Pauci sed boni: New Item Response Theory-based procedures for shortening tests

Ottavia M. Epifania[1,2], Pasquale Anselmi[1], Egidio Robusto[1]
ottavia.epifania@unipd.it
[1]Univerisity of Padova
[2]Catholic University of the Sacred Heart

September 7[th] 2022, Rovereto

Meeting of the European Mathematical Psycology Group

Aim

New IRT-based procedures for shortening tests

Aim

New IRT-based procedures for shortening tests

Equal for all respondents

Aim

New IRT-based procedures for shortening tests

Equal for all respondents

Tailored to specific levels of the latent trait

1. **Introduction**

2. Item Response Theory and information functions

3. IRT procedures for shortening tests
   - Benchmark procedure
   - Procedures based on θ targets

4. Simulation study

5. Some final remarks

Item Response Theory and short test forms

ADAPTIVE SHORT FORMS: *Ad-hoc* tests for each person → The information is maximized for each latent trait level (i.e., for each person)

STATIC SHORT FORMS: Static tests equal for all respondents → The information is maximized across all latent trait levels (i.e., across all respondents)

Item Response Theory and short test forms

ADAPTIVE SHORT FORMS: *Ad-hoc* tests for each person → The information is maximized for each latent trait level (i.e., for each person)

Issue

Different short test forms for each person → Potential fairness issues in assessments, e.g., for recruitment

STATIC SHORT FORMS: Static tests equal for all respondents → The information is maximized across all latent trait levels (i.e., across all respondents)

Item Response Theory and short test forms

ADAPTIVE SHORT FORMS: *Ad-hoc* tests for each person → The information is maximized for each latent trait level (i.e., for each person)

Issue

Different short test forms for each person → Potential fairness issues in assessments, e.g., for recruitment

STATIC SHORT FORMS: Static tests equal for all respondents → The information is maximized across all latent trait levels (i.e., across all respondents)

Issue

Not being tailored to any trait level of interest → Potentially more items are needed to cover a wide range of the latent trait

1 Introduction

2 Item Response Theory and information functions

3 IRT procedures for shortening tests
- Benchmark procedure
- Procedures based on θ targets

4 Simulation study

5 Some final remarks

Item Response Theory
2-PL Model

$$P(x_{pj} = 1|\theta_p, b_j, a_j) = \frac{exp[a_j(\theta_p - b_j)]}{1 + exp[a_j(\theta_p - b_j)]}$$

where:

$P(x_{pj} = 1)$: Probability of a correct response to item j by person p

θ_p: Ability of person p

b_j: Difficulty of item j

a_j: Discrimination of item j

Information functions

Item Information Function

$$IIF_j = a_j^2[P(\theta)(1 - P(\theta))]$$

Information functions

Item Information Function

$$IIF_j = a_j^2[P(\theta)(1 - P(\theta))]$$

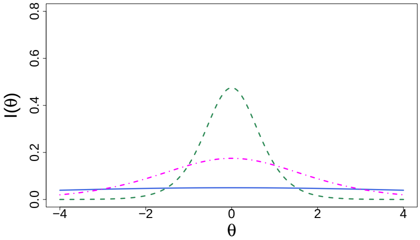


Figure 1: $a = 0.20$, $a = 0.70$, $a = 1.90$, $b = 0$

Information functions

Item Information Function

$$IIF_j = a_j^2[P(\theta)(1 - P(\theta))]$$

Test Information Function

$$TIF = \sum_{j=1}^{J} IIF_j$$

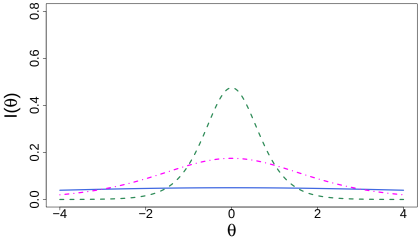


Figure 1: $a = 0.20$, $a = 0.70$, $a = 1.90$, $b = 0$

Information functions

Item Information Function

$$IIF_j = a_j^2[P(\theta)(1 - P(\theta))]$$

Test Information Function

$$TIF = \sum_{j=1}^{J} IIF_j$$

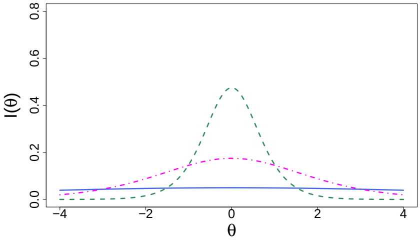


Figure 1: $a = 0.20$, $a = 0.70$, $a = 1.90$, $b = 0$

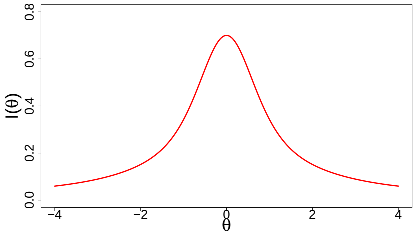


Figure 2: TIF $= IIF_1 + IIF_2 + IIF_3$

1 Introduction

2 Item Response Theory and information functions

3 IRT procedures for shortening tests
- Benchmark procedure
- Procedures based on θ targets

4 Simulation study

5 Some final remarks

1 Introduction

2 Item Response Theory and information functions

3 IRT procedures for shortening tests
- Benchmark procedure
- Procedures based on θ targets

4 Simulation study

5 Some final remarks

Benchmark procedure

Create a short test form composed of N items $\rightarrow$ Select the N items with the highest *IIF*s:

- The *IIF*s of the J items of the full-length test are sorted in decreasing order:

  $$iif = (\max_{1<j<J} IIF_j, \ldots, \min_{1<j<J} IIF_j)$$

- Items with *IIF*s from 1 to N, $N < J$, in *iif* are selected to be included in the short test form

1 Introduction

2 Item Response Theory and information functions

3 IRT procedures for shortening tests
- Benchmark procedure
- Procedures based on θ targets

4 Simulation study

5 Some final remarks

Procedures based on θ targets

Create a short test form composed of N items:

- Define N trait levels of interest $\rightarrow$ θ targets (θ's)
- Choose the items that most precisely assess the identified $\theta's$

$\mathbf{IIF}^{J\times N}$: Matrix of items information functions iif_{jn}:

θ'

		1	2	$\dots$	n	$\dots$	N
	1	iif_{11}	iif_{12}				
	2	iif_{21}	iif_{22}		$\vdots$		
	$\vdots$				$\vdots$		
Items	j	$\dots$	$\dots$	$\dots$	iif_{jn}	$\dots$	$\dots$
	$\vdots$				$\vdots$		
	J				$\vdots$		iif_{JN}

$j = 1, \dots, J$: Full-length test composed of J items;
$n = 1, \dots, N$: Items to be included in the N-item short test form (selected according to the N θ targets, θ's);

$k = 0, \ldots, K$: Scalar denoting the iterations of the procedures ($K = N - 1$);
$S^k \subseteq \{1, \ldots J\}$: Set of items selected to be included in the short test form up to iteration k.
$Q^k \subseteq \{1, \ldots N\}$: Set of θ's satisfied up to iteration k;

$k = 0, \ldots, K$: Scalar denoting the iterations of the procedures ($K = N - 1$);
$S^k \subseteq \{1, \ldots J\}$: Set of items selected to be included in the short test form up to iteration k.
$Q^k \subseteq \{1, \ldots N\}$: Set of θ's satisfied up to iteration k;

At $k = 0$:

- $S^0 = \emptyset$
- $Q^0 = \emptyset$

The procedure cycles steps 1 to 3 until $k = K$:

1. Select $iif_{jn}^k = \max_{j \in J \setminus S^k,\, n \in N \setminus Q^k} \mathbf{IIF}(j, n)$;
2. Compute $S^{k+1} = S^k \cup \{j\}$ as the set of item(s) selected at k;
3. Compute $Q^{k+1} = Q^k \cup \{n\}$ as the set of θ's satisfied at k;

At iteration K, $|Q^{K+1}| = N$ and $|S^{K+1}| = N$

Example

Aim $\rightarrow$ Obtain a 3-item scale ($N = 3$) from a 10-item scale ($J = 10$).

