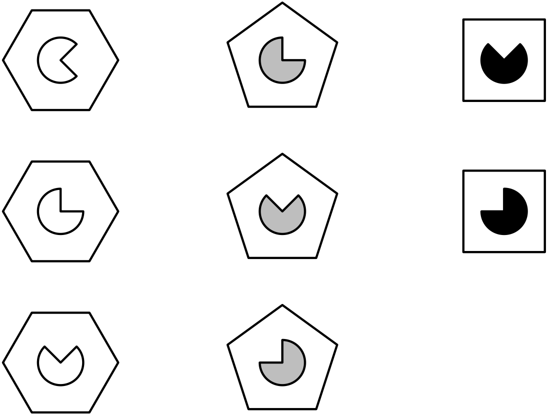
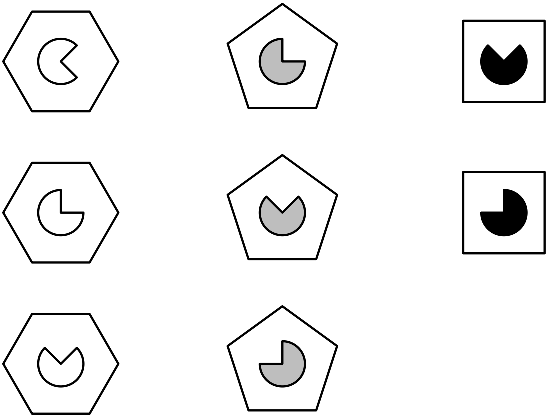
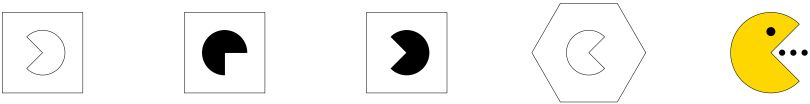
Every response matters:

Nested logit item response theory models for the development of smart item selection algorithms

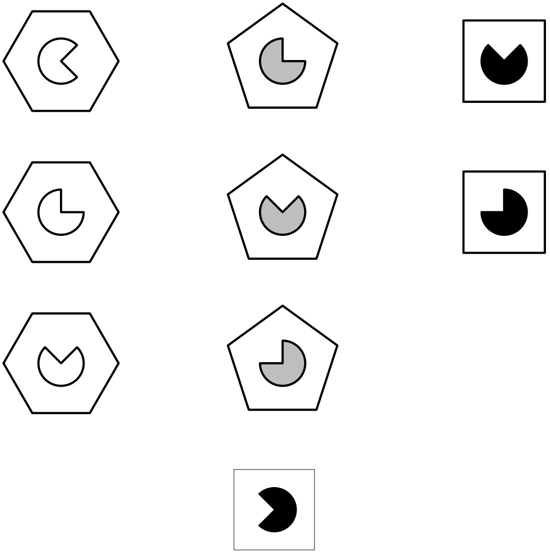
Ottavia M. Epifania[1,2], Andrea Brancaccio[3]
[1] University of Trento (IT), [2] Psicostat (IT)
[3] University of East Anglia (UK)







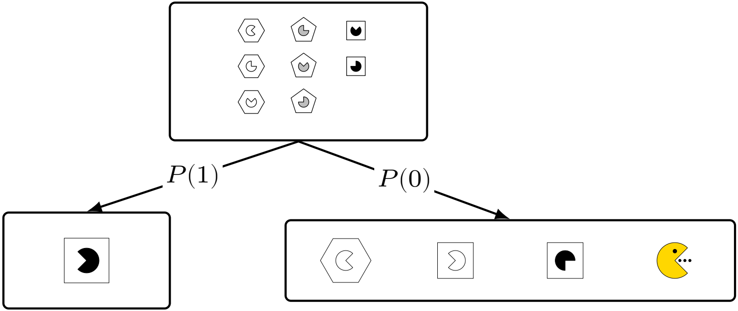
(a) (b) (c) (d) (e)



