Test length doesn't matter, it's how you use the items that counts: An intelligent procedure for item selection in Item Response Theory

Ottavia M. Epifania[1,2], Pasquale Anselmi[3], Egidio Robusto[3]
[1] Psychology and Cognitive Science Department, University of Trento, Italy
[2] Psicostat, University of Padova, Italy
[3] Department of Philosophy, Sociology, Education, and Applied Psychology, University of Padova, Italy

Convegno AIP 2024
Sezione Sperimentale

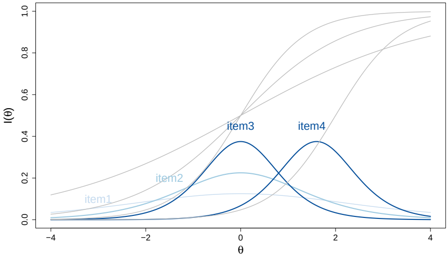
AIM

New automated procedure for item selection in IRT that only requires the definition of the desired characteristics of a test

Item and Test Information Functions

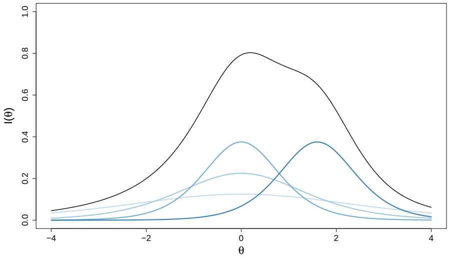
Item Information Function (IIF):

$$I_i(\theta) = a_i^2 P_i(\theta, b_i, a_i)[1 - P_i(\theta, b_i, a_i)]$$



Test Information Function (TIF):

$$I(\theta) = \sum_{i=1}^{N} I_i(\theta)$$

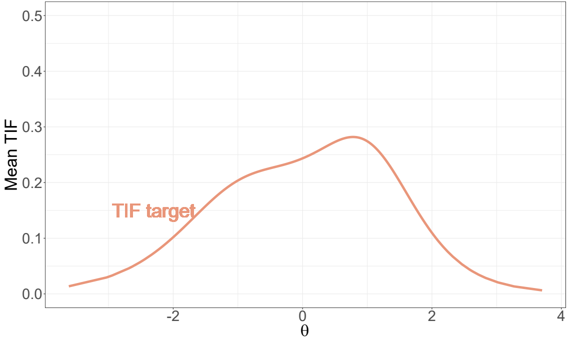


Item Locating Algorithm – ILA

B: Set of items of the item bank , items $i = \{1, 2, \ldots, ||B||\}$

TIF*: TIF target describing the desired characteristic of a test

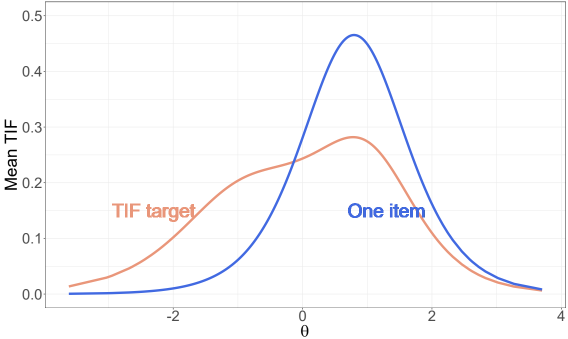
$Q^k_{\mathrm{ILA}} \subset B$: Set of items selected by ILA up to iteration k



$$k = 0,\ Q^0 = \emptyset,$$

$$\theta_{target} := \arg\max |\mathbf{TIF}^* - \mathbf{TIF}^0|, \text{ where } \mathbf{TIF^0} = (0, 0, \ldots, 0)$$

Item Locating Algorithm – ILA

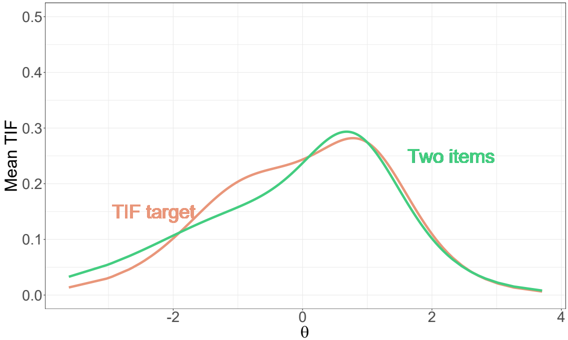


$$k = 1,\ Q^1 = Q^0 \cup \arg\min_{i \in B \setminus Q^0} |\theta_{target} - b_i|,\ ||Q^1|| = 1$$

$$|\mathbf{TIF}^* - \mathbf{TIF}^1| \geq |\mathbf{TIF}^* - \mathbf{TIF}^0| \rightarrow \texttt{false},\ k = 2$$