Example

Aim → Obtain a 3-item scale ($N = 3$) from a 10-item scale ($J = 10$).

$$\mathbf{IIF} = \begin{bmatrix} 0.12 & 0.12 & 0.09 \\ 0.14 & 0.32 & 0.31 \\ 0.02 & 0.01 & 0.01 \\ 0.01 & 0.05 & 0.43 \\ 0.03 & 0.03 & 0.04 \\ 0.35 & 0.07 & 0.01 \\ 0.05 & 0.06 & 0.06 \\ 0.27 & 0.04 & 0.01 \\ 0.05 & 0.05 & 0.04 \\ 0.02 & 0.03 & 0.03 \end{bmatrix}$$

At $k = 0 \rightarrow S^0 = \emptyset$, $Q^0 = \emptyset$

$$k = 0$$

$$\mathbf{IIF} = \begin{bmatrix} 0.12 & 0.12 & 0.09 \\ 0.14 & 0.32 & 0.31 \\ 0.02 & 0.01 & 0.01 \\ 0.01 & 0.05 & 0.43 \\ 0.03 & 0.03 & 0.04 \\ 0.35 & 0.07 & 0.01 \\ 0.05 & 0.06 & 0.06 \\ 0.27 & 0.04 & 0.01 \\ 0.05 & 0.05 & 0.04 \\ 0.02 & 0.03 & 0.03 \end{bmatrix}$$

$$k = 0$$

$$\mathbf{IIF} = \begin{bmatrix} 0.12 & 0.12 & 0.09 \\ 0.14 & 0.32 & 0.31 \\ 0.02 & 0.01 & 0.01 \\ 0.01 & 0.05 & \mathbf{0.43} \\ 0.03 & 0.03 & 0.04 \\ 0.35 & 0.07 & 0.01 \\ 0.05 & 0.06 & 0.06 \\ 0.27 & 0.04 & 0.01 \\ 0.05 & 0.05 & 0.04 \\ 0.02 & 0.03 & 0.03 \end{bmatrix}$$

$$iif^0_{\max} = \max_{j \in J \setminus S^0,\, n \in N \setminus Q^0} \mathbf{IIF} = \mathbf{IIF}(4,3) = 0.43$$

$$S^1 = S^0 \cup \{4\} = \{4\}$$

$$Q^1 = Q^0 \cup \{3\} = \{3\}$$

$$k = 1$$

$$\mathbf{IIF} = \begin{bmatrix} 0.12 & 0.12 & 0.09 \\ 0.14 & 0.32 & 0.31 \\ 0.02 & 0.01 & 0.01 \\ 0.01 & 0.05 & 0.43 \\ 0.03 & 0.03 & 0.04 \\ \mathbf{0.35} & 0.07 & 0.01 \\ 0.05 & 0.06 & 0.06 \\ 0.27 & 0.04 & 0.01 \\ 0.05 & 0.05 & 0.04 \\ 0.02 & 0.03 & 0.03 \end{bmatrix}$$

$$iif^1_{max} = \max_{j \in J \setminus S^1,\, n \in N \setminus Q^1} \mathbf{IIF} = \mathbf{IIF}(6, 1) = 0.35$$

$$S^2 = S^1 \cup \{6\} = \{4, 6\}$$

$$Q^2 = Q^1 \cup \{1\} = \{3, 1\}$$

$k = 2$

$$\mathbf{IIF} = \begin{bmatrix} 0.12 & 0.12 & 0.09 \\ 0.14 & \mathbf{0.32} & 0.31 \\ 0.02 & 0.01 & 0.01 \\ 0.01 & 0.05 & 0.43 \\ 0.03 & 0.03 & 0.04 \\ 0.35 & 0.07 & 0.01 \\ 0.05 & 0.06 & 0.06 \\ 0.27 & 0.04 & 0.01 \\ 0.05 & 0.05 & 0.04 \\ 0.02 & 0.03 & 0.03 \end{bmatrix}$$

$$iif^2_{max} = \max_{j \in J \setminus S^2,\, n \in N \setminus Q^2} \mathbf{IIF} = \mathbf{IIF}(2,2) = 0.32$$

$$S^3 = S^2 \cup \{2\} = \{4, 6, 2\}$$

$$Q^3 = Q^2 \cup \{2\} = \{3, 1, 2\}$$

$k = 2$

$$\mathbf{IIF} = \begin{bmatrix} 0.12 & 0.12 & 0.09 \\ 0.14 & \mathbf{0.32} & 0.31 \\ 0.02 & 0.01 & 0.01 \\ 0.01 & 0.05 & 0.43 \\ 0.03 & 0.03 & 0.04 \\ 0.35 & 0.07 & 0.01 \\ 0.05 & 0.06 & 0.06 \\ 0.27 & 0.04 & 0.01 \\ 0.05 & 0.05 & 0.04 \\ 0.02 & 0.03 & 0.03 \end{bmatrix}$$

