Latent and manifest monotonicity in item response theory[a]

Educational Testing Service
Princeton, NJ
July 7, 1998

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

Klaas Sijtsma
Vakgroep Methoden en
Technieken van Onderzoek
Faculteit der Sociale Wetenschappen
Tilburg University
Tilburg, The Netherlands
k.sijtsma@kub.nl

[a] Junker's 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

Outline

1. Motivation
   (a) Nonparametric Monotone Unidimensional Item Response Model
   (b) Checks
      i. Precursor to Parametric Fit
      ii. Scaling/scale refinement
      iii. If model holds, use SOL for θ-inference
   (c) $P[X_1 = 1|S = s]$ – Sijtsma (1988), Ramsay (1991), Croon (1991), Scheiblechner (1995)
2. Positive Results
   (a) General models: $P[X_j = 1|S_{(-j)} = s]$
   (b) Rasch model: $P[X_j = 1|S = s]$
3. Thinking more generally
   (a) Elementary Symmetric Functions
   (b) Calculation on the Computer
4. Examples
   (a) Snijders
   (b) Simple 2PL
   (c) More realistic 2PL
   (d) Items with IIO
5. Conclusions

Motivation and Background

Monotone Unidimensional IRT model

- $X_1, X_2, \ldots$ binary items
- $P_j(\theta) = P[X_j = 1|\theta]$ *nondecreasing* in θ
- θ unidimensional with density or pmf $f(\theta)$

$$P[X_1 = x_1, \ldots, X_J = x_J] = \int \prod_{j=1}^{J} P_j(\theta)^{x_j}(1 - P_j(\theta))^{1-x_j} f(\theta) d\theta$$

- *Main idea:* Use $P[X_j = 1|\hat{\theta}]$ to check monotonicity of $P_j(\theta)$

Why Check Monotonicity?

- Precursor to parametric model fitting

- Nonparametric scaling / item analysis

- Nonparametric inference: If the model holds then may use

  $$P[\theta > t|S = s] \text{ is nondecreasing in } s \text{ for each } t. \qquad (SOL)$$

  to make inferences (e.g. *mastery*) on θ nonparametrically, where

  $$S = \sum_{j=1}^{J} X_j$$

  is the *total score*.

  (Grayson, 1988; Huynh, 1994; Hemker, Sijtsma, Molenaar, and Junker, 1996, 1997)

How to Check Monotonicity?

Checking monotonicity of the *item-total regression*

$$P[X_j = 1|S = s]$$

has been suggested or explored by

- Sijtsma (1988) — Dissertation on Mokken's IRT model.
- Ramsay (1991) — EDA for IRT using kernel-smoothed estimate of $P[X_j = 1|S = s]$.
- Croon (1991) — Constrains ordered latent class fits to make $P[X_j = 1|S = s]$ monotone.
- Scheiblechner (1995) — Manifest condition for ISOP model.

Junker (1993) suggests checking monotonicity of the *item-rest regression* ($S_{(-j)} = S - X_j$):

$$P[X_j = 1|S_{(-j)|}s]$$

Intuitively, one expects that conditioning on $S = s$ will make $P[X_j = 1|S = s]$ "too monotone".

Reason:

- We usually correct $Corr(X_j, S)$ by replacing S with the *rest-score*

  $$S_{(-j)} = S - X_j$$

  since $Corr(X_j, S)$ is expected to be too large.

Does this intuition really work?

Some positive results

Theorem (Junker, 1993, 1996; others?)

- Under the model (3), $P[X_j = 1 | S_{(-j)} = r]$ is nondecreasing in r.
- Under the Rasch model, $P[X_j = 1 | S = r]$ is also nondecr. in r.

Proof Sketch:

Let R stand for either S or $S_{(-j)}$. We want to show

$$\left.\begin{aligned}
&P[X_j = 1|R = r-1] \\
&= \int P[X_j = 1|R = r-1, \theta]\, f(\theta|R = r-1)d\theta \\
&\overset{?}{\leq} \int P[X_j = 1|R = r, \theta]\, f(\theta|R = r-1)d\theta \\
&\overset{?}{\leq} \int P[X_j = 1|R = r, \theta]\, f(\theta|R = r)d\theta \\
&= P[X_j = 1|R = r],
\end{aligned}\right\} \qquad (1)$$

- ***First*** "$\overset{?}{\leq}$": Sufficient to show
  - $P[X_j = 1|R = r, \theta] \nearrow$ in r for each θ
- ***Second*** "$\overset{?}{\leq}$": Sufficient to show
  - θ stochastically ordered by R;
  - $P[X_j = 1|R = r, \theta] \nearrow$ in θ for each r $\square$

Proof for $R = S_{(-j)}$, monotone IRT model (3):

$$P[X_j = 1 | S_{(-j)} = r - 1]$$

$$= \int P[X_j = 1 | S_{(-j)} = r - 1, \theta] \, f(\theta | S_{(-j)} = r - 1) d\theta$$

$$\overset{?}{\leq} \int P[X_j = 1 | S_{(-j)} = r, \theta] \; , f(\theta | S_{(-j)} = r - 1) d\theta$$

$$\overset{?}{\leq} \int P[X_j = 1 | S_{(-j)} = r, \theta] \, f(\theta | S_{(-j)} = r) d\theta$$

$$= P[X_j = 1 | S_{(-j)} = r],$$

- *First* "$\overset{?}{\leq}$":
  - By LI, $P[X_j = 1 | S_{(-j)} = r, \theta] = P_j(\theta)$, nondecreasing in r for each θ
- *Second* "$\overset{?}{\leq}$":
  - Grayson (1988); Huynh (1994): θ stochastically ordered by $S_{(-j)}$;
  - By LI, $P[X_j = 1 | S_{(-j)} = r, \theta] = P_j(\theta)$, nondecreasing in θ for each r $\square$

Proof for $R = S$, Rasch model:

$$P[X_j = 1|S = r-1]$$

$$= \int P[X_j = 1|S = r-1, \theta]\, f(\theta|S = r-1) d\theta$$

$$\stackrel{?}{\leq} \int P[X_j = 1|S = r, \theta]\, f(\theta|S = r-1) d\theta$$

$$\stackrel{?}{\leq} \int P[X_j = 1|S = r, \theta]\, f(\theta|S = r) d\theta$$

$$= P[X_j = 1|S = r],$$

- <u>*First*</u> "$\stackrel{?}{\leq}$":
  - *In general* $P[X_j = 1|S = r, \theta]$ nondecreasing in r for each θ (Scheiblechner 1995, Theorem 4)
- <u>*Second*</u> "$\stackrel{?}{\leq}$":
  - Grayson (1988); Huynh (1994): θ stochastically ordered by S;
  - Rasch Sufficiency: $P[X_j = 1|S = r, \theta] = P[X_j = 1|S = r]$, $\nearrow$ in θ for each r $\square$

$\underline{P[X_j = 1|S = s]}$ <u>**Outside the Rasch model:**</u>

<u>What can go wrong?</u>

$$
\begin{aligned}
& P[X_j = 1|S = r-1] \\
& = \int P[X_j = 1|S = r-1, \theta]\, f(\theta|S = r-1) d\theta \\
& \stackrel{?}{\leq} \int P[X_j = 1|S = r, \theta]\, f(\theta|S = r-1) d\theta \\
& \stackrel{?}{\leq} \int P[X_j = 1|S = r, \theta]\, f(\theta|S = r) d\theta \\
& = P[X_j = 1|S = r],
\end{aligned}
$$

- First "$\stackrel{?}{\leq}$" always true
- Second "$\stackrel{?}{\leq}$" fails when $P[X_j = 1|S = s, \theta]$ decreases in θ for fixed s.
- Failure must be large to offset first "$\stackrel{?}{\leq}$".

Note that our intuition based on item-total and item-rest correlations was backwards!