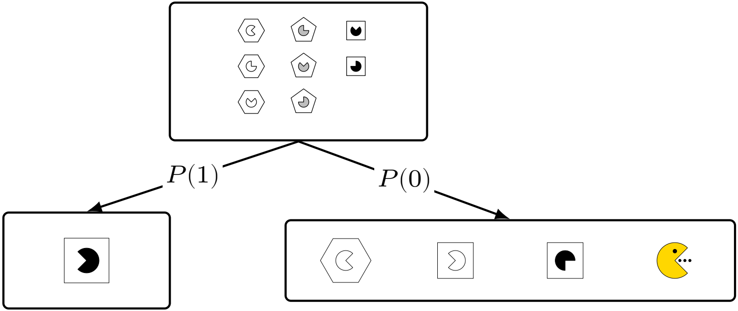
(c)

Modeling

The strategy



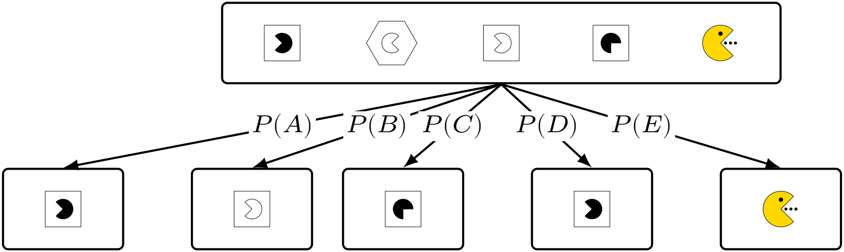
The strategy



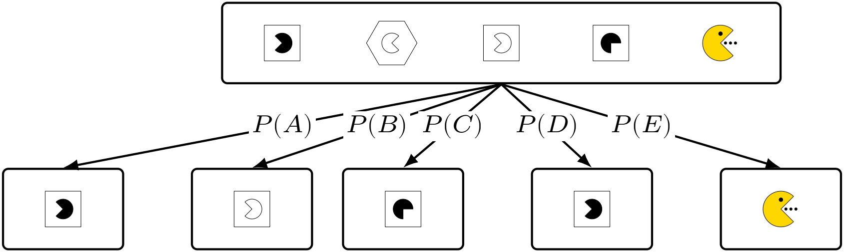
Item Response Theory models for dichotomous responses:

$$P(x_{pi} = 1|\theta_p) = c_i + (1 - c_i)\left[\frac{1}{1 + \exp^{-(b_i + a_i\theta_p)}}\right]$$

The strategy



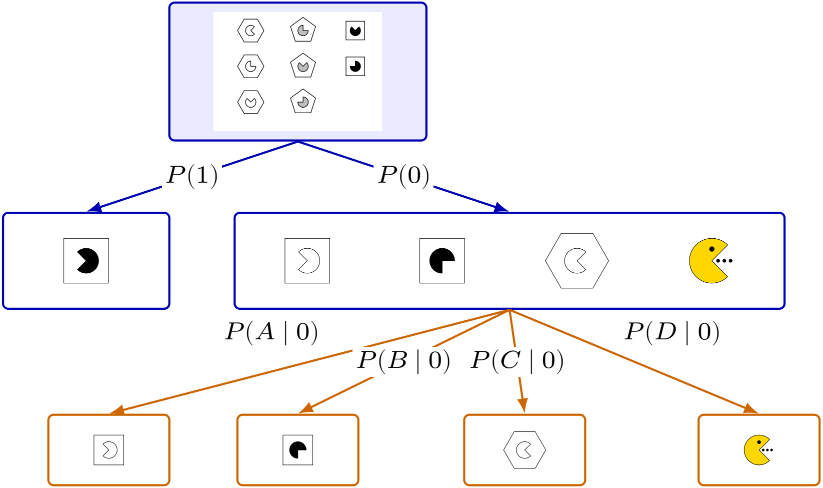
The strategy



Nominal Response Model (NRM, Bock, 1972), where each item i presents $v = 1, \dots, m_i$ *response categories*:

$$(P_{pv} = v|\theta_p) = \frac{\exp Z_{iv}(\theta_p)}{\sum_{k=1}^{m_i} \exp^{Z_{ik}(\theta_j)}}, \qquad Z_{iv}(\theta_p) = \zeta_{iv} + \lambda_{iv}(\theta_p)$$

