Pochi ma buoni: Procedure Item Response Theory per lo sviluppo di forme brevi di test

Ottavia M. Epifania
ottavia.epifania@unitn.it

Psychology and Cognitive Science Department, University of Trento, Italy
Psicostat, University of Padova, Italy

Formazione Metodologica (FORME)
September 27th, 2024

A_{Bart}

Q1

$$4 + 5 = ?$$

$$d_{q1}$$

Q2

$$\frac{3}{2}x^2 + \frac{5}{4}x = ?$$

$$d_{q2}$$

A_{Lisa}

$$P(X_{pi} = 1) = \frac{\frac{A_p}{d_i}}{1 + \frac{A_p}{d_i}}$$

A_{Bart}

Q1

$$4 + 5 = ?$$

$$d_{q1}$$

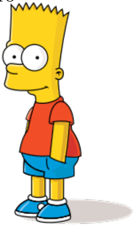
Q2

$$\frac{3}{2}x^2 + \frac{5}{4}x = ?$$

$$d_{q2}$$

A_{Lisa}

$$P(X_{pi} = 1) = \frac{\frac{A_p}{d_i}}{1 + \frac{A_p}{d_i}}$$



A_{Bart}

Q1

$$4 + 5 = ?$$

d_{q1}

Q2

$$\frac{3}{2}x^2 + \frac{5}{4}x = ?$$

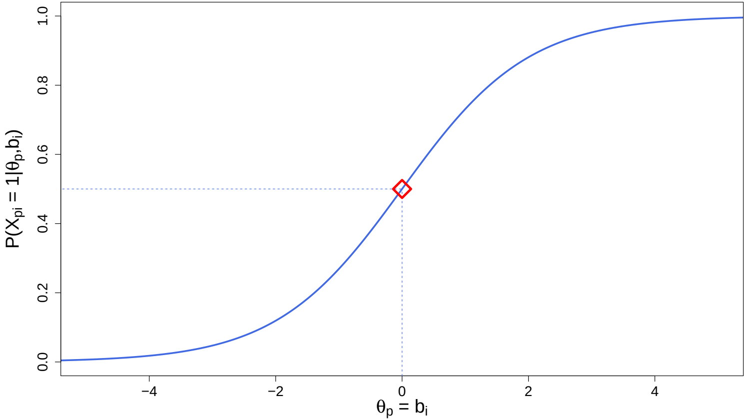
d_{q2}

A_{Lisa}

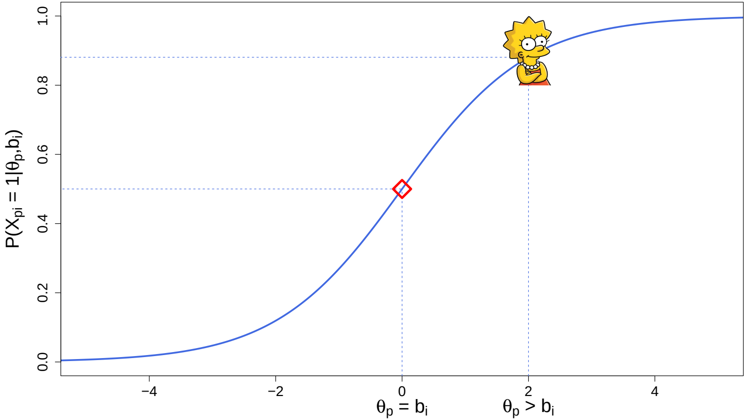
$$P(X_{pi} = 1) = \frac{\frac{A_p}{d_i}}{1 + \frac{A_p}{d_i}}$$

$$P(X_{pi} = 1 | \theta_p, b_i) = \frac{\exp(\theta_p - b_i)}{1 + \exp(\theta_p - b_i)}$$

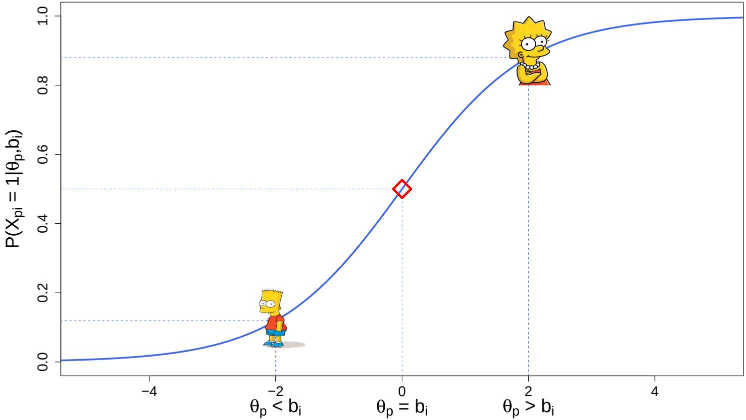
$$P(X_{pi} = 1|\theta_p, b_i) = \frac{\exp(\theta_p - b_i)}{1 + \exp(\theta_p - b_i)}$$



$$P(X_{pi} = 1|\theta_p, b_i) = \frac{\exp(\theta_p - b_i)}{1 + \exp(\theta_p - b_i)}$$



$$P(X_{pi} = 1|\theta_p, b_i) = \frac{\exp(\theta_p - b_i)}{1 + \exp(\theta_p - b_i)}$$



To each its own...IRT model

Different IRT models according to:

- Latent trait:
  - Unidimensional model
  - Multidimensional model
- Response categories:
  - Dichotomous items (Two response categories, e.g., true/falso, agree/disagree)
  - Polytomous items (at least 3 response categories, e.g., Likert-type scale)

To each its own...IRT model

Different IRT models according to:

