = 0|X_k = 1] \end{array}$$

- $h(\underline{Y}) = \sum_{k \neq i,j} X_k \equiv S$

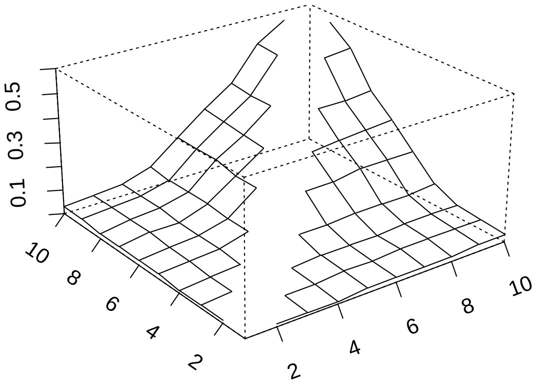
  * Proposal of Sijtsma and Junker (1994, subm.): "Rest-score splitting"

$$\begin{array}{rcl} P[X_i = 1|S > s] & \le & P[X_j = 1|S > s] \\ P[X_i = 0|S \le s] & \le & P[X_j = 0|S \le s] \end{array}$$

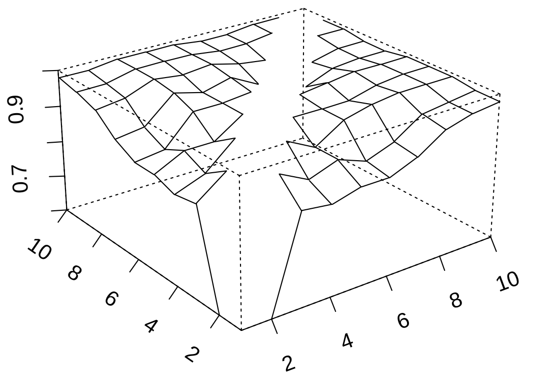
The $P(++)$ and $P(--)$ matrices



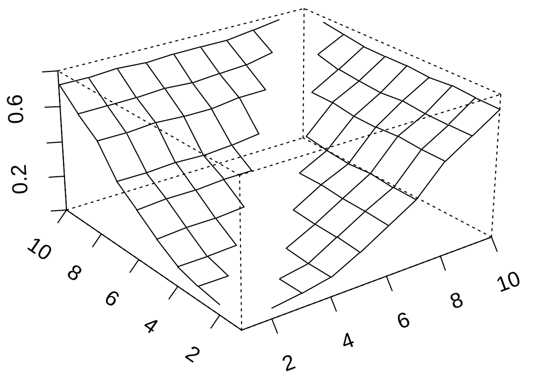
Off-diagonal entries of P(++) matrix (judgement only)



Off-diagonal entries of P(++) matrix (judgement plus)

