A view of nonparametric item response theory[a]

Educational Testing Service
Princeton, NJ
July 6, 1998

Brian W. Junker
Department of Statistics
Carnegie Mellon University
Pittsburgh PA 15213
brian@stat.cmu.edu

[a] Thanks to Sijtsma, Hemker, Molenaar, Grayson, Stout, Ellis. My research was supported in part by the by the National Institutes of Health, Grant CA54852, and the National Science Foundation, Grants DMS-94.04438 and DMS-97.05032

Talk Outline:

- Nonparametric Monotone Unidimensional Item Response Models

- Stochastic Ordering

- Invariant Item Ordering

- Some Approaches to Characterization and Testing

Nonparametric Monotone Unidimensional Item Response Models

- $X_1, X_2, \ldots, X_J$: polytomous; $X_j \in \{0, 1, \ldots, m_j\}$
  θ: the latent variable.

- ICRF's: $P_{jx}(\theta) = P[X_j = x|\theta]$;
  I(S)RF's: $P^*_{js}(\theta) = P[X_j \geq s|\theta] = \sum_{x=s}^{m_j} P_{jx}(\theta)$.

- <u>Basic IRT framework</u>:

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

  assuming

  - *unidimensionality (U)*: $dF(\theta)$ on $\Re$.
  - *local independence (LI)*:
    $P[X_1 = x_1, \ldots, X_J = x_J|\theta] = \prod_{j=1}^{J} P_{jx_j}(\theta)$;

- $X_j \in \{0, 1\} \Rightarrow P^*_{j1}(\theta) = P[X_j = 1|\theta] = P_j(\theta)$.

Nonparametric monotone models

- Nonparametric graded response model (np-GRM; Hemker et al, 1997; Samejima, 1969):

  $P^*_{js}(\theta)$ nondecreasing in θ for each j and s

  Many names: monotone unidimensional (Holland and Rosenbaum, 1996); strictly unidimensional (Stout, 1990; Junker, 1991, 1993); montone homogeneity (Mokken, 1971); ...

- Nonparametric partial credit model (np-PCM; Hemker et al, 1997):

  $\phi_{jx}(\theta) \equiv P[X_j = x | X_j \in \{x-1, x\}, \theta]$
  nondecreasing in θ for each j and x

- Nonparametric sequential model (np-SM; Hemker, 1996; Tutz, 1990; Mellenbergh, 1995; Samejima acceleration model, 1969, 1995):

  $\dfrac{P^*_{js}(\theta)}{P^*_{j(s-1)}(\theta)}$ nondecreasing in θ for each j and s

Nonparametric monotone models

For the np-PCM,

$$\phi_{jx}(\theta) = \frac{1}{1 + P_{j(x-1)}(\theta)/P_{jx}(\theta)}$$

so $\phi_{jx}(\theta) \nearrow$ if and only if $\frac{P_{jx}(\theta)}{P_{j(x-1)}(\theta)} \nearrow$, a monotone likelihood ratio (MLR) property for X_j and θ.

For the np-SM,

$$\frac{P^*_{js}(\theta)}{P^*_{j(s-1)}(\theta)} = \frac{1}{1 + P_{j(s-1)}(\theta)/P^*_{js}(\theta)}$$

and

$$\prod_{s=1}^{t} \frac{P^*_{js}(\theta)}{P^*_{j(s-1)}(\theta)} = P^*_{jt}(\theta)$$

Using these observations, Hemker (1996) shows that

$$\boxed{\text{np-PCM} \Rightarrow \text{np-SM} \Rightarrow \text{np-GRM}}$$

Stochastic Ordering — Binary Items

- Grayson (1988, *Psy*): Let $X_+ = \sum_1^J X_j$; if $P_j(\theta) \nearrow$ then

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

- Corollaries: Stochastic Ordering:

$$s > t \;\Rightarrow\; P[X_+ > c|\theta = s] \geq P[X_+ > c|\theta = t] \tag{SOM}$$

and

$$s > t \;\Rightarrow\; P[\theta > c|X_+ = s] \geq P[\theta > c|X_+ = t] \tag{SOL}$$

- SOM says there is "signal" in X_+
- SOL: $P[\text{mastery}|X_+]$, $E[\theta|X+] \nearrow, \ldots$
- Huynh-Huynh (1994, *Psy*);
- Ghurye and Wallace (1959, *Ann. Math. Stat*) —SOL problem posed by Loevinger in 1959!