The strategy



Level 1:

$$\boxed{P(x_{pi} = 1 \mid \theta_p) = c_i + (1 - c_i)\left[\frac{1}{1 + \exp^{-(b_i + a_i\theta_p)}}\right]}$$

Level 2:

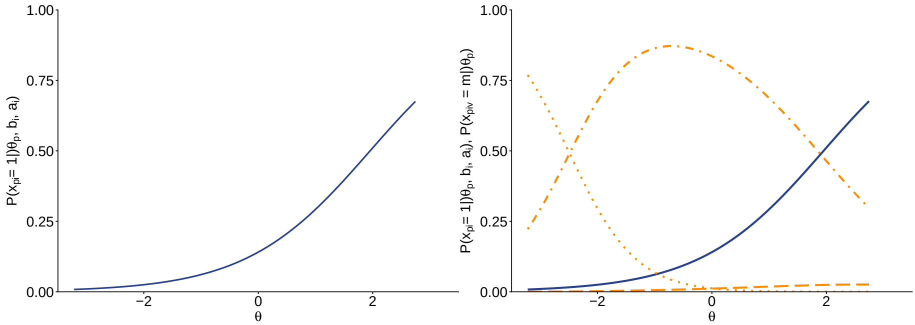
Each item i has $v_1 = 1, \ldots, m_i$ *distractors* $+1$ correct response:

$$D_{ipv} = \begin{cases} 1 & \text{if } x_{pi} = v, \\ 0 & \text{otherwise.} \end{cases}$$

$$\boxed{\begin{aligned} P(x_{ip} = 1,\, D_{ipv} = 1 \mid \theta_p) &= P(x_{ip} = 0 \mid \theta_p) P(D_{ipv} \mid x_{pi} = 0, \theta_p) \\ &= P(x_{pi} = 0 \mid \theta_p)\left[\frac{\exp^{Z_{iv}(\theta_p)}}{\sum_{k=1}^{m_i} \exp^{Z_{ik}(\theta_p)}}\right] \end{aligned}}$$

$Z_{iv}(\theta_p) = \zeta_{iv} + \lambda_{iv}(\theta_p)$

Suh, Y., & Bolt, D. M. (2010). Nested logit models for multiple-choice item response data. *Psychometrika, 75*(3), 454–473. https://doi.org/10.1007/s11336-010-9163-7



(a) 2-PLM (b) 2-PLNM

Figure 1: Comparison between 2Parameter Logistic Model (2-PLM), and the Nested Logistic Model (NLM, Suh & Bolt, 2009)