- *We expect* $Corr(X_j, S) > Corr(X_j, S_{(-j)})$ *<u>but</u>*

- $P[X_j = 1|S]$ *can be <u>nonmonotone</u>, whereas* $P[X_j = 1|S_{(-j)}]$ *<u>must</u> <u>be</u> <u>monotone</u>.*

Elementary Symmetric Functions

- Consider item subset $X_{k_1}, \ldots, X_{k_n}$ from $X_1, \ldots, X_J$
- Item Odds $\varepsilon_{k_j}(\theta) = P_{k_j}(\theta)/[1 - P_{k_j}(\theta)]$
- *Elementary symmetric function* for items $k_1, \ldots, k_n$, total score s, latent variable θ, and IRF's $\underline{P} = (P_1(\theta), \ldots, P_J(\theta))$, is

$$
\begin{aligned}
&\gamma_{(k_1,\ldots,k_n)}(s, \theta, \underline{P}) \\
&\quad = \sum_{s_{(k_1,\ldots,k_n)}=s} \prod_{j=1}^{n} \left[\frac{P_{k_j}(\theta)}{1 - P_{k_j}(\theta)}\right]^{x_{k_j}} \\
&\quad = \sum_{s_{(k_1,\ldots,k_n)}=s} \prod_{j=1}^{n} \varepsilon_{k_j}(\theta)^{x_{k_j}}
\end{aligned}
$$

  – Summation is over all score patterns such that $\sum_{j=1}^{n} X_{k_j} = s$

  – Also write $\gamma_{(k_1,\ldots,k_n)}(s, \theta, \underline{P}) \equiv \gamma_{(k_1,\ldots,k_n)}(s, \theta)$

Some properties of $\gamma_{(k_1,\ldots,k_n)}(s,\theta)$

- $$P[S=s \mid \theta] \;=\; \gamma_{(1,\ldots,J)}(s,\theta,\underline{P}) \cdot \prod_{j=1}^{J}(1-P_j(\theta))\ .$$

- Multiply by $f(\theta)$ and then divide by $P[S=s]$, get

$$\begin{aligned} f(\theta|S=s) &= \frac{P[S=s|\theta]f(\theta)}{P[S=s]} \\ &= \frac{\gamma_{(1,\ldots,J)}(s,\theta,\underline{P}) \cdot \prod_{j=1}^{J}(1-P_j(\theta))f(\theta)}{P[S=s]} \end{aligned}$$

- Integrate wrt. $f(\theta)$ and then multiply and divide by $P[S=0]$, get

$$P[S=s] \;=\; E[\gamma_{(1,\ldots,J)}(s,\Theta,\underline{P}) \mid S=0] \cdot P[S=0],$$

a version of Holland's (1990) *Dutch identity*.

- Standard identities generalize from Rasch (e.g. Fischer 1974):

$$\begin{aligned} \gamma_{(1,\ldots,J)}(s,\theta,\underline{P}) &= \varepsilon_1(\theta)\gamma_{(2,\ldots,J)}(s-1,\theta,\underline{P}) \\ &\quad + \gamma_{(2,\ldots,J)}(s,\theta,\underline{P}) \\ \frac{d}{d\varepsilon_1(\theta)}\gamma_{(1,\ldots,J)}(s,\theta,\underline{P}) &= \gamma_{(2,\ldots,J)}(s-1,\theta,\underline{P}) \end{aligned}$$

Item-total regressions

Note that

$$P[X_1 = 1 \mid S = s, \theta] = \frac{P[X_1 = 1, S = s \mid \theta]}{P[S_{(1,\ldots,J)} = s \mid \theta]}$$

$$= \frac{P[X_1 = 1, S = s - 1 \mid \theta]}{P[S = s \mid \theta]}$$

$$= \frac{\varepsilon_1(\theta)\gamma_{(2,\ldots,J)}(s-1, \theta, \underline{P}) \prod_{j=1}^{J}(1 - P_j(\theta))}{\gamma_{(1,\ldots,J)}(s, \theta, \underline{P}) \prod_{j=1}^{J}(1 - P_j(\theta))}$$

$$= \frac{\varepsilon_1(\theta)\gamma_{(2,\ldots,J)}(s-1, \theta, \underline{P})}{\varepsilon_1(\theta)\gamma_{(2,\ldots,J)}(s-1, \theta, \underline{P}) + \gamma_{(2,\ldots,J)}(s, \theta, \underline{P})}$$

$$= \left[1 + \frac{\gamma_{(2,\ldots,J)}(s, \theta, \underline{P})}{\varepsilon_1(\theta)\gamma_{(2,\ldots,J)}(s-1, \theta, \underline{P})}\right]^{-1}$$

and hence

$$P[X_1 = 1 \mid S = s, \theta] = \frac{1}{1 + \dfrac{\gamma_{(2,\ldots,J)}(s, \theta, \underline{P})}{\varepsilon_1(\theta)\gamma_{(2,\ldots,J)}(s-1, \theta, \underline{P})}} \tag{2}$$

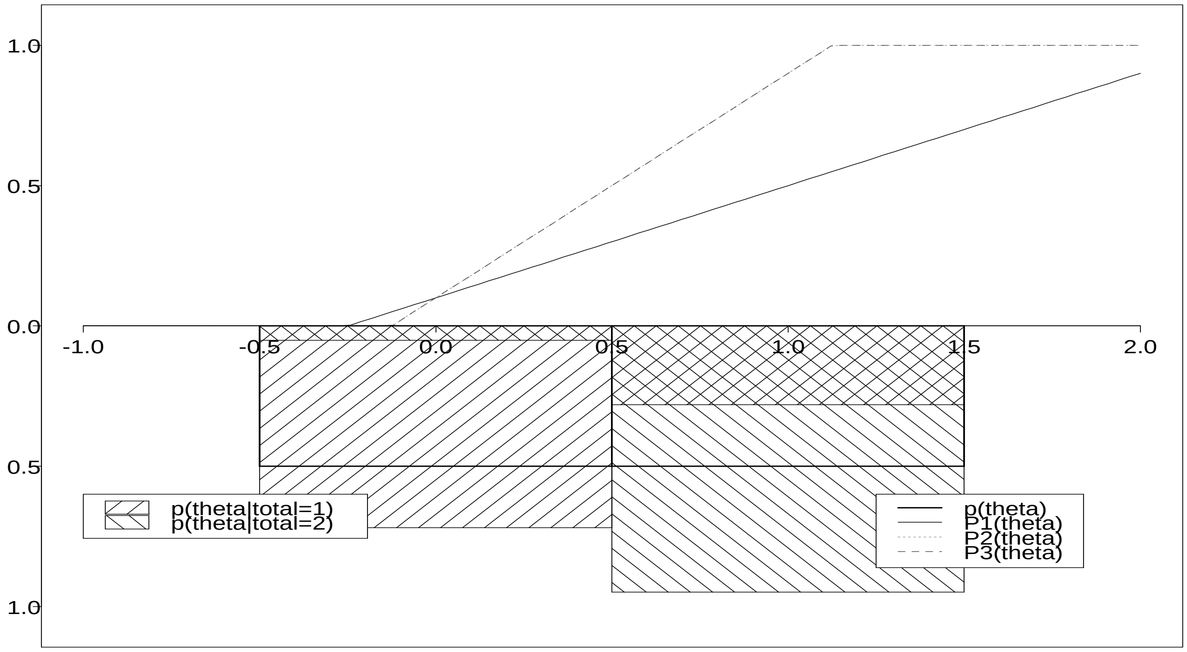
Snijders' Example (Junker, 1993)

- The model:
  - Two-point distribution for Θ, $P(\Theta = \theta_0) = P(\Theta = \theta_1) = 1/2$, where $\theta_0 < \theta_1$.
  - $P_1(\theta_0) = \delta$, $P_1(\theta_1) = 1/2$
  - $P_j(\theta_0) = \delta$, $P_j(\theta_1) = 1-\delta$ for $j = 2, 3, \ldots, J$.
- For $J = 3$, as $\delta \to 0$, $P[X_1 = 1|S = s]$ tends to the values 0, $\frac{1}{4}$, 0, and 1 for $s = 0, 1, 2, 3$, respectively.
- For all $J \geq 3$,

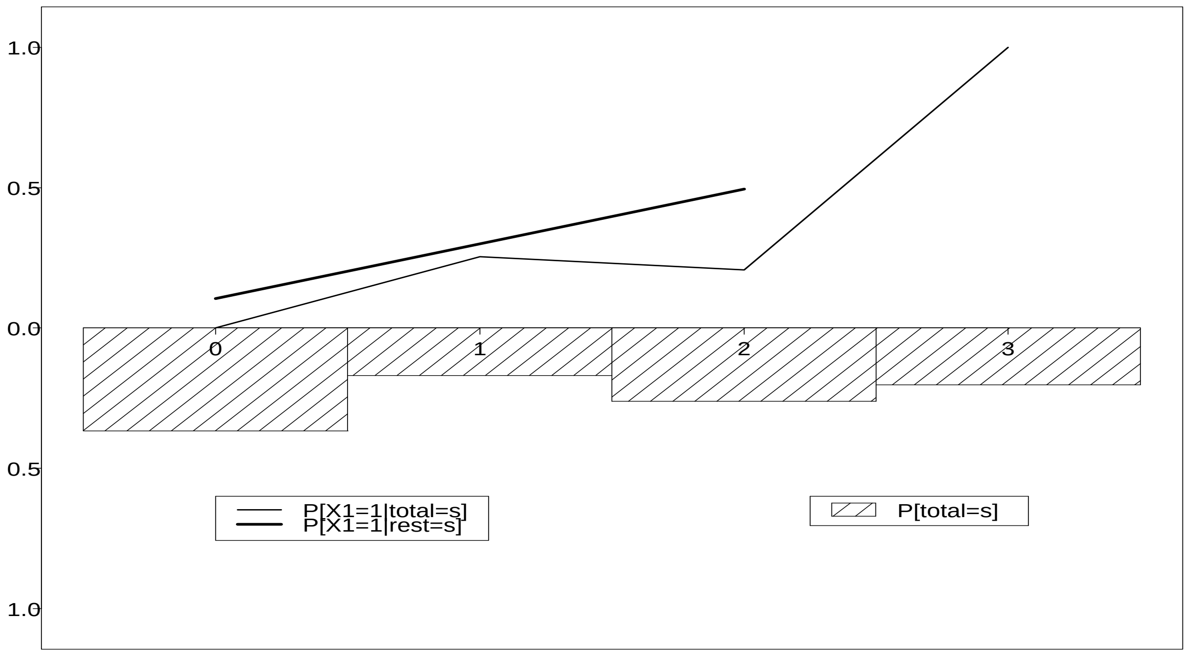
$$
\begin{aligned}
P[X_1|S = s, \theta_0] &= \left[1 + \frac{J-s}{s}\right]^{-1} \\
P[X_1|S = s, \theta_1] &= \left[1 + \frac{1-\delta}{\delta}\frac{J-s}{s}\right]^{-1}
\end{aligned}
$$

  which is *decreasing* in θ, since $\frac{1-\delta}{\delta} > 1$ for $0 < \delta < 1$.