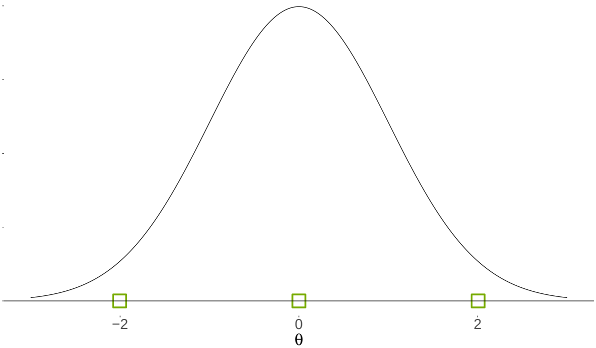
$$iif^2_{max} = \max_{j \in J \setminus S^2,\, n \in N \setminus Q^2} \mathbf{IIF} = \mathbf{IIF}(2,2) = 0.32$$

$$S^3 = S^2 \cup \{2\} = \{4, 6, 2\}$$

$$Q^3 = Q^2 \cup \{2\} = \{3, 1, 2\}$$

- $|S^3| = 3$
- $|Q^3| = 3$
- $K = 2$

$\rightarrow$ END

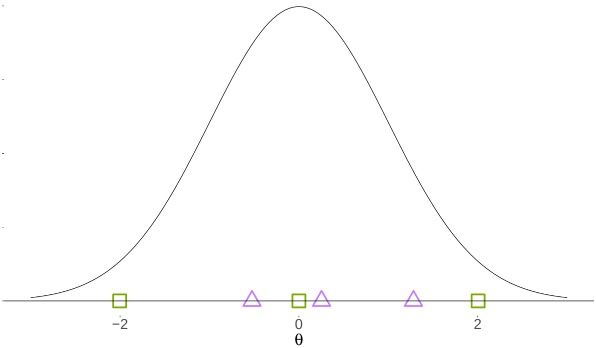


Equal Intervals Procedure (EIP)

Segmentation in $N+1$ intervals of width $w = (\theta_{max} - \theta_{min})/N$

Each interval: $[\theta_{n-1}; \theta_n]$

$\theta'_n = (\theta_{n-1} + \theta_n)/2$



Equal Intervals Procedure (EIP)

Segmentation in $N+1$ intervals of width $w = (\theta_{max} - \theta_{min})/N$

Each interval: $[\theta_{n-1}; \theta_n]$

$\theta'_n = (\theta_{n-1} + \theta_n)/2$

Unequal Intervals Procedure (UIP)

Clustering in N clusters

The c_n centroids are the θ's

1 Introduction

2 Item Response Theory and information functions

3 IRT procedures for shortening tests
- Benchmark procedure
- Procedures based on θ targets

4 Simulation study

5 Some final remarks

Comparison between the item selection procedures:

- **Benchmark procedure (BP)**: The N items with the highest *IIF*s are selected from the full-length test
- **Equal Intervals Procedure (EIP)**: The N items that maximize the information for each θ' obtained by dividing the latent trait into equal intervals are selected
- **Unequal Intervals Procedure (UIP)**: The N items that maximize the information for each θ' obtained by clustering the latent trait are selected
- **Random Procedure (RP)**: N items are randomly selected from the full-length tests

10, 30, 50, 70, 90-item short test forms from a 100-item full-length test

1000 respondents p

1. Normal distribution
   $p \sim \mathcal{N}(0, 1)$
2. Positive skewed distribution
   $p \sim Beta(1, 100)$ (linearly transformed to obtain negative values)
3. Uniform distribution
   $p \sim \mathcal{U}(-3, 3)$

100 items s:

- $b \sim \mathcal{U}(-3, 3)$
- $a \sim \mathcal{U}(0.40, 2)$

An overall look

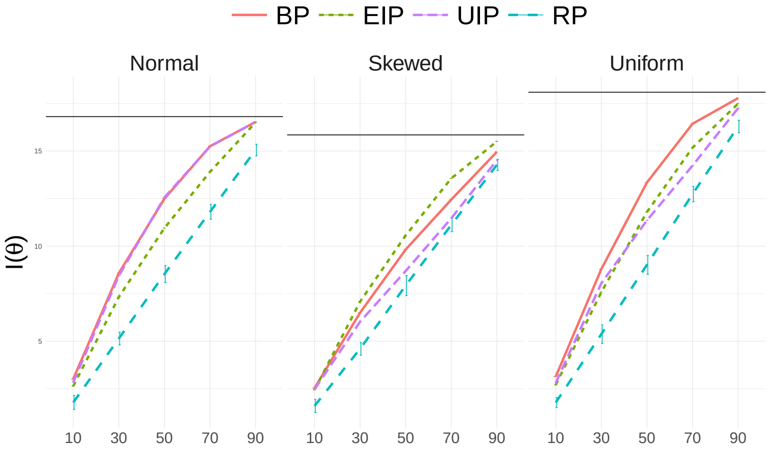


Figure 3: Overall Information of the short test forms

A closer look

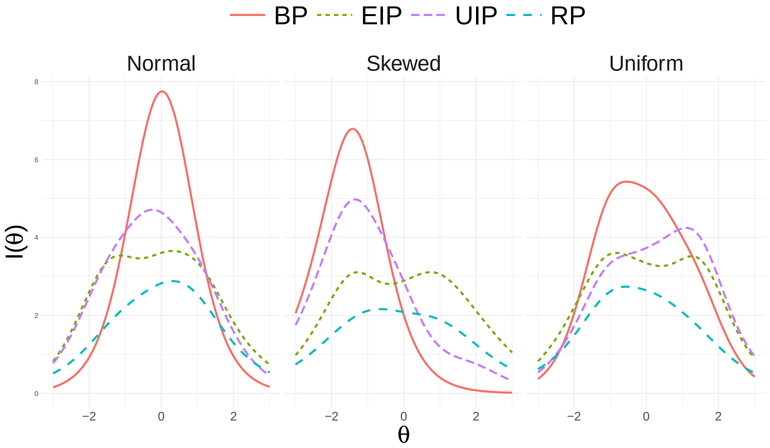


Figure 4: TIF of the 10-item short test form

An even closer look

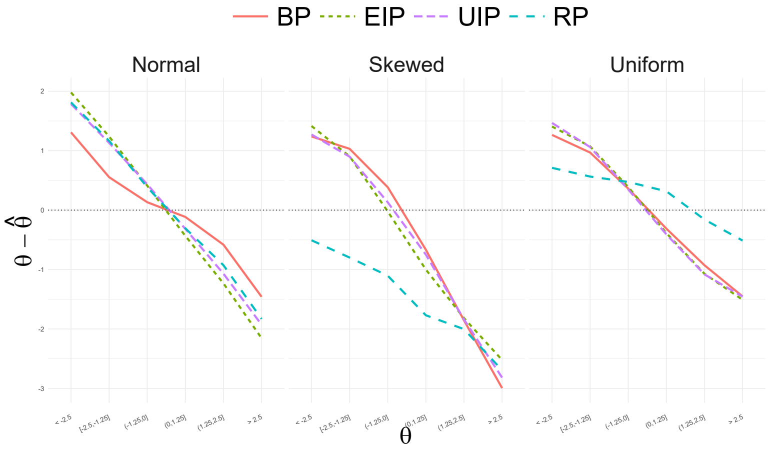


Figure 5: *bias* $= \theta - \hat{\theta}$ of the 10-item short test form

1 Introduction

2 Item Response Theory and information functions

3 IRT procedures for shortening tests
- Benchmark procedure
- Procedures based on θ targets

4 Simulation study

5 Some final remarks

Good!

There's no "one-fits-all" solution

The θ distribution is a key element

Good!

There's no "one-fits-all" solution

The θ distribution is a key element

..but work is still needed

Lack of direct comparison with CAT
Real life applications are missing

Thank you!

ottavia.epifania@unipd.it