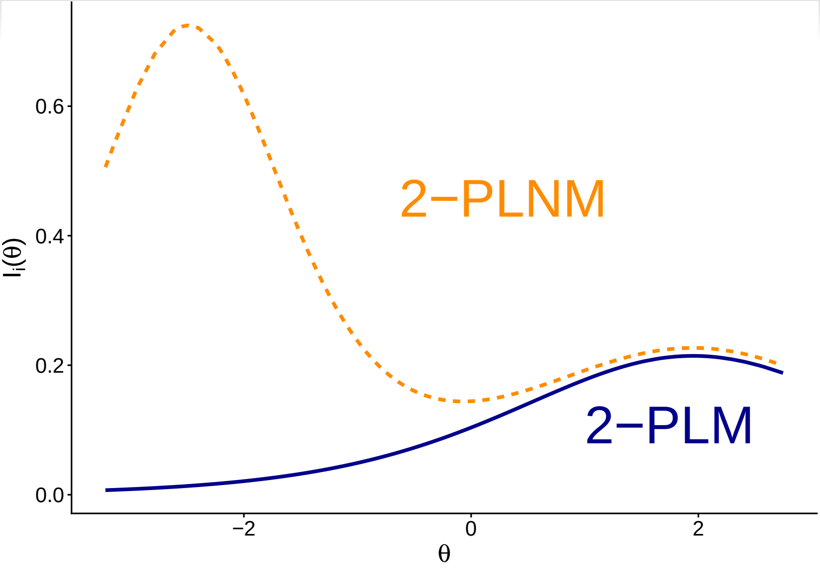


Figure 2: Information Functions

Broader aim of the project

Develop a CAT algorithm able to capitalize on the information provided by the incorrect responses: They are not just "incorrect", they have a meaning and they are informative

- More efficient: Less administered items but with a *at least comparable* level of estimation accuracy of the latent trait

This study

Estimation precision provided by the nested models under different conditions, focusing on the low levels of the latent trait:

Is it really useful to consider the information provided by the distractors?

Is it detrimental to consider the information provided by the distractor when there is actually no information?

Simulation study

Simulation study

Design

$m_1 = 3$ +1 correct response ($b_i \sim \mathcal{U}(-2.5, 2.5)$, $a_i \sim \mathcal{U}(.90, 1.70)$).

Simulation	ζ_1	ζ_2	λ_1	λ_2	θ
1	$\mathcal{U}(-2.5, 2.5)$	$\mathcal{U}(-2.5, 2.5)$	$\mathcal{U}(-2, 2)$	$\mathcal{U}(-2, 2)$	$\mathcal{N}(0, 1)$
2	$\mathcal{U}(-2.5, 2.5)$	$\mathcal{U}(-2.5, 2.5)$	$\mathcal{U}(-2, 0)$	$\mathcal{U}(-2, 0)$	$\mathcal{N}(0, 1)$
3	$\mathcal{U}(0, 2.5)$	$\mathcal{U}(0, 2.5)$	$\mathcal{U}(-2, 0)$	$\mathcal{U}(-2, 0)$	$\mathcal{N}(0, 1)$
4	$\mathcal{U}(-2.5, 2.5)$	$\mathcal{U}(-2.5, 2.5)$	$\mathcal{U}(-2, 2)$	$\mathcal{U}(-2, 2)$	$\mathcal{N}(0, 3)$
5	$\mathcal{U}(-2.5, 2.5)$	$\mathcal{U}(-2.5, 2.5)$	$\mathcal{U}(-2, 0)$	$\mathcal{U}(-2, 0)$	$\mathcal{N}(0, 3)$
6	$\mathcal{U}(0, 2.5)$	$\mathcal{U}(0, 2.5)$	$\mathcal{U}(-2, 0)$	$\mathcal{U}(-2, 0)$	$\mathcal{N}(0, 3)$

Note: $\sum_{k=1}^{m_i} \zeta_{ki} = 0$, $\sum_{k=1}^{m_i} \lambda_{ki} = 0$

$I = 10, 40$

Items with informative distractors

0% **25%** **50%** **75%** **100%**

$m_1 = 3$ +1 correct response ($b_i \sim \mathcal{U}(-2.5, 2.5)$, $a_i \sim \mathcal{U}(.90, 1.70)$).

Simulation	ζ_1	ζ_2	λ_1	λ_2	θ
1	$\mathcal{U}(-2.5, 2.5)$	$\mathcal{U}(-2.5, 2.5)$	$\mathcal{U}(-2, 2)$	$\mathcal{U}(-2, 2)$	$\mathcal{N}(0, 1)$
2	$\mathcal{U}(-2.5, 2.5)$	$\mathcal{U}(-2.5, 2.5)$	$\mathcal{U}(-2, 0)$	$\mathcal{U}(-2, 0)$	$\mathcal{N}(0, 1)$
3	$\mathcal{U}(0, 2.5)$	$\mathcal{U}(0, 2.5)$	$\mathcal{U}(-2, 0)$	$\mathcal{U}(-2, 0)$	$\mathcal{N}(0, 1)$
4	$\mathcal{U}(-2.5, 2.5)$	$\mathcal{U}(-2.5, 2.5)$	$\mathcal{U}(-2, 2)$	$\mathcal{U}(-2, 2)$	$\mathcal{N}(0, 3)$
5	$\mathcal{U}(-2.5, 2.5)$	$\mathcal{U}(-2.5, 2.5)$	$\mathcal{U}(-2, 0)$	$\mathcal{U}(-2, 0)$	$\mathcal{N}(0, 3)$
6	$\mathcal{U}(0, 2.5)$	$\mathcal{U}(0, 2.5)$	$\mathcal{U}(-2, 0)$	$\mathcal{U}(-2, 0)$	$\mathcal{N}(0, 3)$