Snijders: Latent structure



Snijders: Manifest structure



Item-rest Corr = 0.39 Item-total Corr = 0.69

The 2PL model

- The model:
  - $f(\theta)$ = anything reasonable.
  - $P_j(\theta) = \dfrac{e^{a_j(\theta-b_j)}}{1+e^{a_j(\theta-b_j)}}$, so $\varepsilon_j(\theta) = e^{a_j(\theta-b_j)} = e^{a_j\theta-\beta_j}$, where $\beta_j = a_j b_j$.

- Special case: $a_1 = a_1$ and $a_j = a_2$ for all $j = 2, \ldots, J$.

$$
\begin{aligned}
&P[X_1 = 1 \mid S = s, \theta] \\
&\quad = \left[1 + \frac{\gamma_{(2,\ldots,J)}(s, \theta, \underline{P})}{\varepsilon_1(\theta)\gamma_{(2,\ldots,J)}(s-1, \theta, \underline{P})}\right]^{-1} \\
&\quad = \left[1 + e^{(a_2-a_1)\theta} e^{\beta_1} \frac{\gamma_{(2,\ldots,J)}(s, -, \underline{\beta})}{\gamma_{(2,\ldots,J)}(s-1, -, \underline{\beta})}\right]^{-1},
\end{aligned}
$$

where $\gamma_{(2,\ldots,J)}(s, -, \underline{\beta}) = \sum_{s_{(-1)}=s} \prod_{j=2}^{n} e^{-\beta_j x_j}$

- $a_1 = a_2$ (Rasch) $\Longrightarrow P[X_1 = 1 \mid S = s, \theta]$ *const.* in θ.

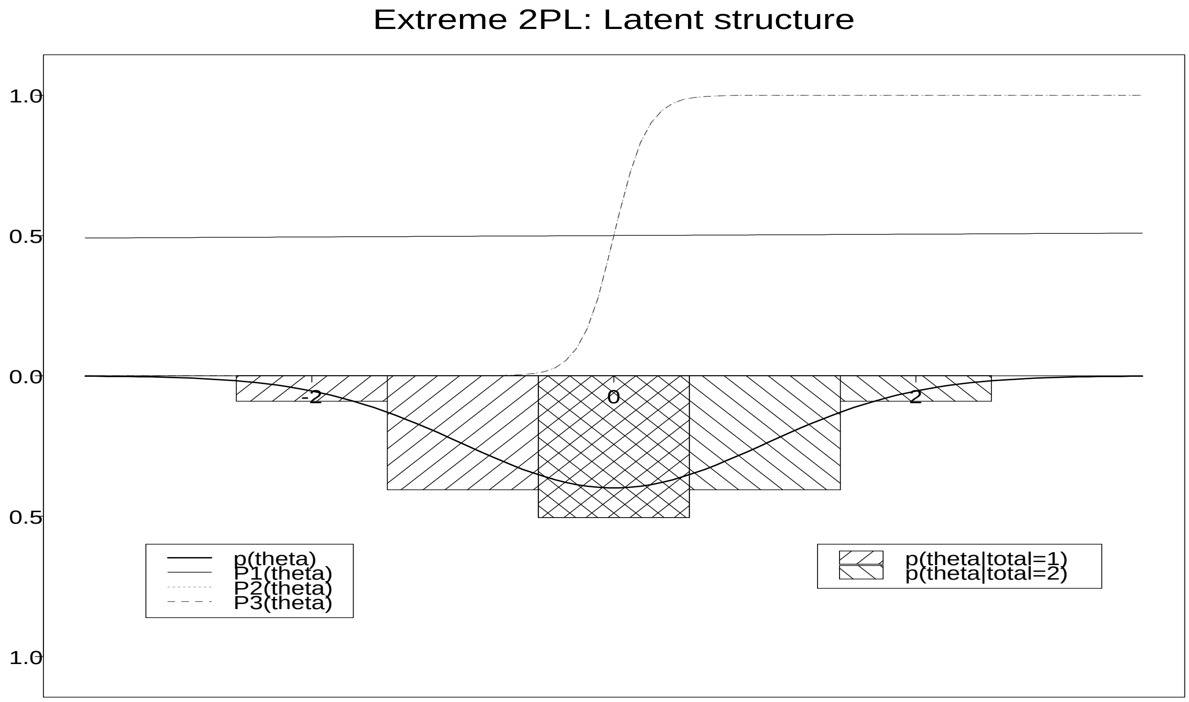
- $a_1 < a_2 \Longrightarrow P[X_1 = 1 \mid S = s, \theta]$ *decreasing* in θ

An extreme 2PL example

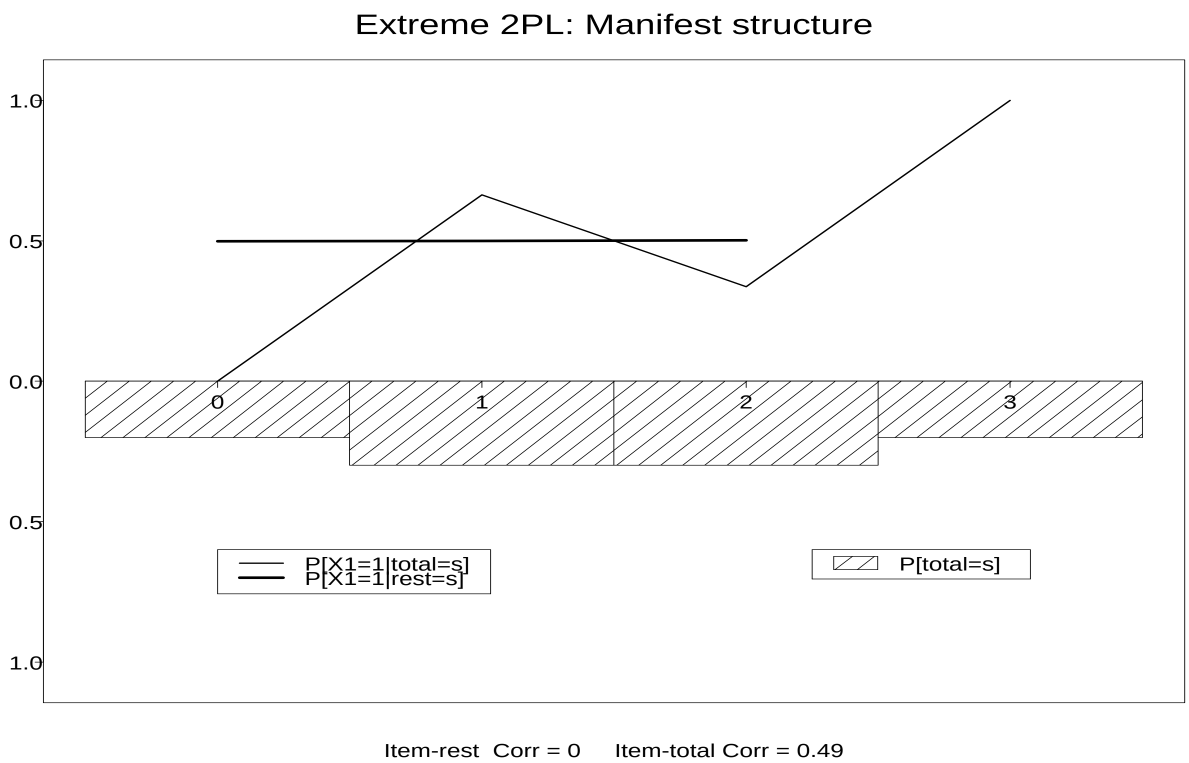
- $f(\theta)$ is symmetric, unimodal on -2, -1, 0, 1, 2.