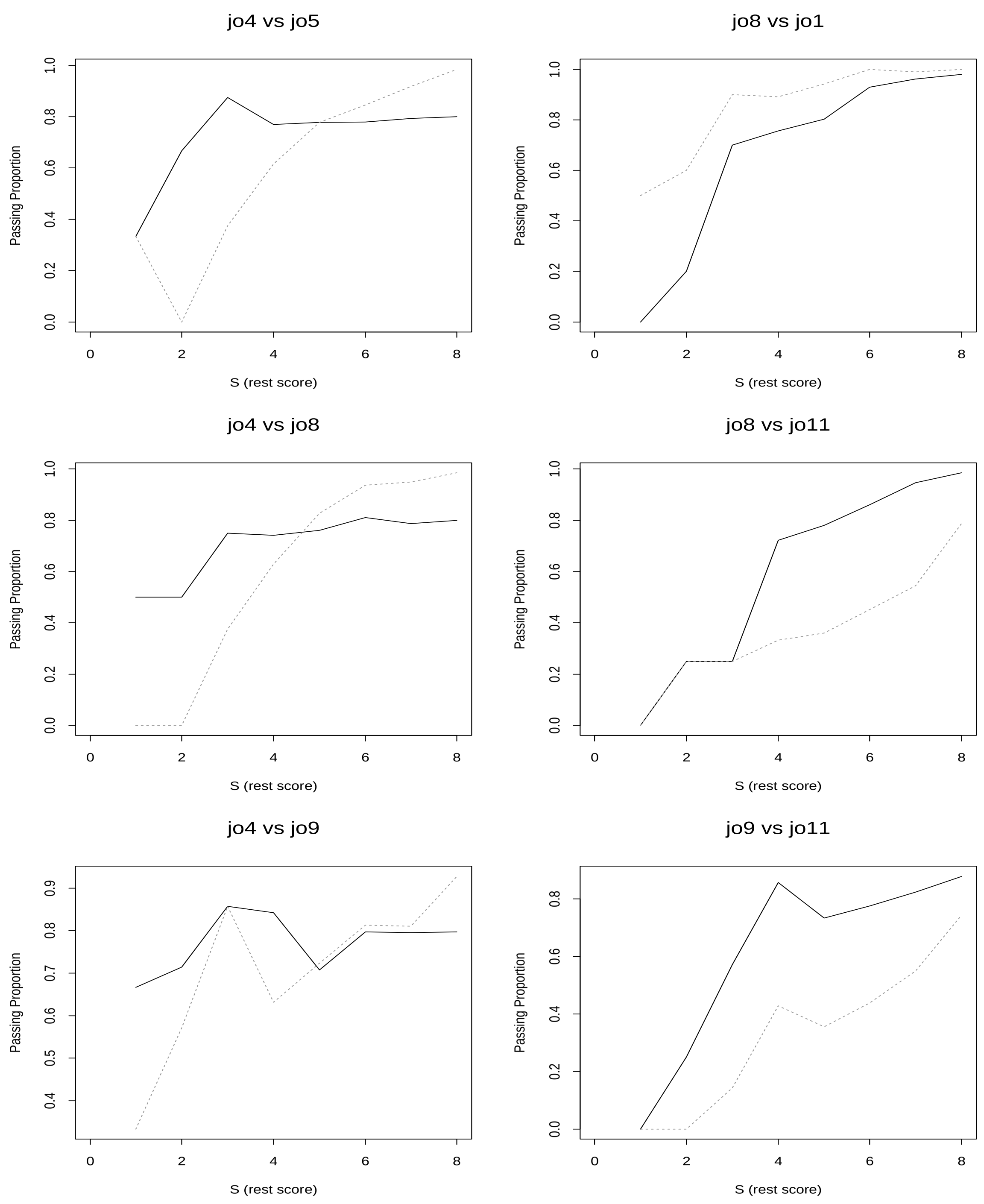


Off-diagonal entries of 1-P(--) matrix (judgement only)



Off-diagonal entries of 1-P(--) matrix (judgement plus)

Rest-score splitting



Invariant Item Order (IO)

- Sijtsma & Meijer (1992, *APM*): H^T

  * Data matrix (of 0's and 1's), $\mathcal{X} = [X_{ij}]$

  * Apply H_{ij}, H_j, H to $\mathcal{X}^T$: get H^T_{ij}, H^T_j, H^T

- $H^T = 0 \Leftrightarrow$ Pairwise Independence of *Persons*
  * Flat person resp. functions; IRF's coincide

- $H^T = 1 \Leftrightarrow$ Guttman scale

- $H^T \in (0, 1)$?

Tests of MH and IO [or DM]

- Item-level tests based on nested hypotheses:

  H_0: independence

  H_1: MH or IO models (e.g. Cov$\geq$0)

  H_2: non-monotone / non-ordered items

"Accepting" H_1 (or H_0) supportive of MH/IO

- Tests based on H, H^T etc. based on:

  H_0: independence

  H_1: positive dependence (including MH/IO)

  H_2: Guttman scale

"Rejecting" H_0 supportive of MH/IO

- Ellis (1994, Ch 5) obtains simple asymptotic distribution theory for H and considers tests based on order-restricted estimates of $\not\Sigma$.

  * Builds on Robertson, Wright and Dykstra (1988).

  * Uses first set of nested hypotheses.

  * Begins to address multiple inference.

Polytomous Generalizations

Let S be a "nondecreasing score"
—maybe $S = X_j$, $S = X_+$ or $S = X_+ - X_j$.

- Two stochastic ordering conditions

$$s > t \Rightarrow P[S > c|\theta = s] \geq P[S > c|\theta = t] \qquad (SO-1)$$

$$s > t \Rightarrow P[\theta > c|S = s] \geq P[\theta > c|S = t] \qquad (SO-2)$$

- SO-1 with $S = X_j$ gives a reasonable generalization of *monotonicity* to polyotomous items.