Note: $\sum_{k=1}^{m_i} \zeta_{ki} = 0$, $\sum_{k=1}^{m_i} \lambda_{ki} = 0$

$I = 10, 40$

Items with informative distractors

0% **25%** **50%** **75%** **100%**

Full model

For all item

$Z_{iv}(\theta_p) = \zeta_{iv} + \lambda_{iv}(\theta_p)$

with $\lambda_{iv} \neq 0$

Restricted model on i

Target item i

$Z_{iv}(\theta_p) = \zeta_{iv} + \lambda_{iv}(\theta_p)$

with $\lambda_{iv} = 0$

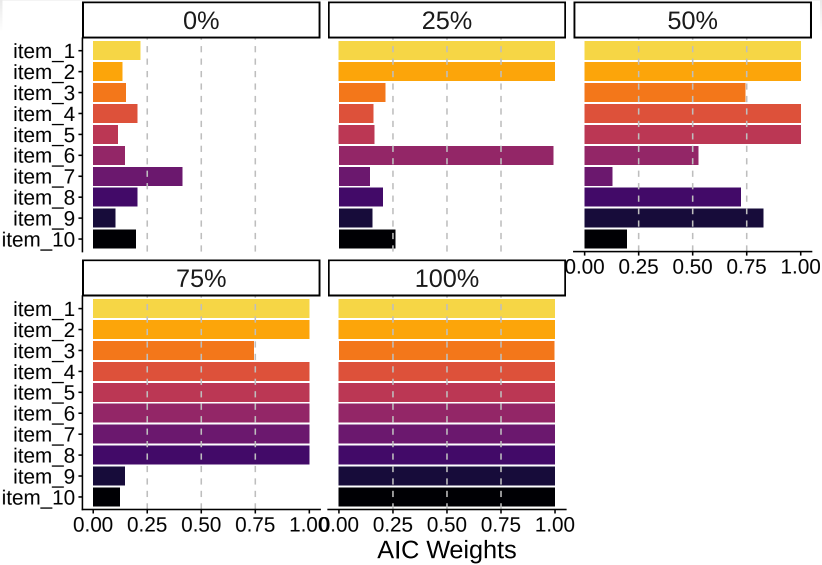
$$w_{\text{Full},i} = \frac{1}{1 + \exp\left[-\frac{AIC_{\text{restricted},i} - AIC_{\text{Full}}}{2}\right]}$$

$$\mathrm{Bias}(\hat{\theta}) = \frac{1}{P}\sum_{p=1}^{P}\left(\hat{\theta}_p - \theta_p\right), \qquad \mathrm{RMSE}(\hat{\theta}) = \sqrt{\frac{1}{P}\sum_{p=1}^{P}\left(\hat{\theta}_p - \theta_p\right)^2}$$

$$TIF(\theta_p) = \sum_{i=1}^{I}\left[\boxed{P_{pi}I_{ix}(\theta_j)} + \boxed{\sum_{v=1}^{m_i} P_{piv}I_{iv}(\theta_j)}\right]$$

Simulation study

Results ($I = 10$)


0%
25%
50%
75%
100%
item_1
item_2
item_3
item_4
item_5
item_6
item_7
item_8
item_9
item_10
0.00
0.25
0.50
0.75
1.00
AIC Weights