Stochastic Ordering — Polytomous Items

Let $X_+ = \sum_1^J X_j$ again; SOM, SOL and MLR look the same:

$$s > t \;\Rightarrow\; P[X_+ > c|\theta = s] \geq P[X_+ > c|\theta = t] \quad (SOM)$$

$$s > t \;\Rightarrow\; P[\theta > c|X_+ = s] \geq P[\theta > c|X_+ = t] \quad (SOL)$$

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

- Still have MLR $\Rightarrow$ SOM, SOL
- np-GRM $\Rightarrow$ SOM (Holland and Rosenbaum, 1986, Lemma 2).
- What about SOL? Two approaches (Hemker et al, 1996, 1997):
  - Try to prove MLR
  - Try to work directly with SOL

MLR — Polytomous Items (Hemker et al, 1996):

Suppose $\frac{d}{d\theta}P_{jx}(\theta)$ exists; then MLR holds,
iff $s > t$ implies $\forall\ \theta$

$$\begin{aligned}\frac{d}{d\theta}P(X_+ = s|\theta) \cdot P(X_+ = t|\theta)& \\ - \ P(X_+ = s|\theta) \cdot \frac{d}{d\theta}P(X_+ = t|\theta) &\geq 0;\end{aligned}$$

iff $s > t$ implies $\forall\ \theta$

$$\begin{aligned}\sum_{x_+=s}\sum_{y_+=t}\Bigg\{\left(\sum_{j=1}^{J}\frac{d}{d\theta}\log P_{jx_j}(\theta)/P_{jy_j}(\theta)\right)& \\ \cdot \quad \left(\prod_{j=1}^{J}P_{jx_j}(\theta)\cdot P_{jy_j}(\theta)\right)\Bigg\} &\geq 0\end{aligned}$$

For the np-PCM one can show that

- $P_{jx}(\theta) = \phi_{j0}\prod_{k=1}^{x}\frac{\phi_{jk}(\theta)}{1+\phi_{jk}(\theta)} = \frac{R_{jx}(\theta)}{1+\sum_{k=1}^{m_j}R_{jk}(\theta)}$
- The ugly triple sum condition for MLR is also true if $P_{jx}(\theta)$ is replaced with $R_{jx}(\theta)$ throughout.

MLR — Polytomous Items (Hemker et al, 1996):

- Masters' (1982) PCM $\Rightarrow$ MLR: $\log R_{jx}(\theta) = x\theta - \beta_{jx}$; the ugly sum cleans right up.

- np-GRM, np-PCM $\not\Rightarrow$ MLR;

- Even nice parametric logistic GRM's (Samejima) and GPCM's (Muraki) fail to have MLR.

- In the nice parametric models the parameters chosen to violate MLR are rather extreme. It may be that the parameters may be restricted in some reasonable way to get MLR. *(But this is cheating...)*