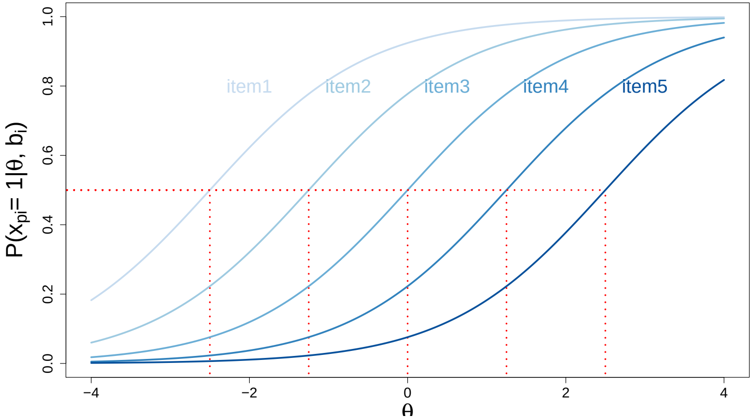
- Latent trait:
  - Unidimensional model
  - Multidimensional model
- Response categories:
  - Dichotomous items (Two response categories, e.g., true/falso, agree/disagree)
  - Polytomous items (at least 3 response categories, e.g., Likert-type scale)

Models for dichotomous responses

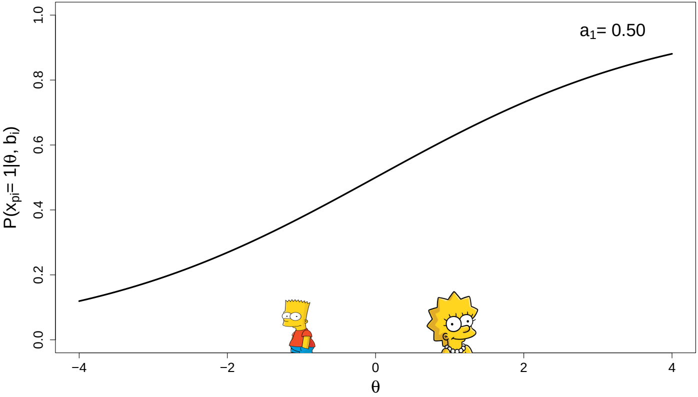
IRT models can be distinguinshed according to the number of parameters describing the charcateristics of the items.

- One-Parameter Logistic Model (1-PL)
- Two-Parameter Logistic Model (2-PL; Birnbaum, 1968)
- Three-Parameter Logistic Model (3-PL; Lord, 1980)
- Four-Parameter Logistic Model (4-PL; Barton & Lord, 1981)

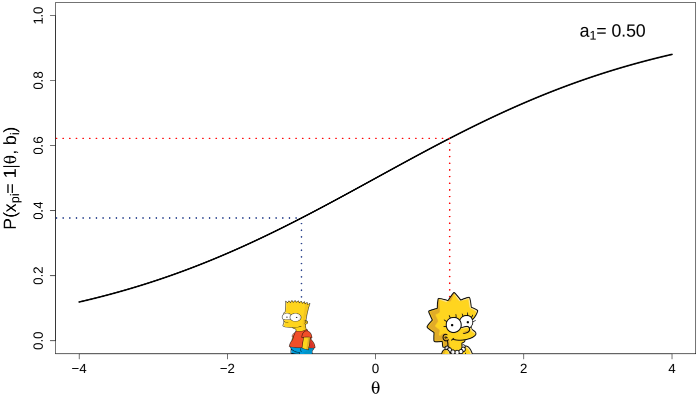
$$P(x_{pi} = 1|\theta_p, b_i) = \frac{\exp(\theta_p - b_i)}{1 + \exp(\theta_p - b_i)}$$



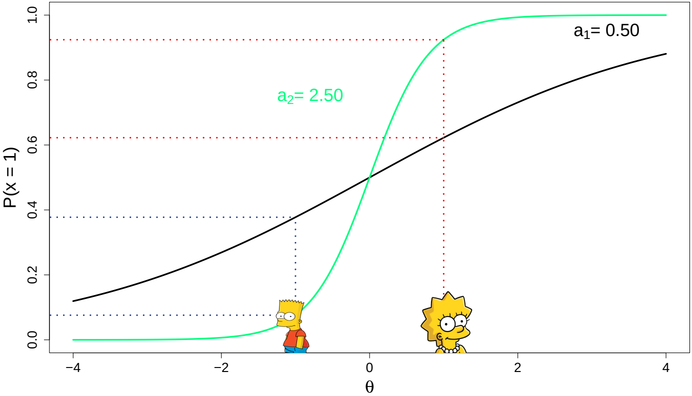
$$P(x_{pi} = 1|\theta_p, b_i, a_i) = \frac{\exp[a_i(\theta_p - b_i)]}{1 + \exp[a_i(\theta_p - b_i)]}$$



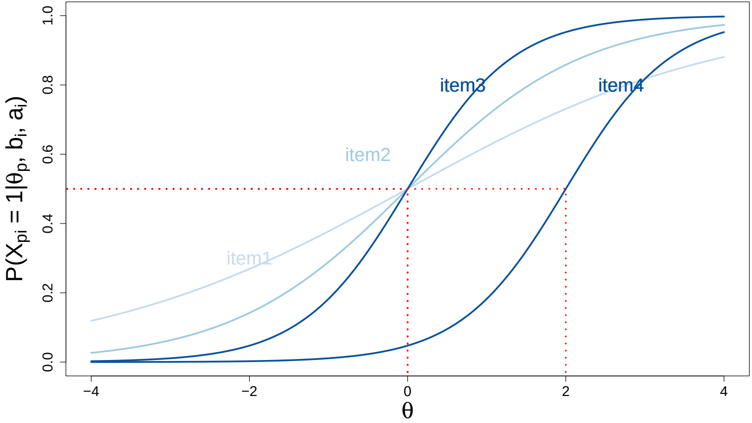
$$P(x_{pi} = 1|\theta_p, b_i, a_i) = \frac{\exp[a_i(\theta_p - b_i)]}{1 + \exp[a_i(\theta_p - b_i)]}$$