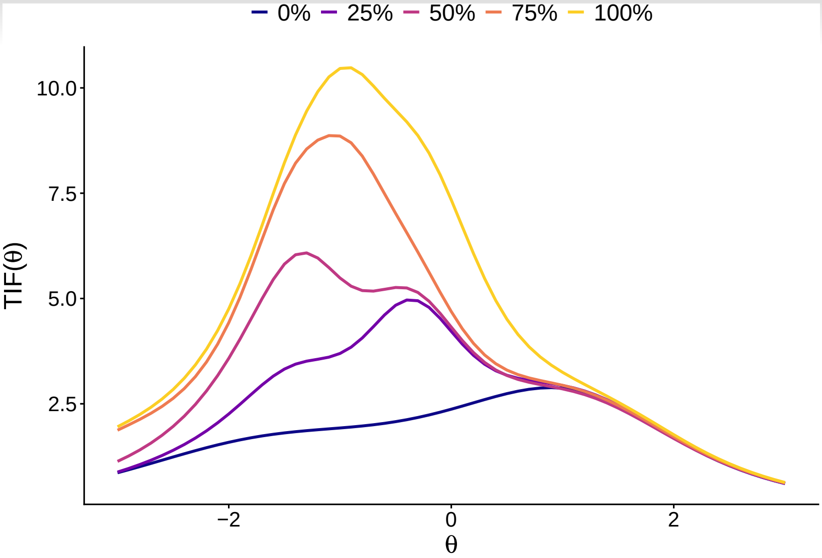


Figure 3: Information Function

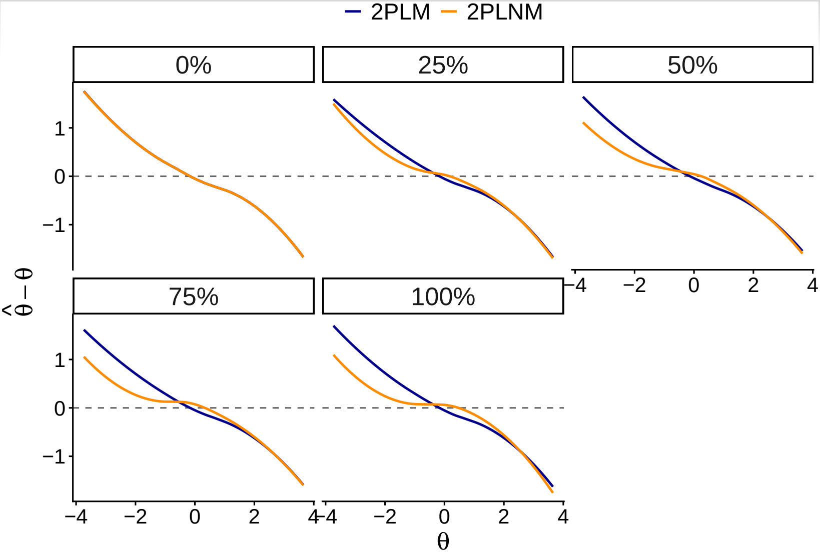


Figure 4: Recovery – Bias

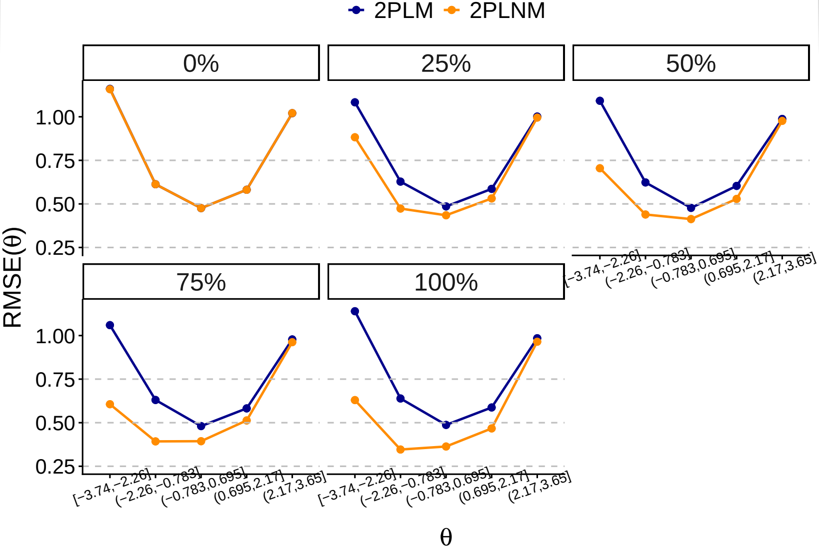


Figure 5: Recovery – RMSE

Take home

Positive outcomes

Attending to the distractors is useful and allows for gaining information and obtaining better estimates of the latent trait.

The nested nature of these models allows for a direct comparison with the model used for the dichotomous responses.

With many items: Attending to the distractors does improve the estimation accuracy, but results are less evident than with fewer items.