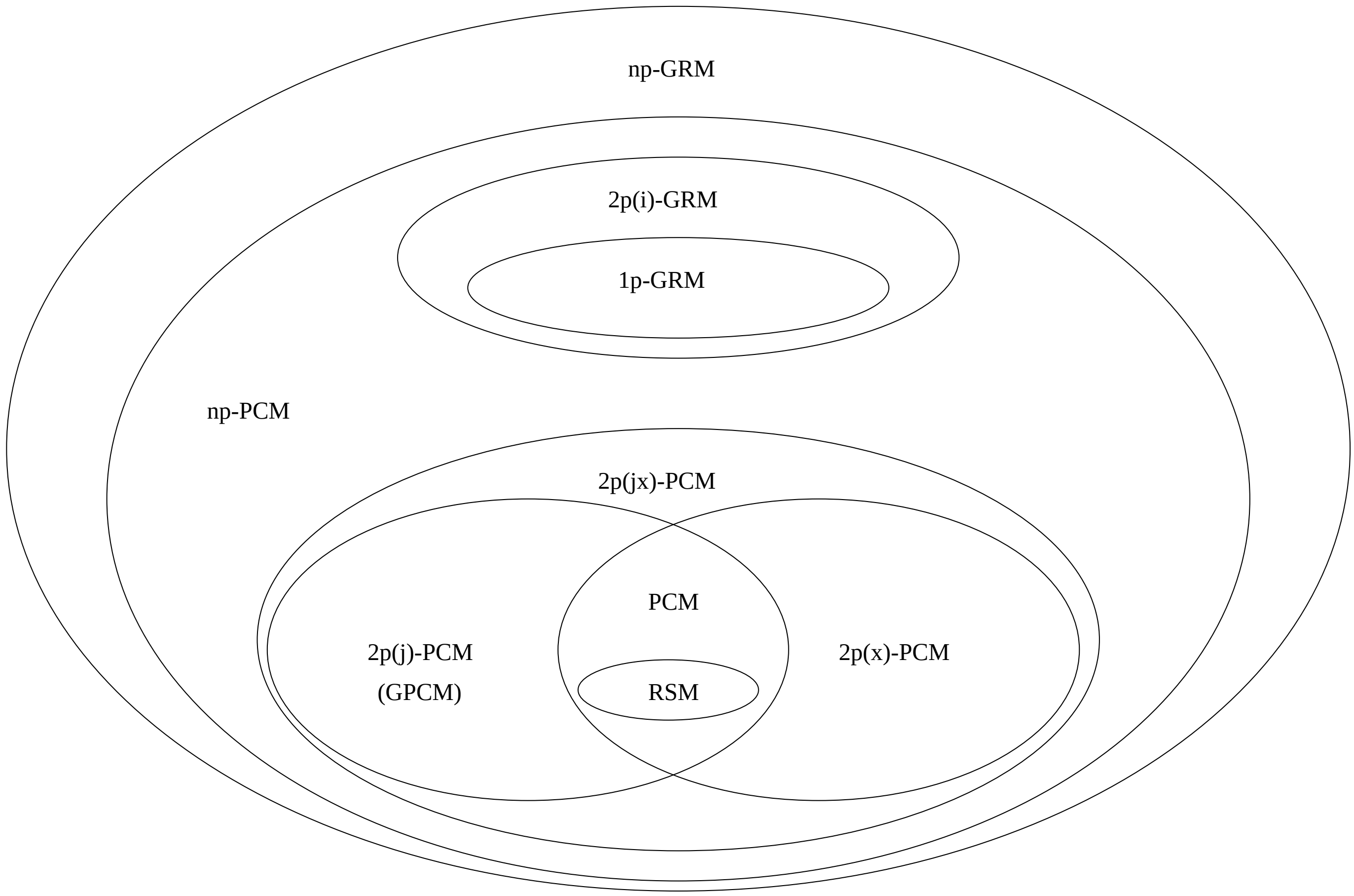
Bottom line:

Among commonly used polytomous IRT models, only the PCM has MLR of X_+ and θ.

SOL — *Polytomous items* (Hemker et al, 1997):

$$P[\theta > c | X_+ = s] \quad \nearrow \quad ?$$

- Masters' PCM ⇒ MLR ⇒ SOL
- Taxonomy of common polytomous IRT models



Bottom line:

Among commonly used polytomous IRT models, only the PCM has SOL.

Aside: Replace X_+ with something else?

- Grayson (unpublished, ca. 1990): **<u>binary IRT</u>**
  * $X_1, \ldots, X_J$ from LI, M, U model
  * $P_1(\theta), \ldots, P_J(\theta)$ **<u>any old</u>** 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:

$$t > s \;\Rightarrow\; P[\hat{\theta} > c | \theta = t] \geq P[\hat{\theta} > c | \theta = s] \qquad (SOMT)$$

  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:

$$t > s \;\Rightarrow\; P[\theta > c | \hat{\theta} = t] \geq P[\theta > c | \hat{\theta} = s] \qquad (SOLT)$$

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

- *Aside:* Correlation studies of recovery of θ by $\hat{\theta}$ offer weak evidence of success!