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

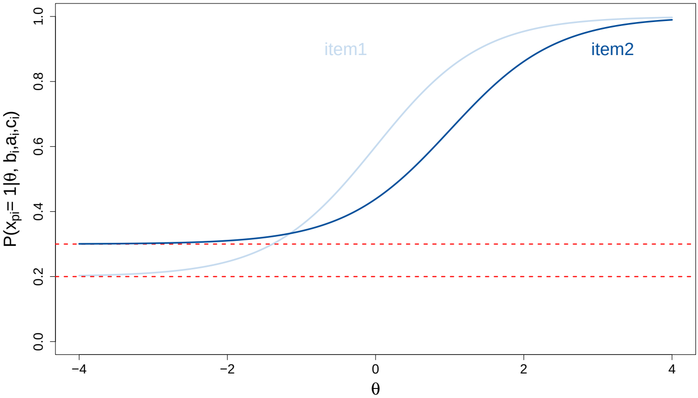


2PL



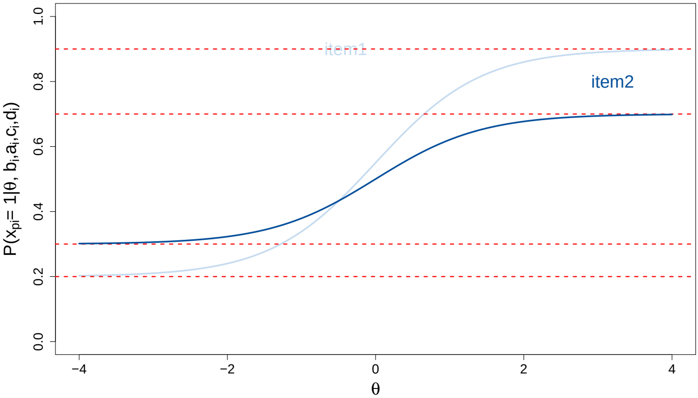
3-PL

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



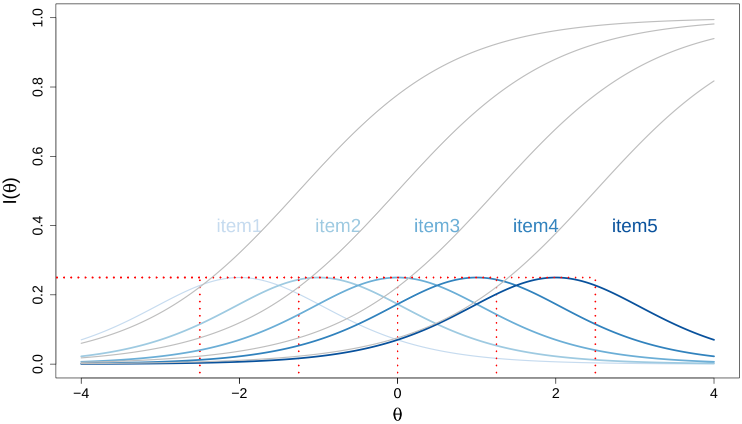
4-PL

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

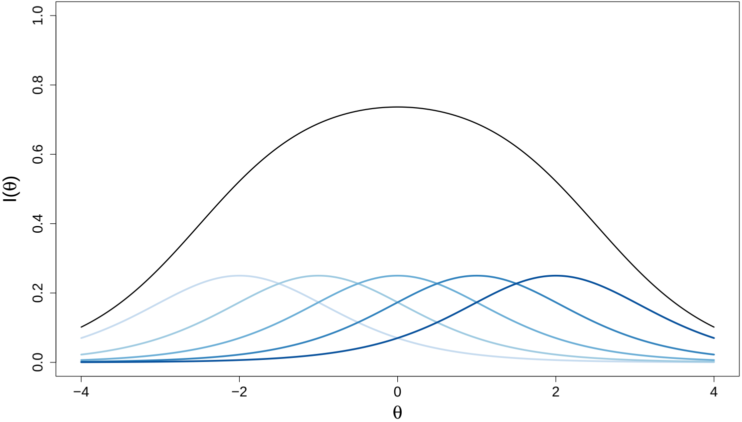


Item Information Function:

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

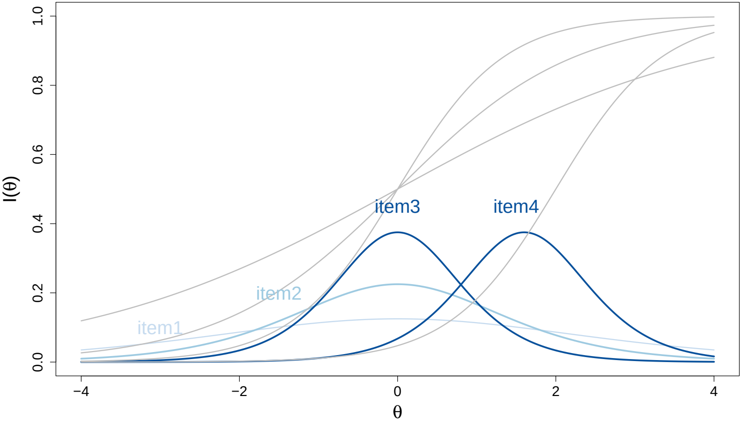


Test Information Function: $I(\theta) = \sum_{i=1}^{I} I_i(\theta)$

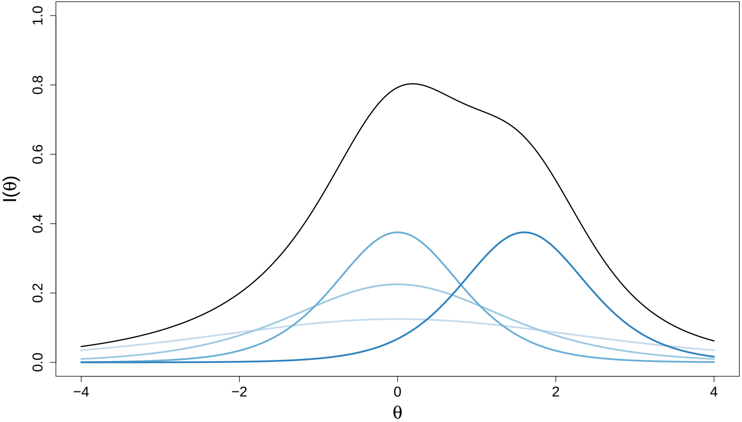


Item Information Function:

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



Test Information Function:
$I(\theta) = \sum_{i=1}^{I} I_i(\theta)$



Why?