$$\theta_{target} := \arg\max |\mathbf{TIF}^* - \mathbf{TIF}^1|$$

Item Locating Algorithm – ILA

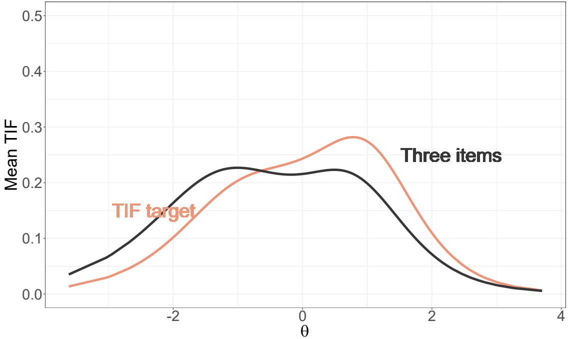


$$Q^2 = Q^1 \cup \arg\min_{i \in B \setminus Q^1} |\theta_{target} - b_i|, \; ||Q^2|| = 2$$

$$|\mathbf{TIF}^* - \mathbf{TIF}^2| \geq |\mathbf{TIF}^* - \mathbf{TIF}^1| \rightarrow \texttt{false}, \; k = 3$$

$$\theta_{target} := \arg\max |\mathbf{TIF}^* - \mathbf{TIF}^2|$$

Item Locating Algorithm – ILA



$$Q^3 = Q^2 \cup \arg\min_{i \in B \setminus Q^2} |\theta_{target} - b_i|, \ ||Q^3|| = 3$$

$$|\textbf{TIF}^* - \textbf{TIF}^3| \geq |\textbf{TIF}^* - \textbf{TIF}^2| \rightarrow \texttt{true} \rightarrow \textbf{end}, \ Q_{\textbf{ILA}} = Q^2$$

Brute Force Procedure – BFP

Item bank B

$Q_m \subset B$ item combinations of different lengths $l = 1, 2, \dots, N-1$

Total number of item combinations $2^{||B||} - 2$

$$\overline{\Delta}_{\mathbf{TIF}^{Q_m}} = mean(|\mathbf{TIF}^* - \frac{\sum_{i \in Q_m} IIF_i}{||Q_m||}|)$$

$$Q_{BFP} = \arg\min_{\emptyset \neq Q_m \subset Q} \overline{\Delta}_{\mathbf{TIF}^{Q_m}}$$

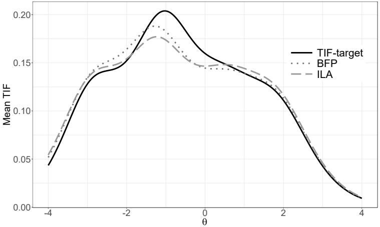
100 data frames:

1. Generate an item bank B of $N = 6$ items:
   - Difficulty parameters: $\mathcal{U}(-3, 3)$
   - Discrimination parameters: $\mathcal{U}(.90, 2.0)$
2. Random item selections of lengths l from B ($M_l = 3.34 \pm 1.13$) + modification parameters $\mathcal{U}(-0.20, 0.20) \rightarrow \mathbf{TIF}^*$
3. Considering $\mathbf{TIF}^*$ at Step 2 and item parameters at Step 1:
   - ILA $\rightarrow$ *Forwardly searches*
   - BFP $\rightarrow$ *Systematically tests*

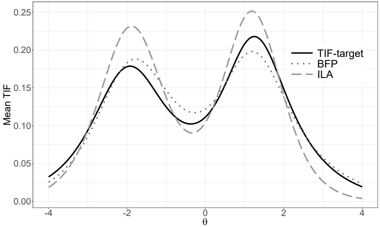
Comparison:

- $||Q_{\text{BFP}}|| - ||Q_{\text{ILA}}||$
- Percentile rank (RP) of the distance $\mathbf{TIF}_{\text{BFP}} - \mathbf{TIF}_{\text{ILA}}$

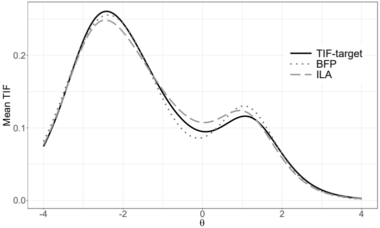
$$||Q_{\mathrm{BFP}}|| = ||Q_{\mathrm{ILA}}||,\ Q_{\mathrm{BFP}} \neq Q_{\mathrm{ILA}},\ RP = 3.17$$



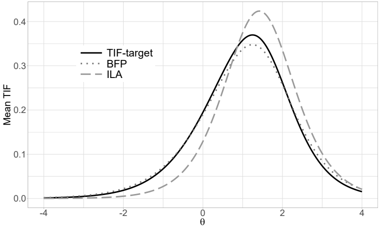
$$||Q_{\mathrm{BFP}}|| > ||Q_{\mathrm{ILA}}||,\ RP = 4.76$$



$$||Q_{\mathrm{BFP}}|| < ||Q_{\mathrm{ILA}}||,\ RP = 3.17$$



$$||Q_{\mathrm{BFP}}|| > ||Q_{\mathrm{ILA}}||,\ RP = 12.70$$



Pros of ILA

- It selects items that are able to recreate the desired characteristics of a test (usually)
- It is computationally "Light"

Cons of ILA

- It grounds its selection on a single θ_{target} at a time $\rightarrow$ it might select items minimizing the distance on that target but that are not very useful for the test
- It only forwardly searches an item $\rightarrow$ once it is in, it can't get out
- It does not account for the discrimination parameters of the items