Invariant Item Ordering (Sijtsma and Hemker, 1998):

- **IIRFO:** For every j, k, s, t: If $P^*_{js}(\theta) \geq P^*_{kt}(\theta)$ holds for some $\theta = \theta_0$, then it holds for all θ.
- **IIO:** For every j, k: If $E[X_j|\theta] \geq E[X_k|\theta]$ holds for some $\theta = \theta_0$, then it holds for all θ.
- Why bother?
  - All examinees agree on the relative difficulty of the items.
  - Simplifies some adaptive testing, person fit, and DIF detection problems.
  - Some psychological (and social survey) scales are intended to be constructed so that later items layer on more skills (uniformly in θ); tests for IIO then are tests for the success of the item construction method.
- When $P^*_{js}(\theta)$ is also $\nearrow$, the model is *doubly monotone (DM)* (Mokken, 1971; Mokken and Lewis, 1982; Molenaar, 1991).

Invariant Item Ordering — Binary Items

For <u>*binary*</u> items, the two definitions coincide (and we say "IIO"); e.g.

$$P_1(\theta) \leq P_2(\theta) \leq \cdots \leq P_J(\theta)$$

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

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

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

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

$$\begin{aligned} P[X_i = 1, X_k = 1] &\leq P[X_j = 1, X_k = 1] \\ P[X_i = 0, X_k = 0] &\geq P[X_j = 0, X_k = 0] \end{aligned}$$

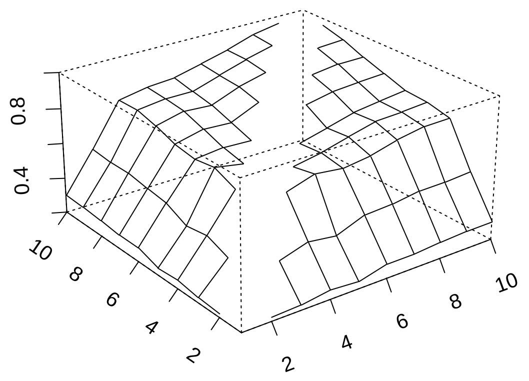
- Let $S = X_+ - X_i - X_j$. If

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

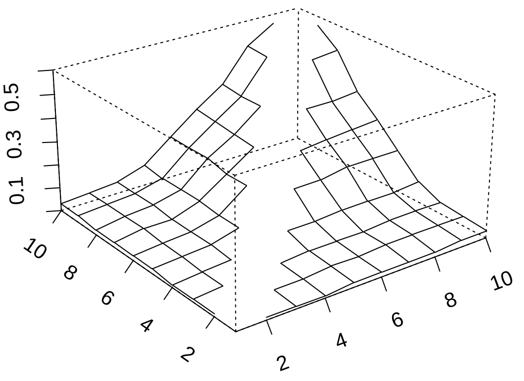
then

$s_a \leq s_b \Rightarrow P(X_i = 1|S = s_a) \leq P(X_j = 1|S = s_b)$, and the converse holds if the test is *essentially unidimensional* in the sense of Stout (1987, 1990). See Junker (1993) and Sijtsma and Junker (1996).