Limitations & future directions

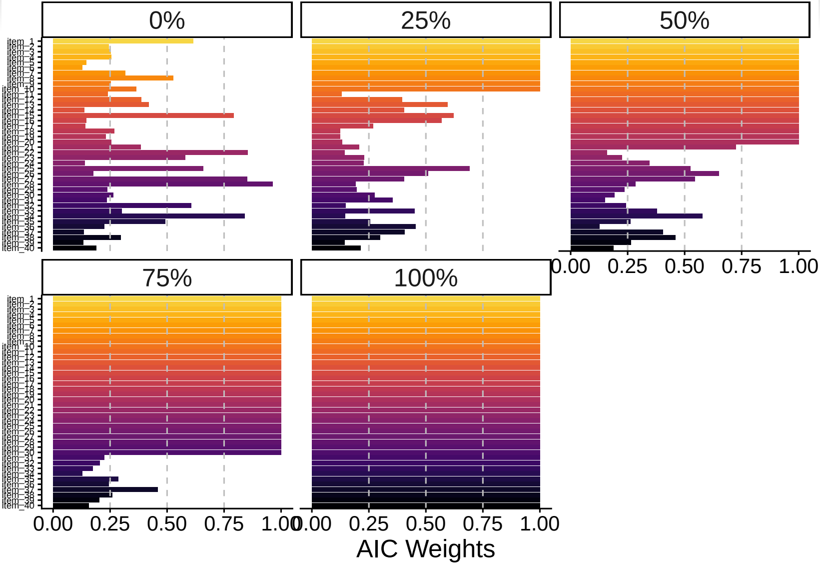
The latent trait is unidimensional – Also for the distractors

Few works on CAT algorithm based on the NRM, no works on CAT algorithm based on NLM

Appendix

Appendix

Simulation results ($I = 40$)


0%
25%
50%
75%
100%
item_1
item_2
item_3
item_4
item_5
item_6
item_7
item_8
item_9
item_10
item_11
item_12
item_13
item_14
item_15
item_16
item_17
item_18
item_19
item_20
item_21
item_22
item_23
item_24
item_25
item_26
item_27
item_28
item_29
item_30
item_31
item_32
item_33
item_34
item_35
item_36
item_37
item_38
item_39
item_40
0.00 0.25 0.50 0.75 1.00
AIC Weights