- $j = 1$: $P_1(\theta) \equiv 1/2$

- $j > 1$: $P_j(\theta) = \dfrac{e^{a\theta}}{1 + e^{a\theta}}$ where $a = 9$

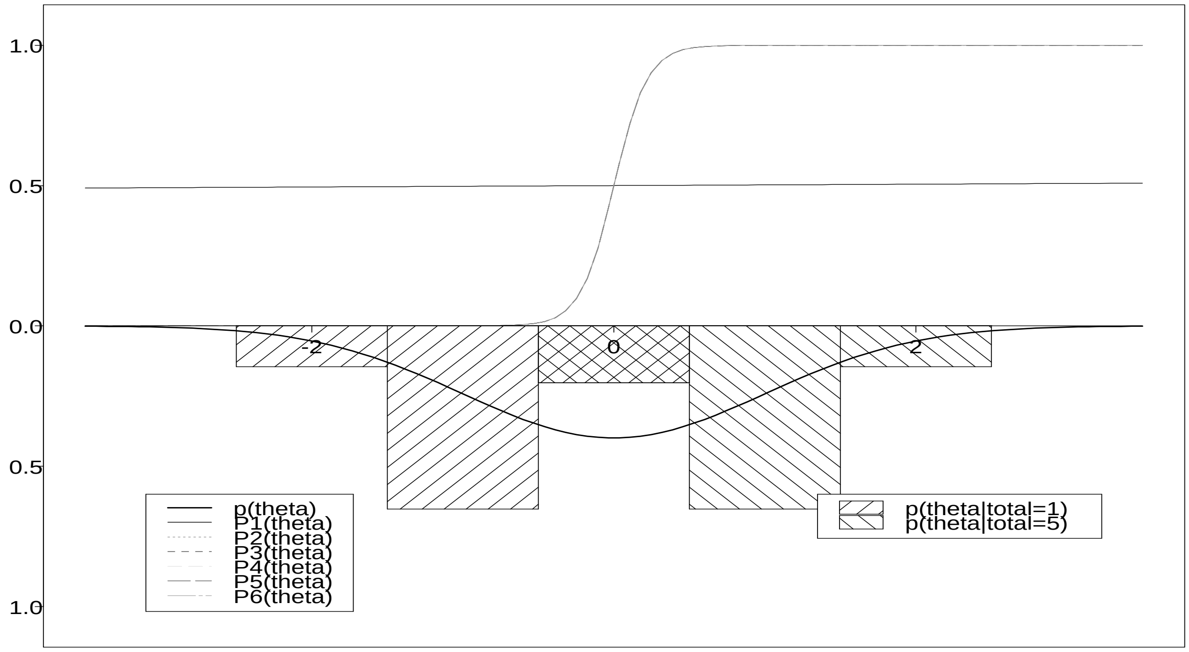


Extreme 2PL: Latent structure
1.0
0.5
0.0
0.5
1.0
-2
0
2
p(theta)
P1(theta)
P2(theta)
P3(theta)
p(theta|total=1)
p(theta|total=2)

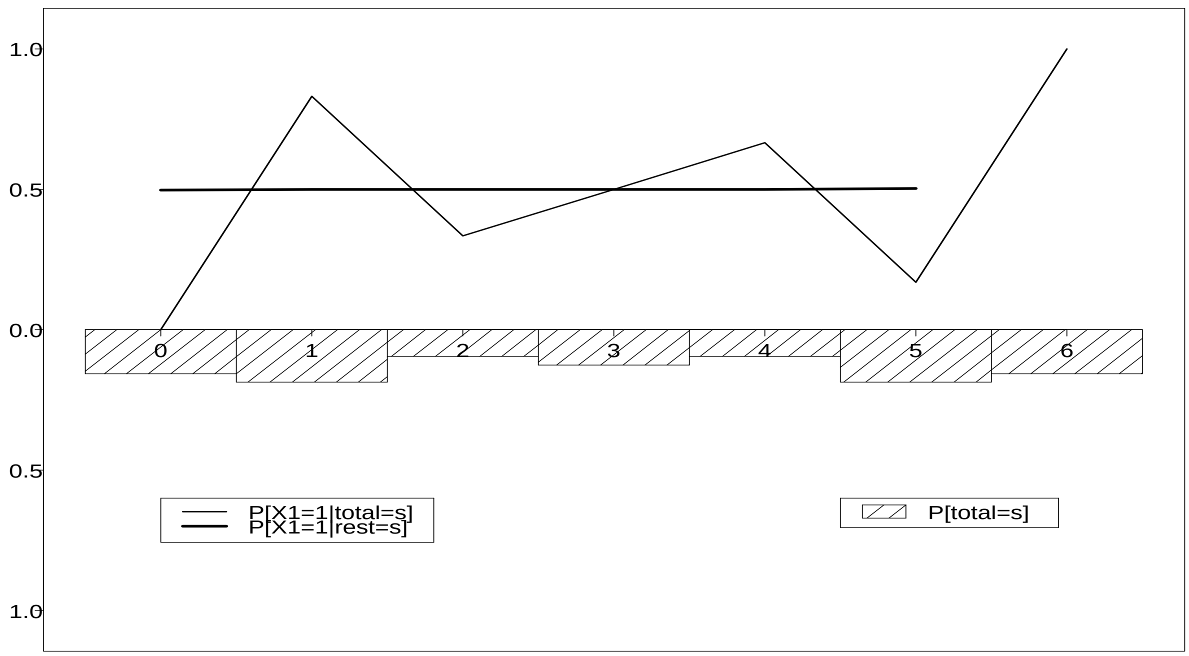


Extreme 2PL: Manifest structure
1.0
0.5
0.0
0.5
1.0
0
1
2
3
P[X1=1|total=s]
P[X1=1|rest=s]
P[total=s]
Item-rest Corr = 0 Item-total Corr = 0.49