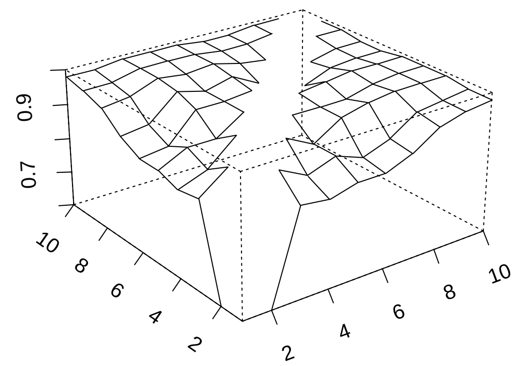
The $P(++)$ and $P(--)$ matrices



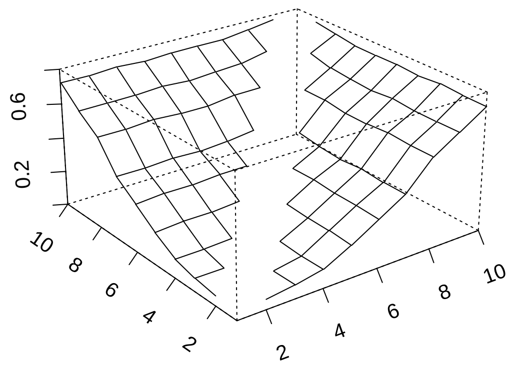
Off-diagonal entries of P(++) matrix (judgement only)



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)

Invariant Item Ordering — Polytomous Items

Sijtsma and Hemker (1998) consider three classes of monotone unidimensional models:

- **Weak DM:** IISRFO: For every j, k, s, t: If

  $$P^*_{js}(\theta) \geq P^*_{kt}(\theta)$$

  holds for some $\theta = \theta_0$, then it holds for all θ.

- **Strong DM:** IISRFO, and the "strong separation" condition: For every j, k, if

  $$P[X_j \geq x|\theta] \geq P[X_k \geq x|\theta] \qquad (SS)$$

  holds for some $\theta = \theta_0$ and some $x = x_0$, then it holds for all θ and all x.

- **Scheiblechner's ISOP model:** Satisfies the SS condition but not necessarily IISRFO.

and various parametric models.

Invariant Item Ordering — Polytomous Items

Using the identity

$$E[X_j|\theta] - E[X_k|\theta] = \sum_x (P[X_j \geq x|\theta] - P[X_k \geq x|\theta])$$

for items with the same number of categories, Sijtsma and Hemker (1998) show:

- The weak DM fails to have IIO
- Both the Strong DM and the ISOP model have IIO
- The RSM and a restricted RS-GRM have IIO

The Strong DM model makes unrealistic constraints on the ISRF's.

Not much is known yet about the ISOP model, though it is promising in many respects.

Again the bottom line seems to be

Among commonly used polytomous IRT models, only the RSM and a restricted RS-GRM have IIO.

Stochastic Ordering and Invariant Item Ordering — Summary

Dichotomous Data	SOL	IIO	Polytomous Data	SOL	IIO
Rasch	+	+	RSM	+	+
2PL	+	−	PCM	+	−
3PL	+	−	GPCM	−	−
4PL	+	−	RS-GRM*	−	−
OPLM	+	−	OPLM-po	−	−
1PNO	+	+	GRM	−	−
2PNO	+	−	MHM-po	−	−
3PNO	+	−	WEAK DM	−	−
MHM-di	+	−	STRONG DM	−	+
DMM-di	+	+	ISOP	−	+

*When RS-GSM is constrainted to have a single slope, it satisfies IIO.

Presence (+) or absence (−) of SOL and IIO in several common IRT models (Sijtsma and Hemker, 1998).

Some Approaches to Characterization and Testing

> *Want:* Necessary and Sufficient manifest conditions for nonparametric, monotone unidimensional model.

Think dichotomous again...

- *Necessary conditions:*
  - LI, M, $\dim(\theta) = 1 \Rightarrow$ MLR, SOM, SOL, etc.
  - LI, IIO, $\dim(\theta) \geq 1 \Rightarrow$ Rosenbaum's $\neq$, $P(++)$, $P(--)$, etc.

- *Sufficient conditions* 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.*), Douglas (1996, *Psy*): Parametric JMLE requires *both* $N \to \infty$ *and* $J \to \infty$.

Similarly. . .

Levine (tech rept), 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*

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

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

While this is a big space it is not infinite-dimensional and nonparametric estimation places $F(t) = P[\Theta \leq t]$ in an infinite-dimensional setting.

Some Recent Approaches to characterization and testing

- Stout (1987, 1990): least conditions for consistent estimation of θ; asymptotic in the items.