Many items $\rightarrow$ good measurement precision, great reliability and so on

Not always!

People might get tired and frustrated

IRT models for the win

Being focused on the item information and on the ability of each item to measure different levels of the latent trait, IRT models provide an ideal framework for developing STF (and not torturing people)

Static STFs

- Equal for all respondents
- Can be administered paper-and-pencil/computerized versions
- Might not provide adequate measurement precision of certain regions of the latent trait

Adaptive STFs

- Tailored on the actual level of ability of each respondent
- Avoid frustration and boredom
- Fairness issues in specific evaluation contexts

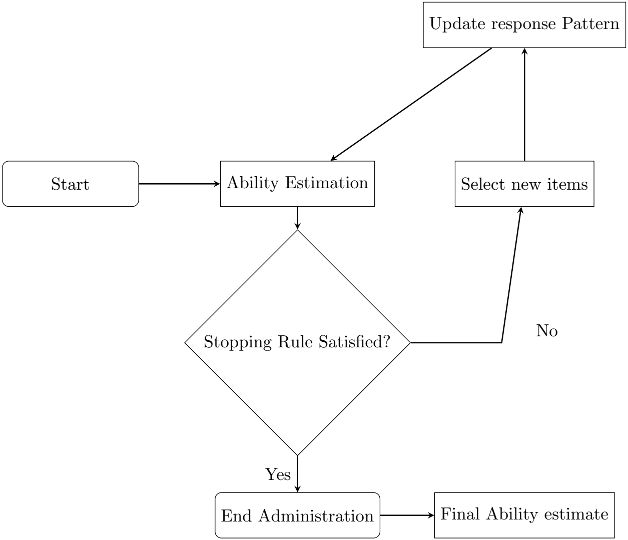
Static STFs

- Equal for all respondents
- Can be administered paper-and-pencil/computerized versions
- Might not provide adequate measurement precision of certain regions of the latent trait

Adaptive STFs

- Tailored on the actual level of ability of each respondent
- Avoid frustration and boredom
- Fairness issues in specific evaluation contexts

- **Item bank B:** Set of items which are administered in a CAT
- $||B||$: Number of items in the item bank
- **An item bank must be calibrated**
- **Item calibration:** estimation of item parameters according to IRT models


Update response Pattern
Start
Ability Estimation
Select new items
Stopping Rule Satisfied?
No
Yes
End Administration
Final Ability estimate

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

- The *IIF*s of the items of item bank 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 *e.g.: 3-item short form from 10-item full-length test*

item	b_i	a_i	IIF_i
1	−0.67	0.71	0.08
2	0.50	1.19	0.15
3	−2.43	0.25	0.01
4	2.12	1.98	0.24
5	1.72	0.39	0.03
6	−2.28	1.62	0.19
7	0.64	0.50	0.05
8	−2.51	1.68	0.19
9	−0.66	0.44	0.04
10	0.72	0.33	0.02

e.g.: 3-item short form from 10-item full-length test

item	b_i	a_i	IIF_i
4	2.12	1.98	0.24
8	−2.51	1.68	0.19
6	−2.28	1.62	0.19
2	0.50	1.19	0.15
1	−0.67	0.71	0.08
7	0.64	0.50	0.05
9	−0.66	0.44	0.04
5	1.72	0.39	0.03
10	0.72	0.33	0.02
3	−2.43	0.25	0.01

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}:

$$
\begin{array}{cc|cccccc|}
 & & & & \theta' & & & \\
 & & 1 & 2 & \dots & n & \dots & N \\
\hline
 & 1 & iif_{11} & iif_{12} & & & & \\
 & 2 & iif_{21} & iif_{22} & & \vdots & & \\
 & \vdots & & & & \vdots & & \\
\text{Items} & j & \dots & \dots & \dots & iif_{jn} & \dots & \dots \\
 & \vdots & & & & \vdots & & \\
 & J & & & & \vdots & & iif_{JN} \\
\hline
\end{array}
$$

$j = 1, \dots, J$: Items in the item bank B composed of $||B|| = J$ 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;

At $k = 0$:

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

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

1. Select $iif^k_{jn} = \max\limits_{j \in J \setminus S^k,\, n \in N \setminus Q^k} \mathbf{IIF}(j, n)$;
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 $|S^{K+1}| = N$

$k = 0, \ldots, K$: 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 $iif^k_{jn} = \max\limits_{j \in J \setminus S^k,\, n \in N \setminus Q^k} \mathbf{IIF}(j, n)$;
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 $|S^{K+1}| = N$