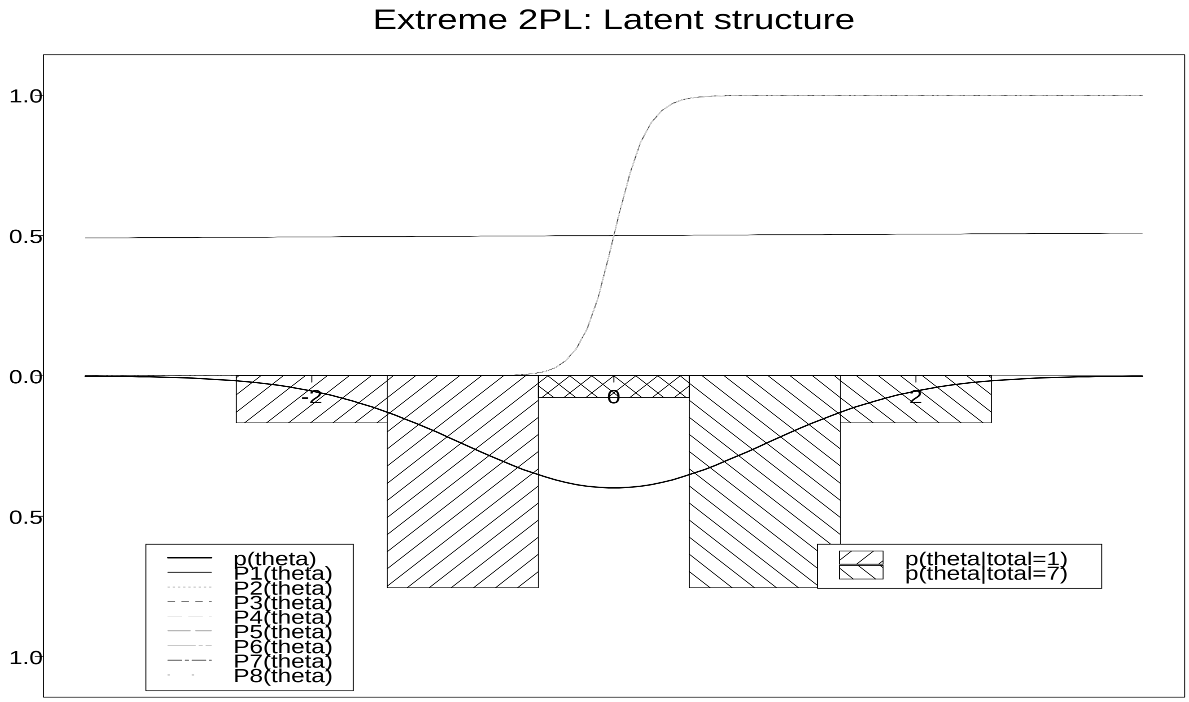
Extreme 2PL: Latent structure

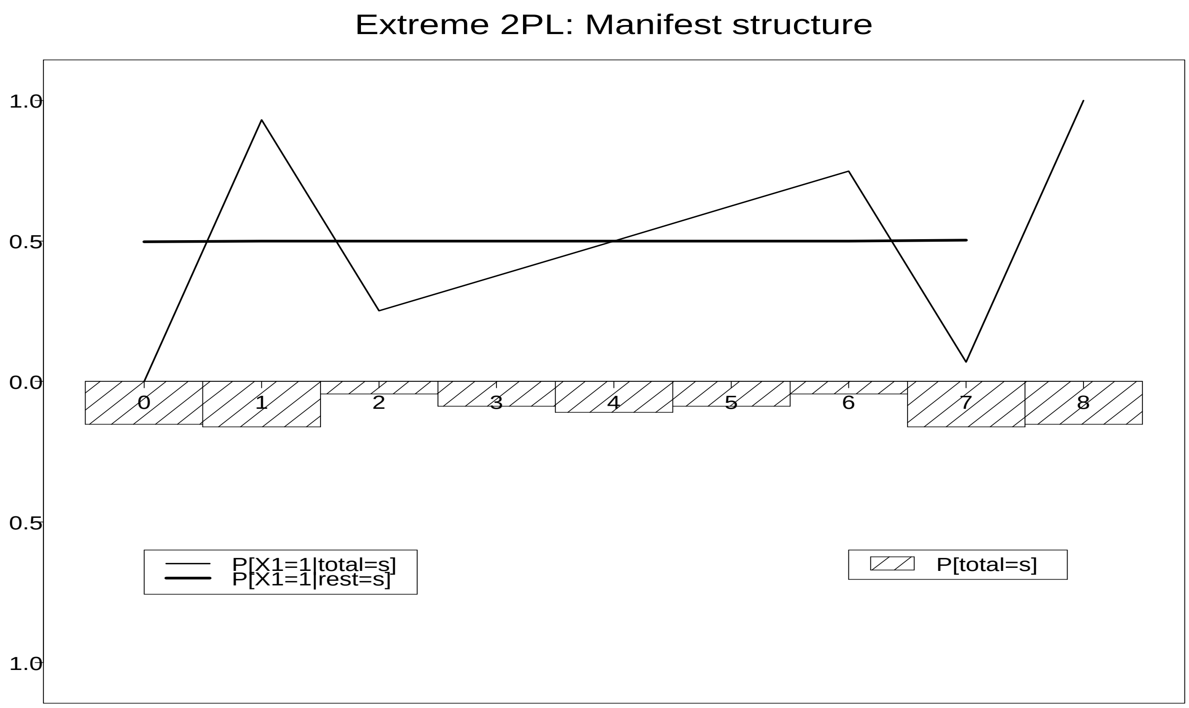


Extreme 2PL: Manifest structure



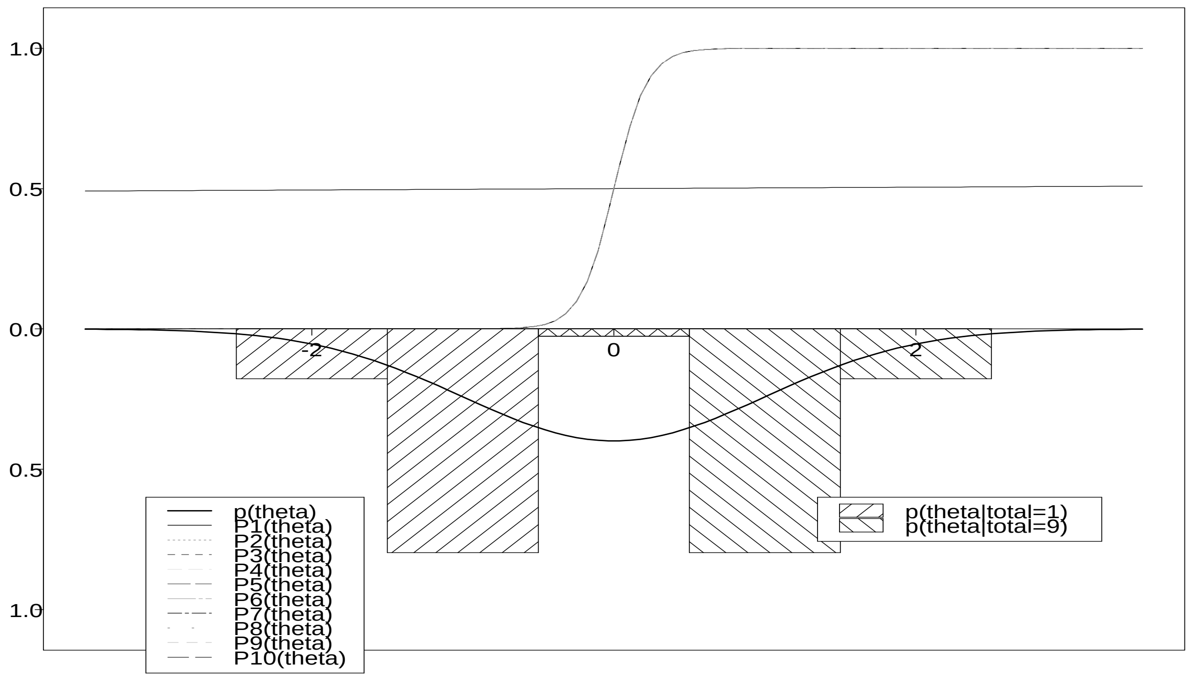
Item-rest Corr = 0 Item-total Corr = 0.24



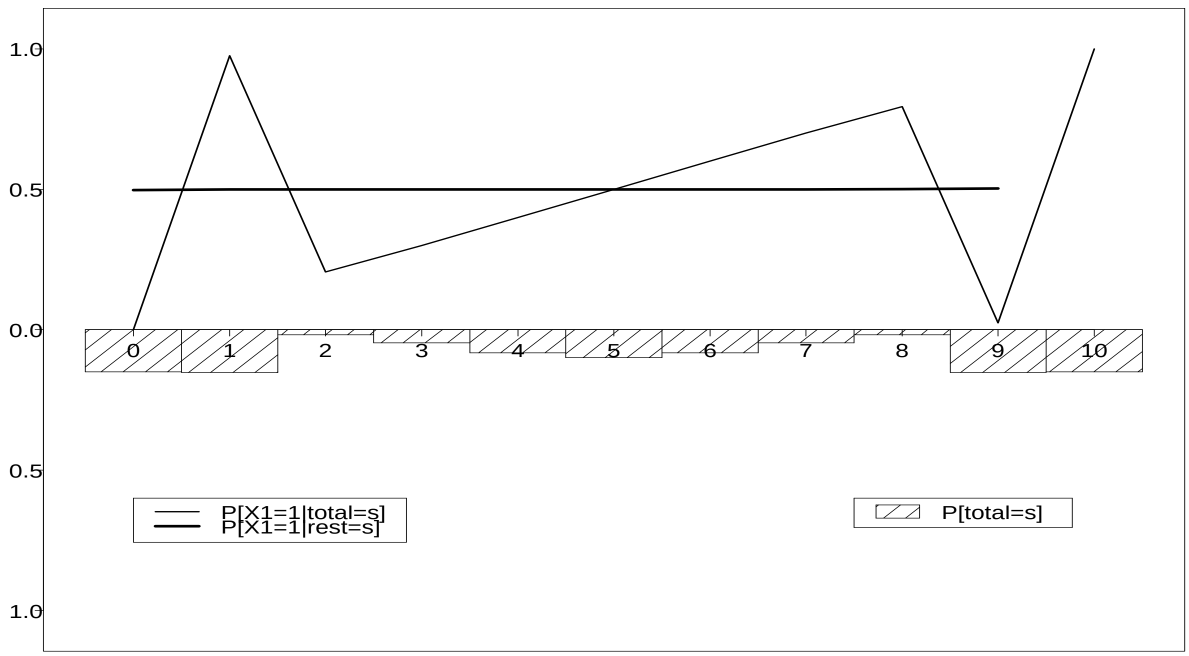


Item-rest Corr = 0 Item-total Corr = 0.18

Extreme 2PL: Latent structure



Extreme 2PL: Manifest structure



Item-rest Corr = 0 Item-total Corr = 0.14

2PL with varying difficulties

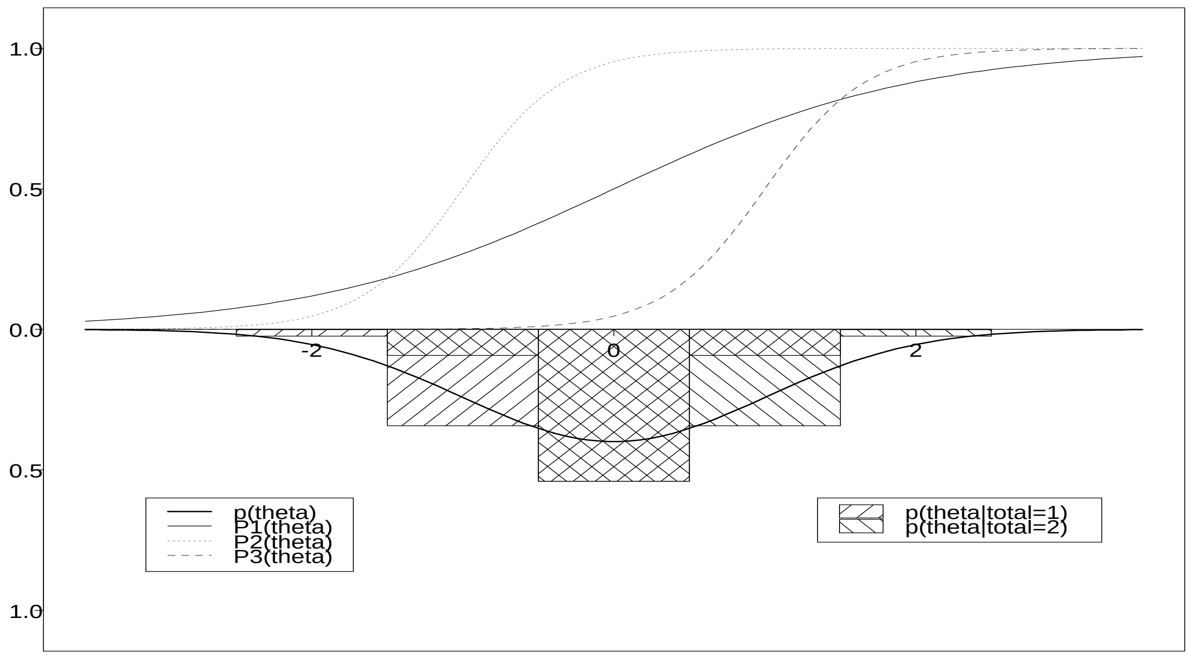
- $f(\theta)$ is symmetric, unimodal on -2, -1, 0, 1, 2.