  EI: As $J \to \infty$,

  $$\binom{J}{2} \sum\sum_{1\leq i<j\leq J} |\mathrm{Cov}\,(X_i, X_j|\theta)| \to 0$$

  WM/LAD As $J \to \infty$,

  $$\frac{d}{d\theta}\left(\frac{1}{J}E[X_+|\theta]\right) \geq \epsilon_\theta > 0$$

- Holland and Rosenbaum (1986): strongest observable conditions from monotone unidimensional latent variable model. *Finitely many items considered.*

  CA Conditional Association; to be defined below.

  They note that this seems to be more restrictive as J grows.

- Junker (1993), Junker and Ellis (1997), Ellis and Junker (1997): Combine the Stout and Holland/Rosenbaum approaches:

  CA (for U and M):

  $$\mathrm{Cov}\,(f(Y_1,\ldots,Y_{J_1}),\ g(Y_1,\ldots,Y_{J_1}) \\ \mid \underline{h}(Z_1,\ldots,Z_{J_2})) \geq 0$$

  for all nondecreasing f, g; all vector-valued $\underline{h}$, and all partitions $(X_1,\ldots,X_J) = (Y_1,\ldots,Y_{J_1}, Z_1,\ldots,Z_{J_2})$.

  VCD (for LI):

  $$\mathrm{Cov}\,(f(Y_1,\ldots,Y_{J_1}),\ g(Z_1,\ldots,Z_{J_2}) \\ \mid X_{J+m},\ldots,X_{J+n}) \to 0$$

  as $m, n \to \infty$, for all J, all f and g, and all partitions $(X_1,\ldots,X_J) = (Y_1,\ldots,Y_{J_1}, Z_1,\ldots,Z_{J_2})$.

  - Characterization of LI, U, M latent structure in terms of CA, VCD.
  - "Measuring θ" $\approx$ "measuring long-run examinee response behavior." ($\sim$ De Finetti)

– Importance of co-monotonicity arguments in manifest-to-latent arguments (e.g. Ellis and Van den Wollenberg, 1993).

- Scheiblechner (1995): Reformulate the np-GRM axioms and symmetrize examinees and items:

  (W1) If there exist t_0 and j_0 such that $P[X_{j_0} > t_0|\theta_1] < P[X_{j_0} > t_0|\theta_2]$, then for all t and j $P[X_j > t|\theta_1] \leq P[X_j > t|\theta_2]$;

  (W2) If there exist t_0 and w_0 such that $P[X_{j_1} > t_0|\theta_0] < P[X_{j_2} > t_0|\theta_0]$, then for all t and θ $P[X_{j_1} > t|\theta] \leq P[X_{j_2} > t|\theta]$.

  W1 $\approx$ "$P^*_{jx}(\theta) \nearrow$" in the np-GRM; W2 is the non-intersection axiom we saw before. *Finitely many items considered.*

- Both Scheiblechner (1995) and Junker and Ellis (1997) construct θ from the item response variables

  – Scheiblechner's θ is a function of the first J items in an item pool.

  – Junker and Ellis's θ is a function of "the items you haven't seen yet", no matter how many you've seen.

Conclusions:

- Nonparametric Monotone Unidimensional Item Response Models
  - Has been a very fruitful area for thinking about the general features of IRT models, and inferences based upon them.

- Stochastic Ordering
  - dichotomous: MLR, SOM, SOL well known and well understood
  - polytomous:
    * SOM carries over
    * MLR and SOL seem to be very restrictive
    * Other scores besides X_+ have not been investigated

- Invariant Item Ordering
  - dichotomous:
    - Reasonably well understood
    - Diagnostics; most based on Rosenbaum's inequality; some based on SOL and similar properties
  - polytomous:
    - Not as restrictive as SOL? (ISOP and Strict DM models satisfy)
    - Few good tests and diagnostics (more on this tomorrow)

- Some Approaches to Characterization and Testing
  - Infinite item pool formulation needed for clear thinking
  - Ellis/Junker characterization best "so far"...
  - Has not lead to any diagnostics beyond those explored in the late 80's for CA.