$k = 0$

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

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

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

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

$k = 0$

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

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

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

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

$k = 1$

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

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

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

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

$k = 2$

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

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

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

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

$|S^3| = 3, \quad |Q^3| = 3, \; K = 2 \rightarrow$ end

$k = 2$

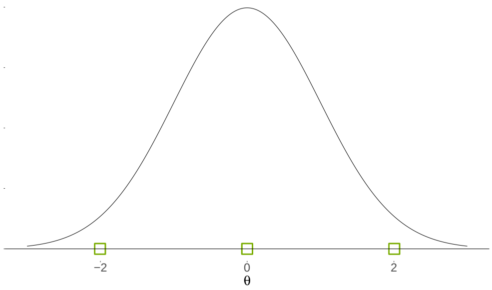
$$\mathbf{IIF} = \begin{bmatrix} 0.12 & 0.12 & 0.09 \\ 0.14 & \mathbf{0.32} & 0.31 \\ 0.02 & 0.01 & 0.01 \\ 0.01 & 0.05 & 0.43 \\ 0.03 & 0.03 & 0.04 \\ 0.35 & 0.07 & 0.01 \\ 0.05 & 0.06 & 0.06 \\ 0.27 & 0.04 & 0.01 \\ 0.05 & 0.05 & 0.04 \\ 0.02 & 0.03 & 0.03 \end{bmatrix}$$

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

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

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

$$|S^3| = 3, \quad |Q^3| = 3, \; K = 2 \rightarrow \texttt{end}$$



Equal Intervals Procedure (EIP)

Unequal Intervals Procedure (UIP)

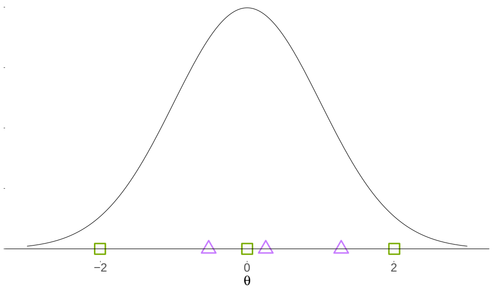
Segmentation in $N+1$ intervals of width $w = (\theta_{max} - \theta_{min})/N$

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

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

Clustering in N clusters

The c_n centroids are the θ's



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

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 i:

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

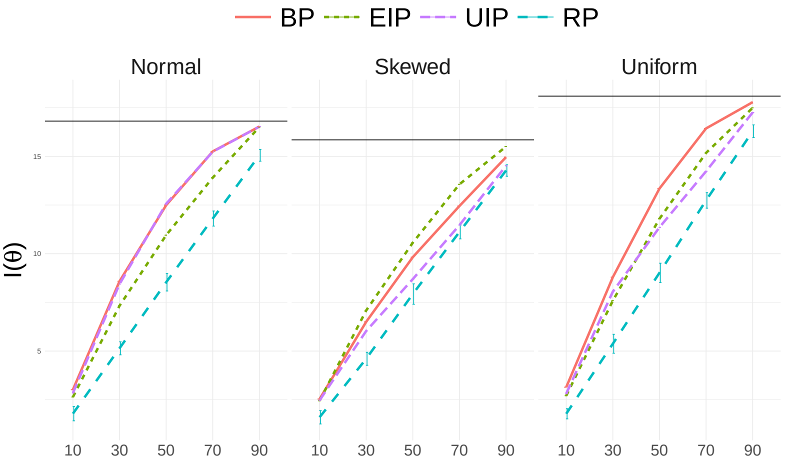


Figure: Overall Information of the short test forms

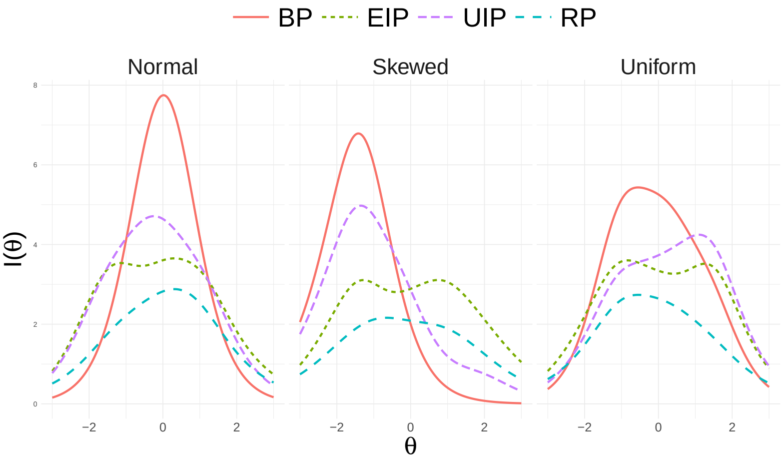


Figure: TIF of the 10-item short test form

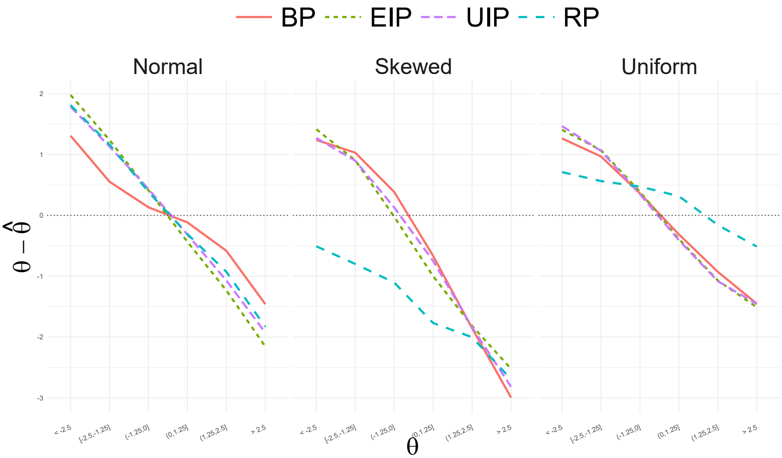


Figure: $bias = \theta - \hat{\theta}$ of the 10-item short test form

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

The number of items included in the STF strictly depends on the θ targets

$\rightarrow$ One item for each target

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
The number of items included in the STF strictly depends on the θ targets
$\rightarrow$ One item for each target