- $j = 1$: $P_1(\theta)\dfrac{e^{\theta}}{1 + e^{\theta}}$

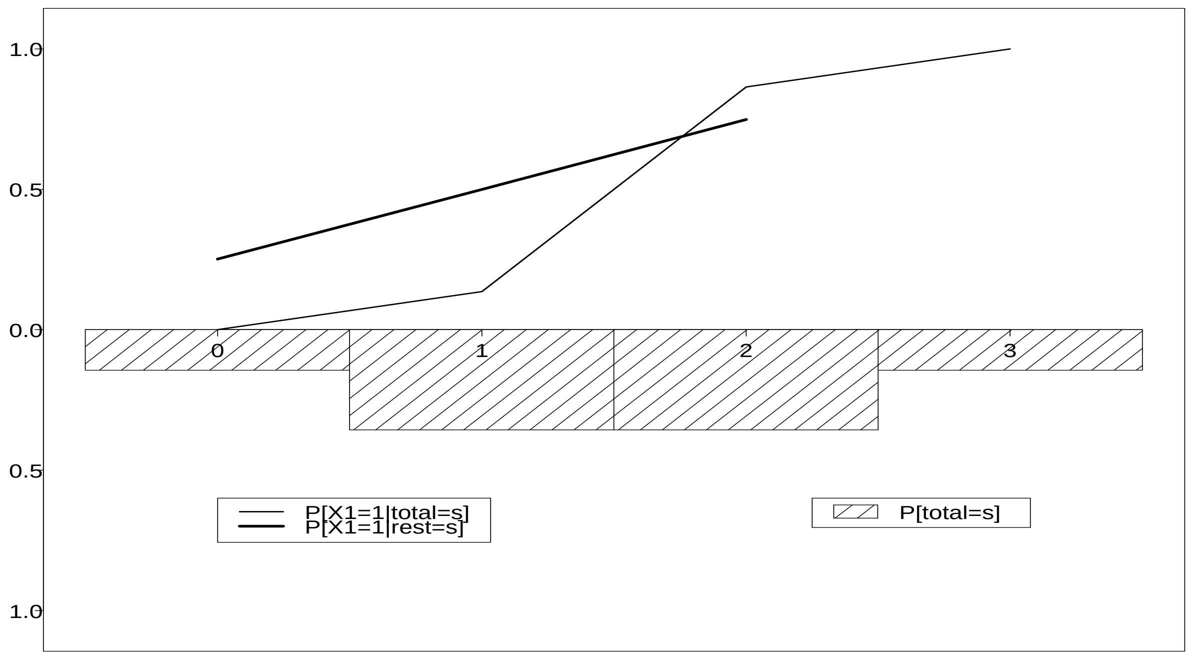
- $j > 1$: $P_j(\theta) = \dfrac{e^{a(\theta - b_j)}}{1 + e^{a(\theta - b_j)}}$,

  where $a = 3$ and b_j equally spaced from -1 to $+1$.