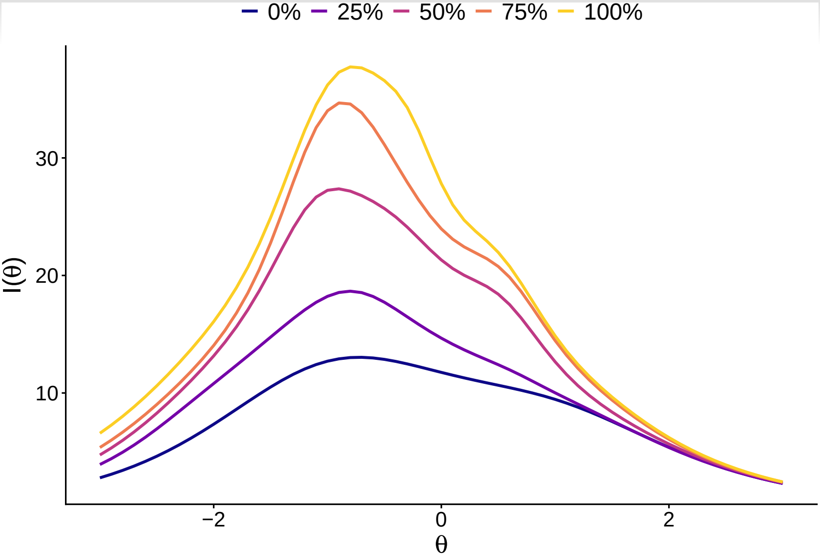


Figure 6: Information Function

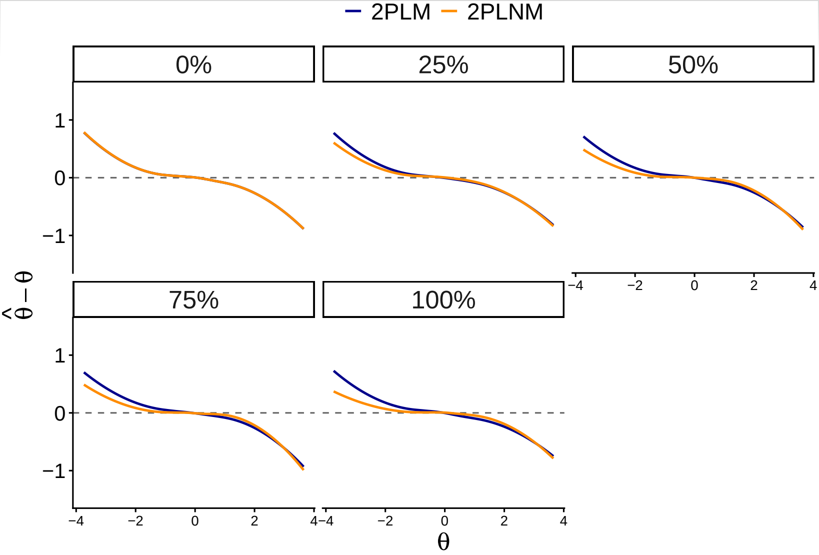


Figure 7: Recovery – Bias

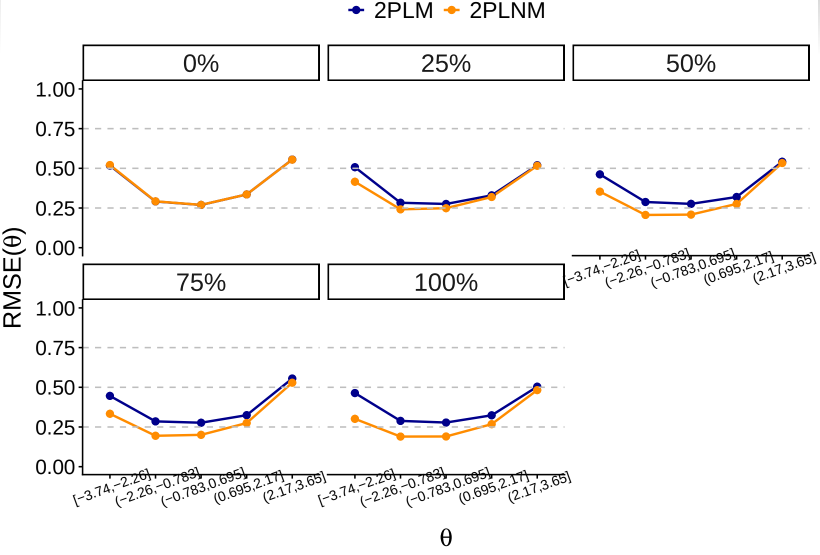


Figure 8: Recovery – RMSE