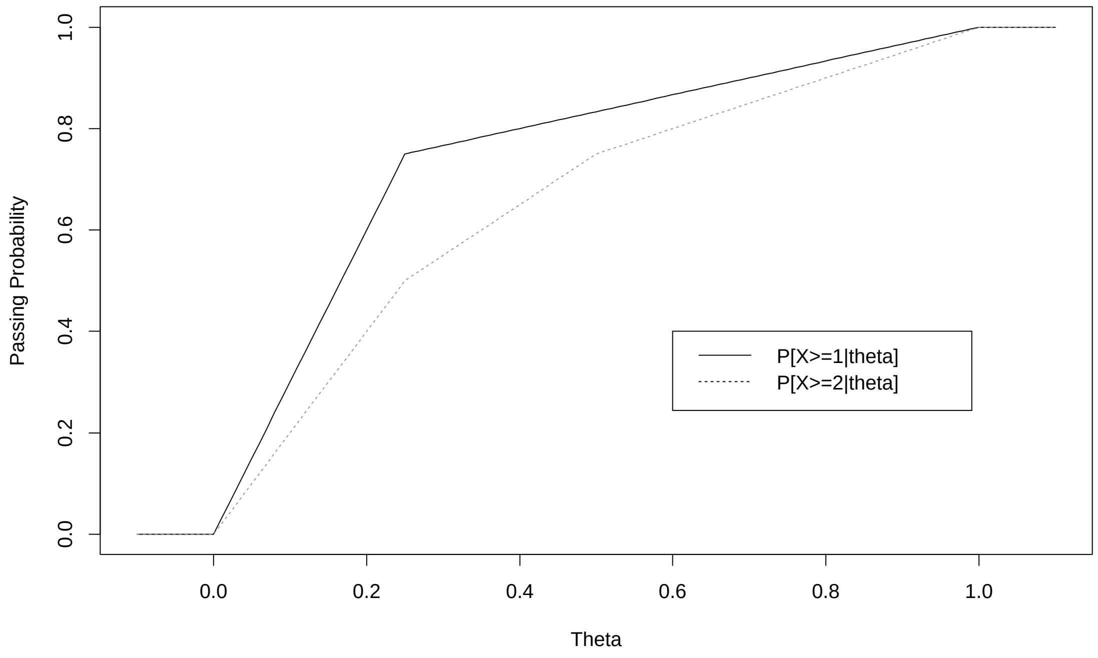
- SO-2 gives Manifest Monotonicity (nondecreasing item/rest-score regressions)

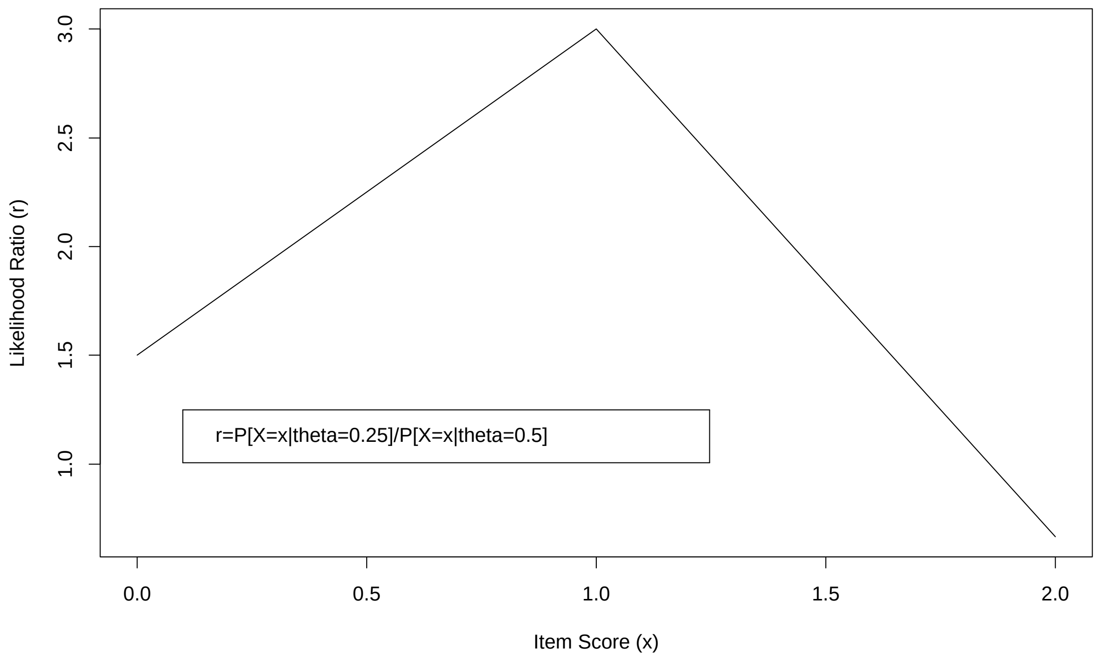
- Both SO-1 and SO-2 are implied by *monotone likelihood ratio*

$$s > t \Rightarrow \frac{P[S = c|\theta = s]}{P[S = c|\theta = t]} \text{ nondecr. in } c \quad (MLR)$$

Polytomous Generalizations

- SO doesn't always imply MLR





Polytomous Generalizations

LI, M, d=1 (MH):

• Hemker and Sijtsma (in progress): investigating MLR in Rasch-based polytomous models (and their generalizations).