2PL: Latent structure

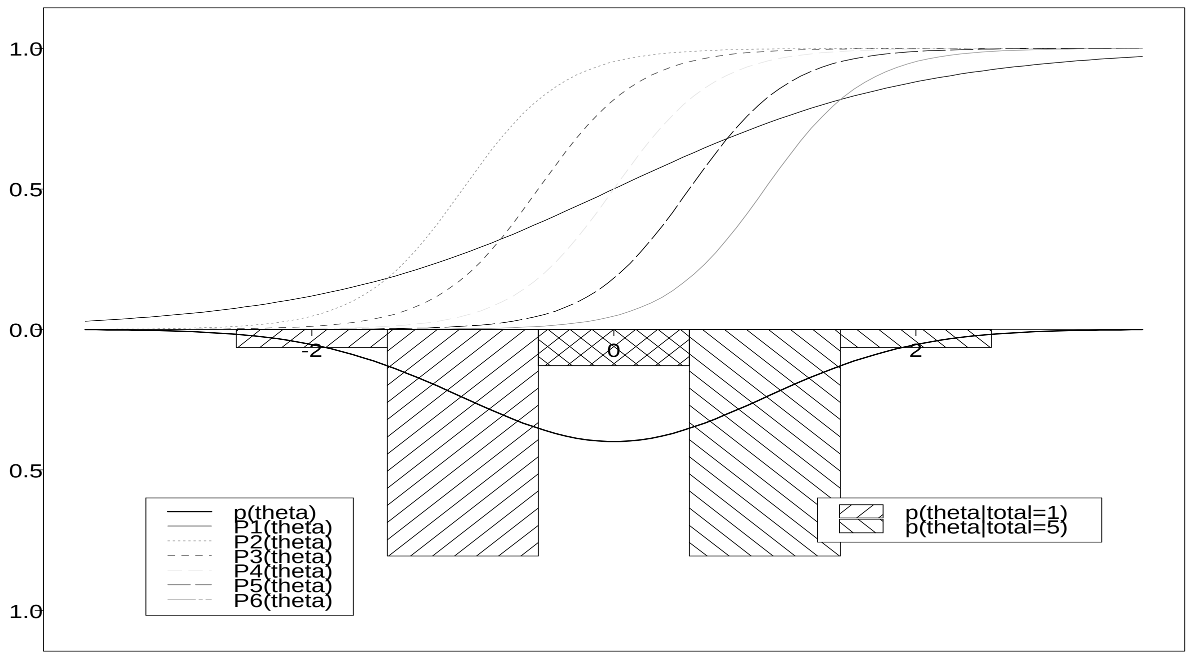


2PL: Manifest structure

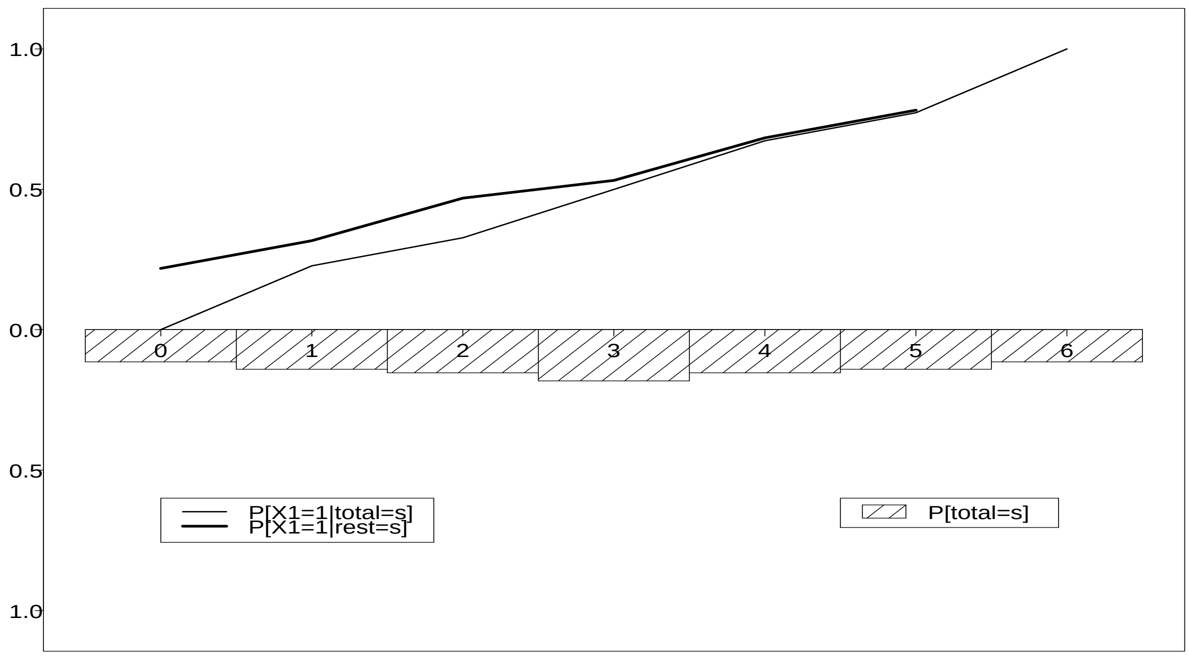


Item-rest Corr = 0.31 Item-total Corr = 0.76

2PL: Latent structure

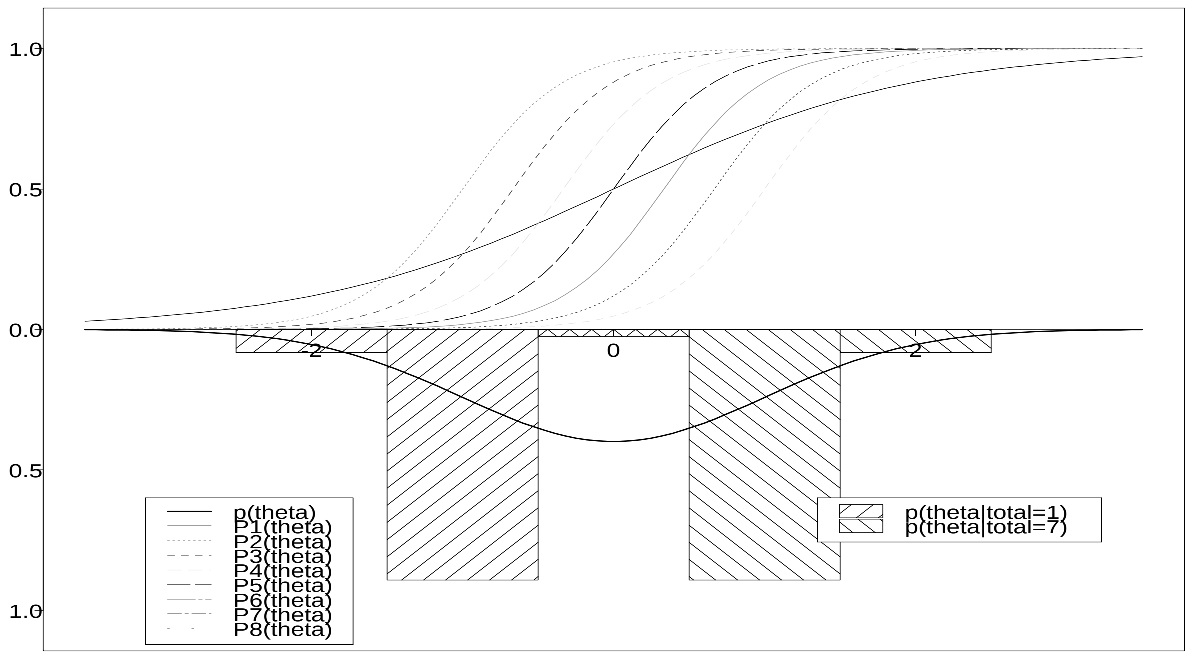


2PL: Manifest structure

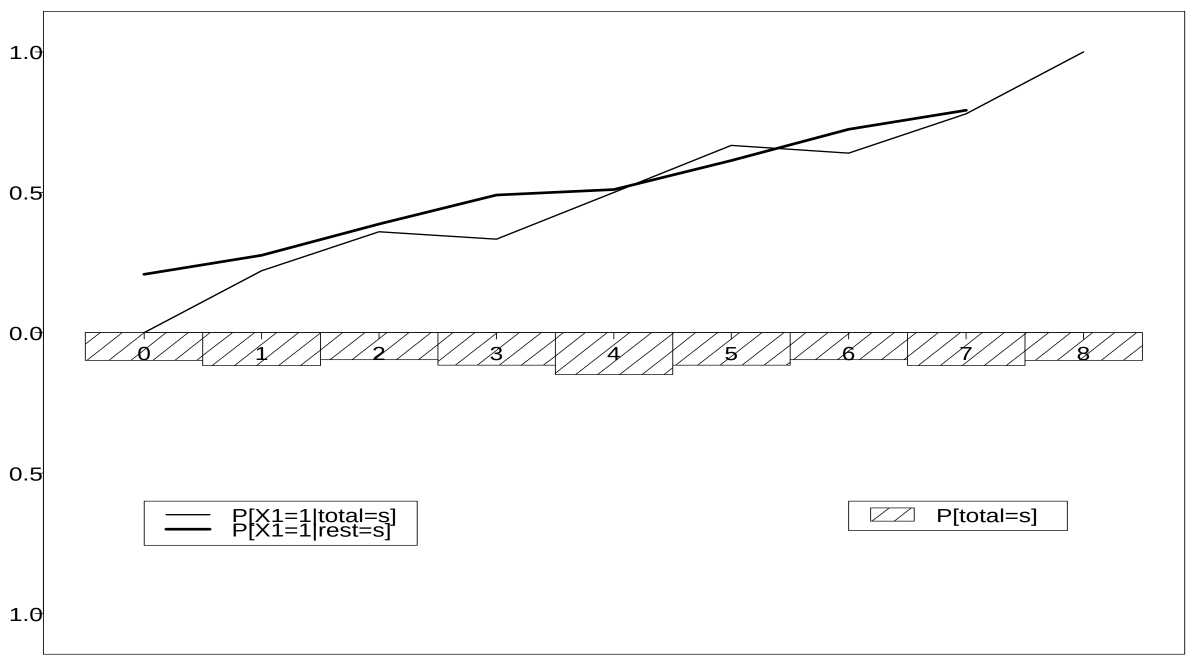


Item-rest Corr = 0.37 Item-total Corr = 0.59

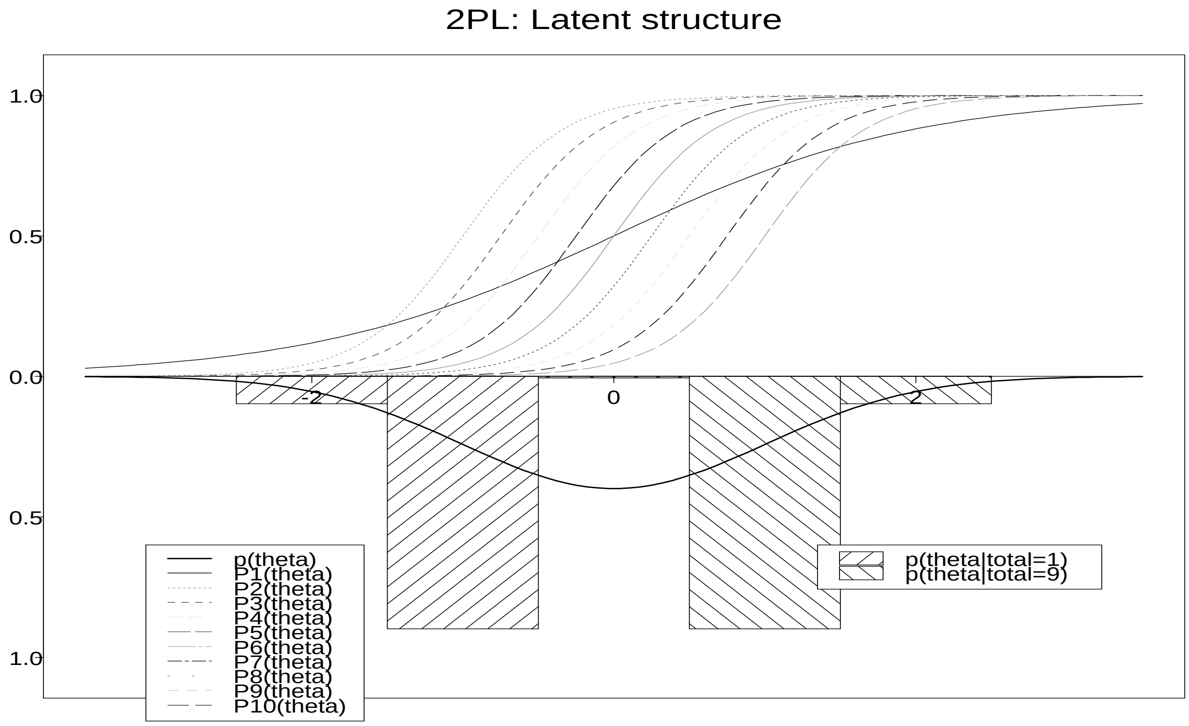
2PL: Latent structure

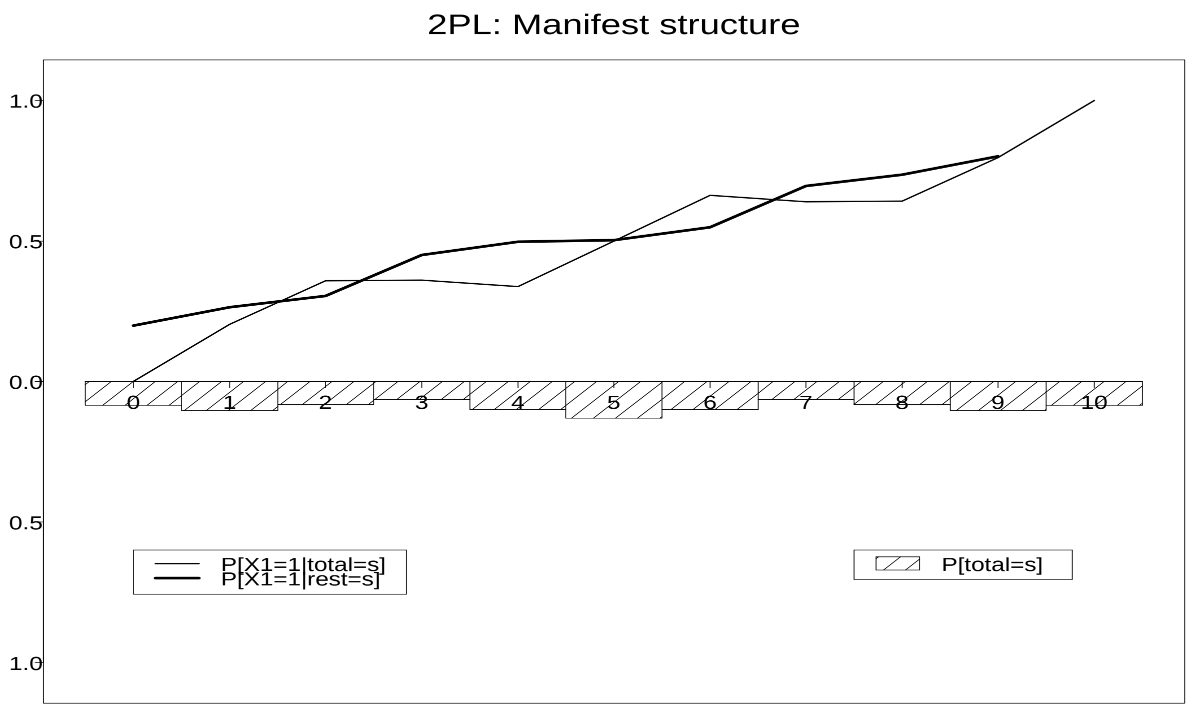


2PL: Manifest structure



Item-rest Corr = 0.38 Item-total Corr = 0.54





Item-rest Corr = 0.38 Item-total Corr = 0.52

Noncrossing IRF's

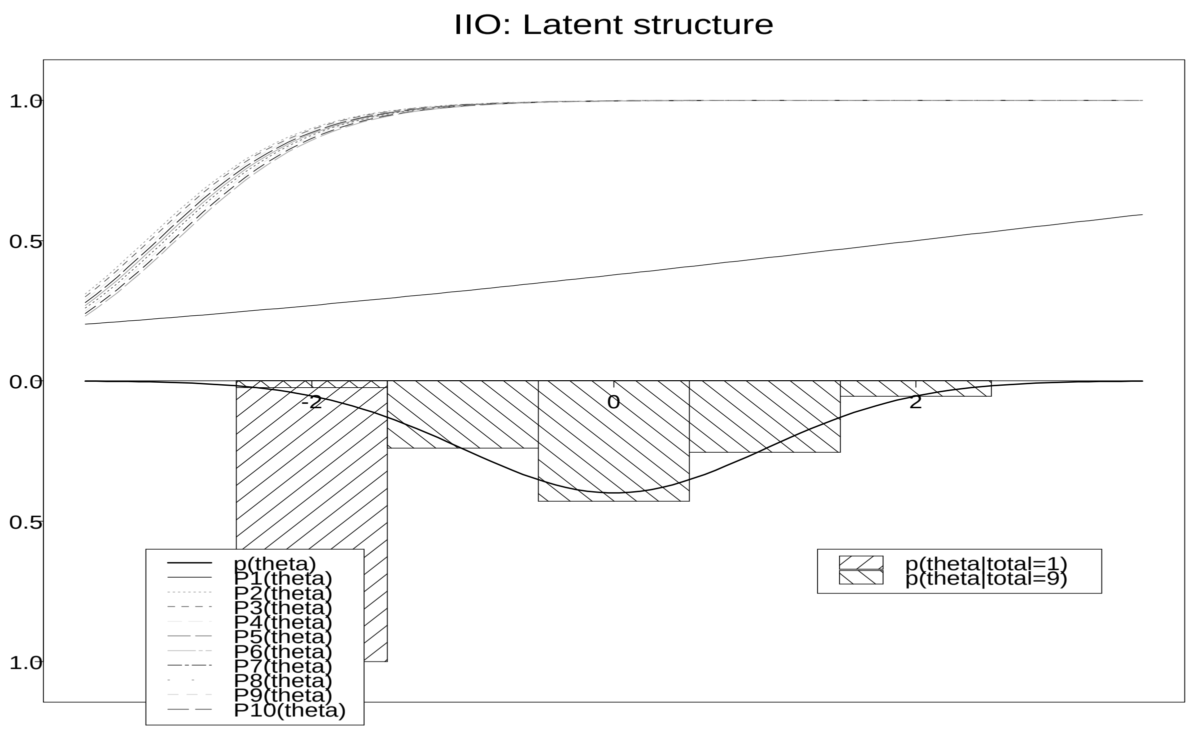
- $f(\theta)$ is symmetric, unimodal on -2, -1, 0, 1, 2.

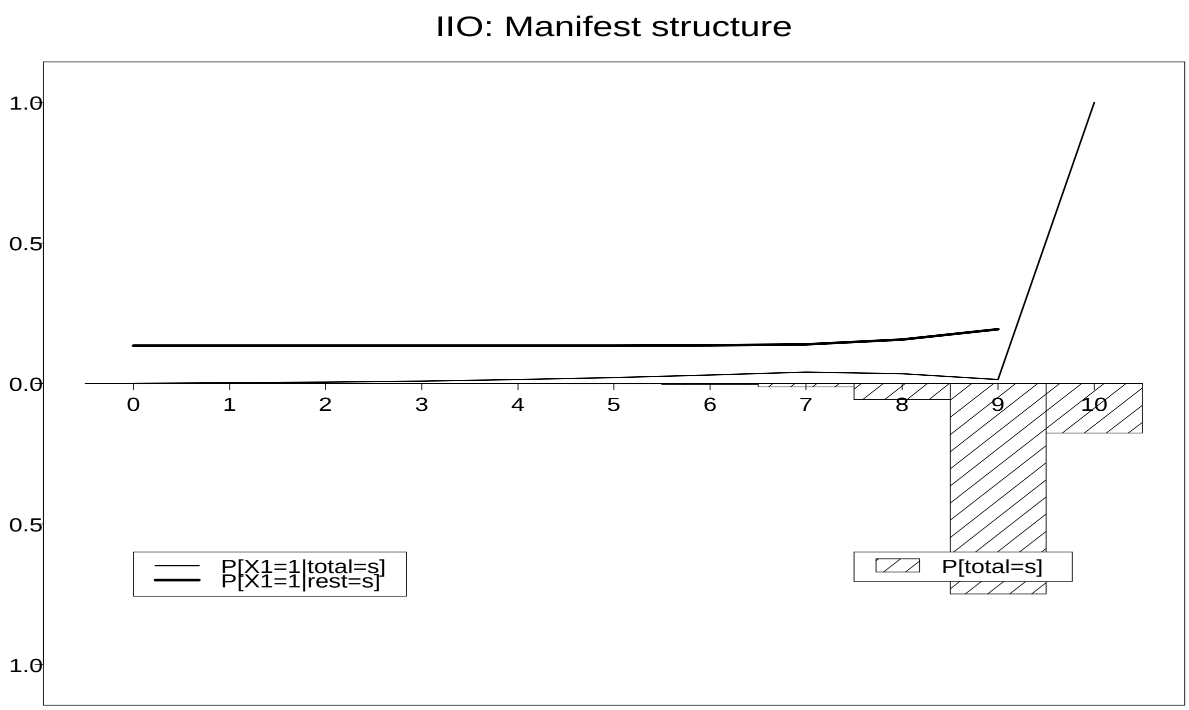
- $P_j(\theta) = \dfrac{e^{a_j(\theta - b_j)}}{1 + e^{a_j(\theta - b_j)}}$

- $j = 1$: $a_1 = 0.25$, $b_1 = 2$

- $j > 1$: $a_j = 2$, b_j equally spaced between -3.1 and -2.9

- Nonintersecting over the (finite) range of θ





Item-rest Corr = 0.03 Item-total Corr = 0.71

Conclusions

Under the monotone unidimensional IRT model for binary items. . .

- $P[X_j = 1|S_{(-j)} = s]$ is always monotone in s
- $P[X_j = 1|S = s]$ monotone in the Rasch model
- Outside the Rasch model $P[X_j = 1|S = s]$ need not be monotone