1. Item Response Theory models
2. Item and Test Information Functions
3. Short Test Forms
4. Computerized Adaptive Testing
5. Benchmark procedure
6. θ-target Procedure
7. The `shortIRT` package
8. Item Locating Algorithm
9. Wrapping up

The shortIRT package

shortIRT: Procedures Based on Item Response Theory Models for the Development of Short Test Forms

Implement different Item Response Theory (IRT) based procedures for the development of static short test forms (STFs) from a test. Two main procedures are considered, specifically the typical IRT-based procedure for the development of STF, and a recently introduced procedure (Epifania, Anselmi & Robusto, 2022 <doi:10.1007/978-3-031-27781-8_7>). The procedures differ in how the most informative items are selected for the inclusion in the STF, either by considering their item information functions without considering any specific level of the latent trait (typical procedure) or by considering their informativeness with respect to specific levels of the latent trait, denoted as theta targets (the newly introduced procedure). Regarding the latter procedure, three methods are implemented for the definition of the theta targets: (i) theta targets are defined by segmenting the latent trait in equal intervals and considering the midpoint of each interval (equal interval procedure, eip), (ii) by clustering the latent trait to obtain unequal intervals and considering the centroids of the clusters as the theta targets (unequal intervals procedure, uip), and (iii) by letting the user set the specific theta targets of interest (user-defined procedure, udp). For further details on the procedure, please refer to Epifania, Anselmi & Robusto (2022) <doi:10.1007/978-3-031-27781-8_7>.

Version:	0.1.3
Imports:	stats, TAM, dplyr, ggplot2
Suggests:	rmarkdown, sirt, testthat (≥ 3.0.0), V8
Published:	2024-02-16
DOI:	10.32614/CRAN.package.shortIRT
Author:	Ottavia M. Epifania [aut, cre, cph] Pasquale Anselmi [aut, ctb] Egidio Robusto [ctb]
Maintainer:	Ottavia M. Epifania <ottavia.epifania at unipd.it>
License:	MIT + file LICENSE
NeedsCompilation:	no
Materials:	README NEWS
CRAN checks:	shortIRT results

Documentation:

Reference manual: shortIRT.pdf

https://cran.r-project.org/web/packages/shortIRT/index.html

The shortIRT package

`bp()`	Benchmark Procedure
`eip()`	Equal Interval Procedure
`uip()`	Unequal interval procedure
`plot_difference()`	Plot the difference between θs
`plot_tif()`	Plot Test Information Functions
`diff_theta()`	Difference between thetas

```
# install the shortIRT package. Must be run once
install.packages("shortIRT")
# make the package available
library(shortIRT)
```

The shortIRT package

`bp()`	Benchmark Procedure
`eip()`	Equal Interval Procedure
`uip()`	Unequal interval procedure
`plot_difference()`	Plot the difference between θs
`plot_tif()`	Plot Test Information Functions
`diff_theta()`	Difference between thetas

```
# install the shortIRT package. Must be run once
install.packages("shortIRT")
# make the package available
library(shortIRT)
```

AIM:
Generate a 10-item STF from an item bank B of 100 items

Generate data

```
library(sirt) # quite convenient to generate data
set.seed(999) # set a seed to replicate the results
# simulate the true thetas from a Normal distribution
true_theta = rnorm(1000)
# simulate the item parameters
parameters = data.frame(b = runif(100, -3, 3), a = runif(100, 0.6, 2))
# simulate the responses given the true theta and the item parameters
data = sirt::sim.raschtype(true_theta, b = b, fixed.a = a)
```

AIM:

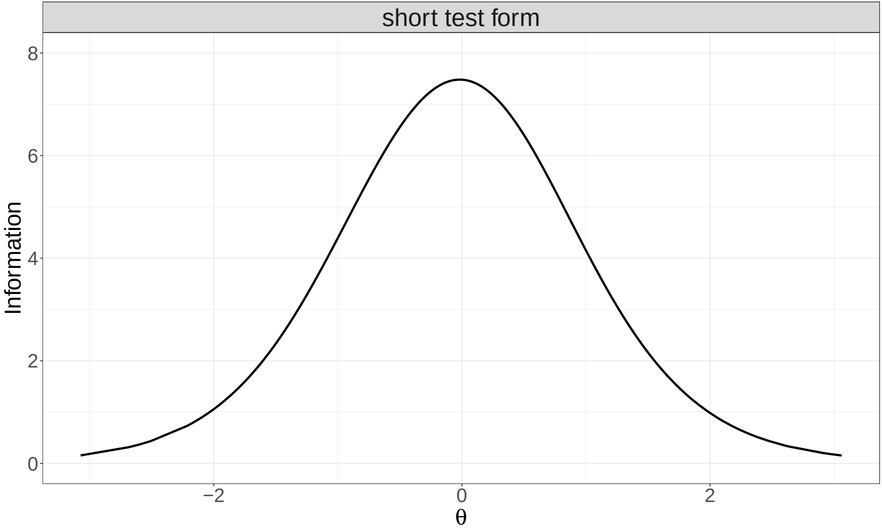
Generate a 10-item STF from an item bank B of 100 items