  * Things look bad for nonparametric MLR.

• Where do we go from here? MLR is too strong a property. Must think about SO-2 directly. Much harder problem!

Summary Measures:

• Molenaar (1991, *QM*): same definitions for H for polytomous items; can be interpreted as a ratio of error counts, weighted by severity of errors (incorporated into MSP).

Polytomous Generalizations

Invariant item order (IO and DM):

- Two definitions possible:

(1) X_i uniformly harder than X_j:

$$P[X_i \geq x|\theta] \leq P[X_j \geq y|\theta] \ \forall \ x, y \ \forall \ \theta$$

(2) X_i and X_j are "order-able": For each x, y:

$$P[X_i \geq x|\theta] \leq P[X_j \geq y|\theta] \ \forall \ \theta$$

or

$$P[X_j \geq x|\theta] \leq P[X_i \geq y|\theta] \ \forall \ \theta$$

(depending on the values of x and y).

- No general results are known for (1) or (2).

- Can apply existing methods (based on Rosenbaum's inequality) to various "dicotomizations" of the items—not satisfying.

Finitely many items

Think dichotomous again...

- *Necessary conditions:*
  * LI, M, d=1 $\Rightarrow$ CA, MLR, SO-1, SO-2, MM, etc.
  * LI, IO, d=k $\Rightarrow$ Rosenbaum's $\neq$, $P(++)$, $P(--)$, rest-score splitting, etc.

- *Sufficient conditions:*
  * Rasch model iff

$$\log p(\underline{x}) = \alpha + \sum_{j=1}^{J} \beta_j x_j + \gamma(x_+)$$

where $\gamma(k)$ satisfies "moment inequalities": Tjur (1982, *St.N.*); Cressie and Holland (1983, *Psy*); de Leeuw and Verhest (1983, *JES*); Lindsay, Clogg and Grego (1991, *JASA*).

- *Sufficient conditions* for nonparametric IRT?

Finitely many items

• <u>*Sufficient conditions*</u> for nonparametric IRT?

* Equivalent to estimating <u>*not only*</u> IRF's $P_j(\theta)$ <u>*but also*</u> θ distribution nonparametrically.

* Haberman (1977, *Ann. Stat.*) Parametric JMLE requires *both* $N \to \infty$ *and* $J \to \infty$.

* Levine (subm.), summarized by Ramsay (1991, *Psy*) and Drasgow et al. (1989, *APM*): *Suppose the IRF's* $P_j(\theta)$ *known. Then all that can be known about* θ *and its distribution must be in the linear span of the* 2^J *functions*

$$P_1(\theta), P_2(\theta), \ldots, P_J(\theta),$$
$$P_1(\theta)P_2(\theta), P_1(\theta)P_3(\theta), \ldots, P_{J-1}(\theta)P_J(\theta),$$
$$\vdots$$
$$P_1(\theta)P_2(\theta)\cdots P_J(\theta)$$

the <u>canonical space</u> *for Levine's "multilinear formula score theory".*

The infinite item pool formulation

Letting $J \to \infty \ldots$

- Embed the J items, $\underline{X} = (X_1, X_2, \ldots, X_J)$ in an *infinite item pool*

$$\underline{X}_\infty = X_1, X_2, \ldots, X_J, X_{J+1}, \ldots$$

- *LI, M, d=1 (MH)* holds for $\underline{X}_\infty$ iff it holds in every J-item subset (for all J), with the same IRF's and the same Θ distribution.

- *Invariant item ordering (IO)* holds for $\underline{X}_\infty$ iff the IRF's of any two items do not cross.

- Most clearly formulated by Stout (1987, 1990 *Psy*); polytomous case: Junker (1991, *Psy*).

The infinite item pool formulation

- LI requires $X_i \coprod X_j|\theta$: $\text{Cov}\,(X_i, X_i|\theta) = 0$.

- <u>*Essential Independence (EI):*</u> Let $\overline{A}_J = \frac{1}{J}\sum_{j=1}^J a_j X_j$. Then $\underline{X}_\infty$ satisfies EI iff

$$\lim_{J\to\infty} \text{Var}\,(\overline{A}_J|\theta) = 0$$

for all θ and all bounded sets of weights a_j.