  - Snijders' counter-example
  - Easy to construct 2PL and related counter-examples
- The nonmonotonicities clearly can get worse as J grows.

- There are conflicting and confusing traditions about using $S_{(-j)}$:
  - Stout (1990) and standard MLE theory suggest that estimators based on S are, at least, pointwise consistent for θ;
  - Clarke & Ghosh (1995): $f(\theta|X_+)$ tightens up around true θ as J grows.
  - Ananstasi (1988), Sijtsma (1988), Ramsay (1991) advocate regressing on S to check item monotonicity
  - Croon (1991) and Scheiblechner (1998) suggest making monotonicity of $P[X_j = 1|S]$ part of the modeling assumptions
  - Lord and Novick (1968), Cureton (1966) and Wolfe (1967) advocate $S_{(-j)}$
  - Lord (1965) uses item-rest regressions but when they are reproduced in Lord (1980) they are labelled item-total regressions
  - Wainer (1983) and Wainer, Wadkins and Rogers (1984) employ item-rest regressions (correctly!) to explore methods of identifying incorrectly-keyed items.

- We think it is clear from our examples that one should use $S_{(-j)}$.
- Thissen and Orlando (1997) develop mosaic plot diagnostics that can be used with either S or $S_{(-j)}$, and point out that 3PL guessing cannot be by regressing on S.