Generate data

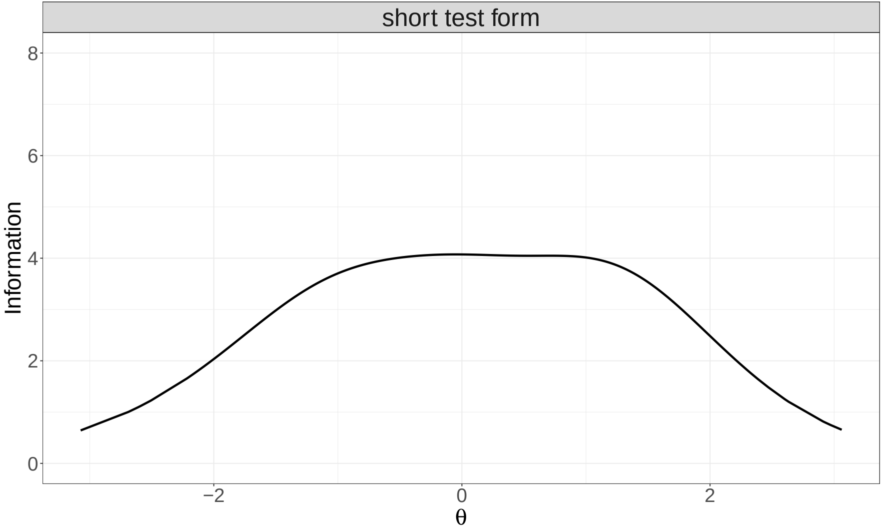
```
library(sirt) # quite convenient to generate data
set.seed(999) # set a seed to replicate the results
# simulate the true thetas from a Normal distribution
true_theta = rnorm(1000)
# simulate the item parameters
parameters = data.frame(b = runif(100, -3, 3), a = runif(100, 0.6, 2))
# simulate the responses given the true theta and the item parameters
data = sirt::sim.raschtype(true_theta, b = b, fixed.a = a)
```

```
# STF according to BP
stf_bp = bp(data, starting_theta = true_theta,
   item_par = parameters,
   num_item = 10)
# STF according to EIP
stf_eip = eip(data, starting_theta = true_theta,
   item_par = parameters,
   num_item = 10)
# STF according to UIP
stf_uip = uip(data, starting_theta = true_theta,
   item_par = parameters,
   num_item = 10)
```

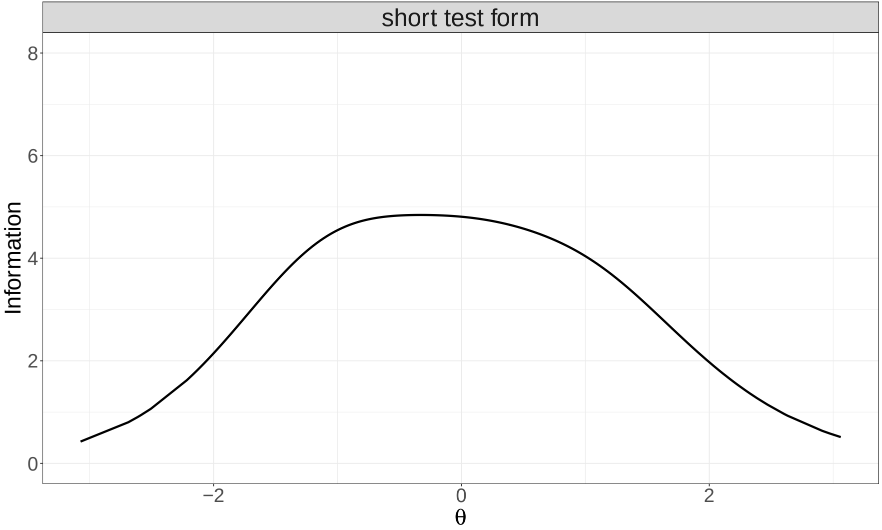
plot_tif(stf_bp)



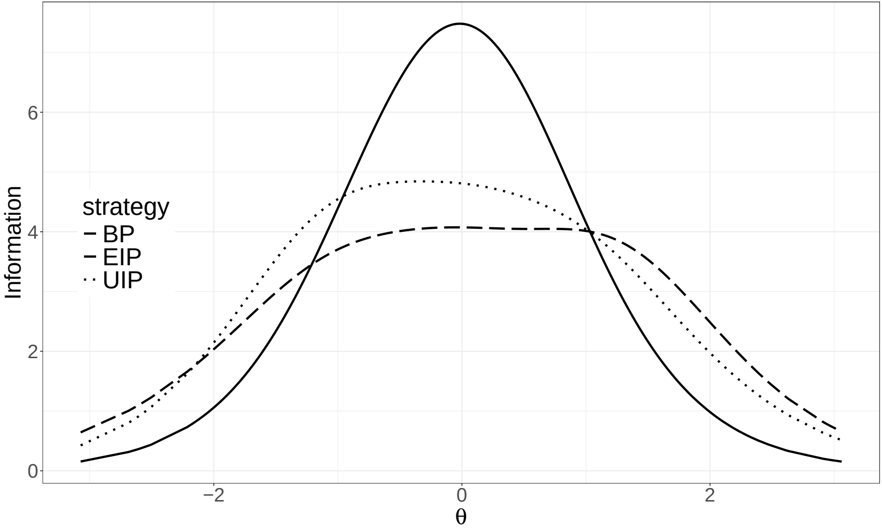
`plot_tif(stf_eip)`



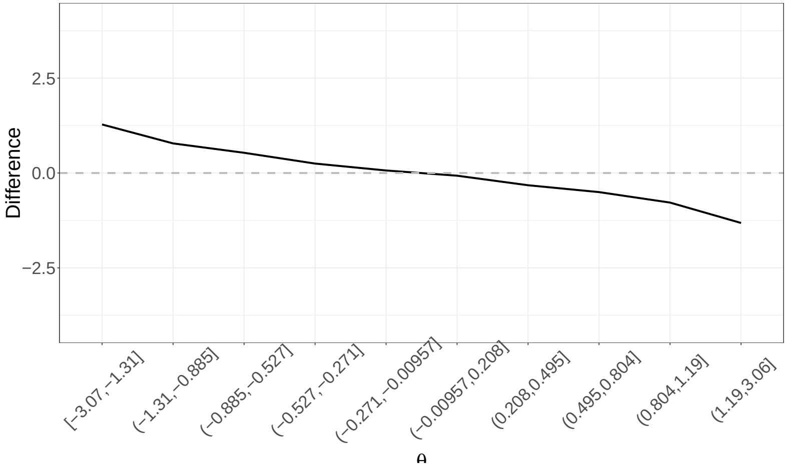
plot_tif(stf_uip)