- Equivalently,

$$\binom{J}{2}^{-1} \sum\sum_{1\le i<j\le J} \text{Cov}\,(a_i X_i, a_j X_j|\theta) \to 0.$$

- <u>*Locally Asymptotically Discriminating (LAD)*</u>: Let $\overline{A}_J(\theta) = E[\overline{A}_J|\theta]$ Then $\underline{X}_\infty$ satisfies LAD iff

$$A'_J(\theta) \ge \varepsilon_{\theta,\{a_j\}} > 0$$

as $J \to \infty$, for all θ and all bounded sets of weights a_j.

The infinite item pool formulation

- *Useful latent trait model* (Junker 1993):
  (1) Consistent estimation of θ as $J \to \infty$
  (2) $E[\overline{A}_J|\theta]$ nondecreasing
  (3) Nontrivial θ distribution (more than one examinee!)

- Examples:
  * EI, LAD: <u>*Essential unidimens.*</u> ($d_E = 1$)
  * LI, M, LAD: <u>*Strict unidimens.*</u> ($d_L = 1$)

- Consistent estimation of θ (Stout, 1990 *Psy*) and of $E[f(\underline{X})|\theta]$ (Junker, 1993 *Ann Stat*)

- $d_E = 1$: Unique trait θ, up to monotone (ordinal) transformation (Stout, 1990 *Psy*)

- $d_L = 1$: θ describes exactly all of the long-run features of examinee responses (Junker, 1993 *Ann Stat*)

A combination of approaches

- *Manifest Monotonicity*

  * We know LI, M $\Rightarrow$ MM

  * How about EI, LAD, MM $\Rightarrow$ M?

  Let $S_J = X_+ - X_j$ and construct a consistent estimator $\widehat{\theta}(S_J)$. If $\theta_1 < \theta_2$ then

$$\begin{aligned} P_j(\theta_1) &\approx E[X_j|\widehat{\theta}(S_J) \approx \theta_1] \\ &\leq E[X_j|\widehat{\theta}(S_J) \approx \theta_2] \\ &\approx P_j(\theta_2) \end{aligned}$$

  * Hence: LI, LAD: M $\Leftrightarrow$ MM

  Junker (1993, *Ann Stat*)

- *Double Monotonicity:* Let $S = X_+ - X_i - X_j$

$$\theta_a \leq \theta_b \Rightarrow P(X_i = 1|\theta_a) \leq P(X_j = 1|\theta_b) \qquad (DM)$$

iff

$$s_a \leq s_b \Rightarrow P(X_i = 1|S = s_a) \leq P(X_j = 1|S = s_b) \qquad (MDM)$$

Sijtsma and Junker (1994, subm.)

A combination of approaches

- *Strict Unidimensionality*

  $d_L = 1 \Rightarrow d_E = 1$, CA, MM, LCSN

LCSN (Covariance given Score is Negative):

$$\text{Cov}\,(X_i, X_j | \overline{X}_J, \theta) \le 0$$

Rasch model: θ disappears ($\overline{X}_J = X_+/J$ is sufficient!), so we get CSN:

$$\text{Cov}\,(X_i, X_j | \overline{X}_J) \le 0$$

- *Partial characterization* of $d_L = 1$:

  $d_L = 1 \Leftrightarrow d_E = 1$, CA, MM, LCSN

  $d_L = 1 \Leftarrow d_E = 1$, CA, MM, CSN

  (Junker 1993, *Ann Stat*)

- *Partial characterization* of $d_L = 1$, DM:

  $d_L = 1 \Leftrightarrow d_E = 1$, CA, DM, LCSN

  $d_L = 1 \Leftarrow d_E = 1$, CA, MDM, CSN

  (Sijtsma and Junker 1994, subm.)

Concluding thoughts

- *Nonparametric IRT:*

  * Item-level tests of MH: CA, MM, $d_E = 1$, (L)CSN

  * Inferences: $s > t \Rightarrow P[\theta > c|S = s] \geq P[\theta > c|S = t]$ for appropriate score S.

  * Item-level tests of IO/DM: Rosenbaum's $\neq$, $P(++)/P(--)$, rest-score splitting...

  * Summary statistics H for MH, H^T for DM.

- *Testing:*

  * Testing should *accept* MH/DM, not *infer it from rejection* of another model.

  * Item-level tests already do this; inference for H should follow suit.

  * Multiple inference $\Rightarrow$ tests not serious one-by-one.
    – Analogous to "regression diagnostics"
    – Graphical summaries tell the story

- *Polytomous:* Much to do!

- *Necessary and sufficient conditions:* Use the infinite item pool formulation.