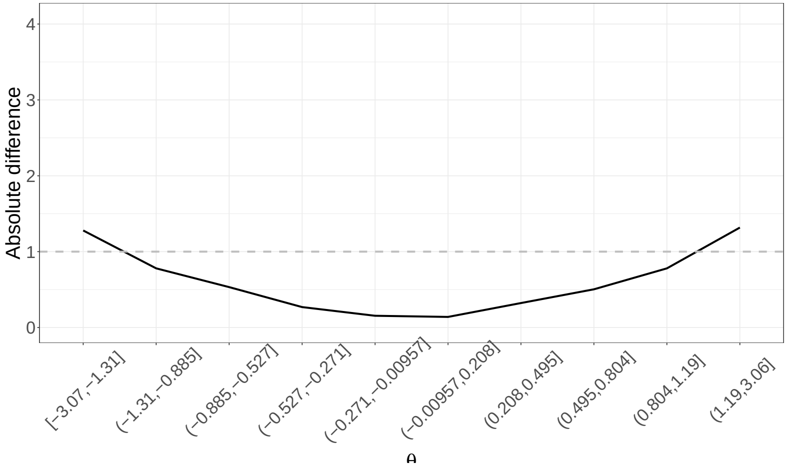
All infos:



```
eip_difference  = diff_theta(stf_eip, starting_theta = true_theta)
plot_difference(eip_difference, type = "diff", levels = 10)
```



```
eip_difference  = diff_theta(stf_eip, starting_theta = true_theta)
plot_difference(eip_difference, type = "absolute_diff", levels = 10)
```



1. Item Response Theory models
2. Item and Test Information Functions
3. Short Test Forms
4. Computerized Adaptive Testing
5. Benchmark procedure
6. θ-target Procedure
7. The `shortIRT` package
8. **Item Locating Algorithm**
9. Wrapping up

Set up:

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

$Q^k \subset B$: Set of item selected for inclusion in the STF up to iteration k ($Q^0 = \emptyset$)

$\mathbf{TIF}^*$: TIF target

$\mathbf{TIF}^0 = (0, 0, \ldots, 0)$

ILA Algorithm:

Set up:

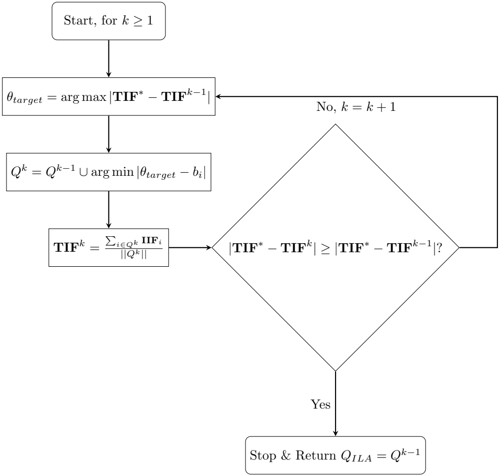
B: Set of items of the item bank, items $i = \{1, 2, \dots, ||B||\}$

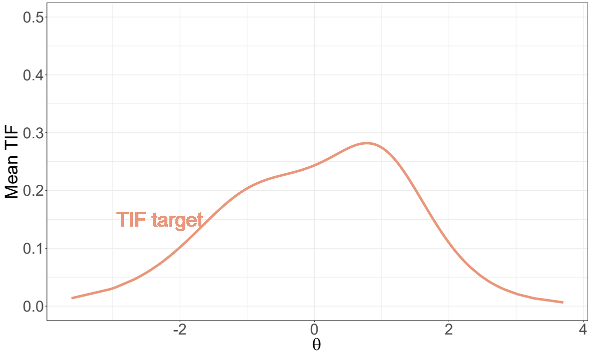
$Q^k \subset B$: Set of item selected for inclusion in the STF up to iteration k ($Q^0 = \emptyset$)

$\mathbf{TIF}^*$: TIF target

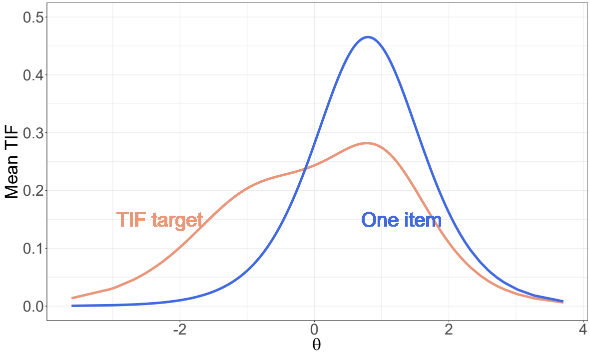
$\mathbf{TIF}^0 = (0, 0, \dots, 0)$

ILA Algorithm:





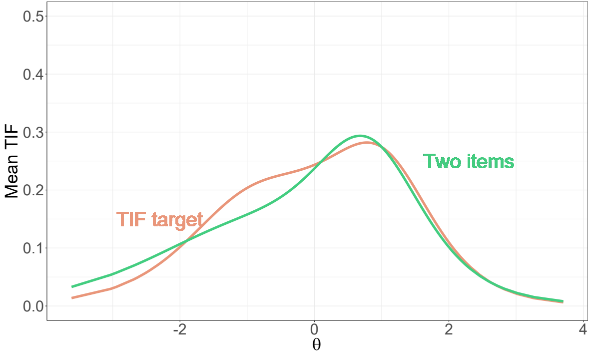
$$k = 0,\ Q^0 = \emptyset,$$
$$\theta_{target} := \arg\max |\mathbf{TIF}^* - \mathbf{TIF}^0|, \text{ where } \mathbf{TIF^0} = (0, 0, \dots, 0)$$



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

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

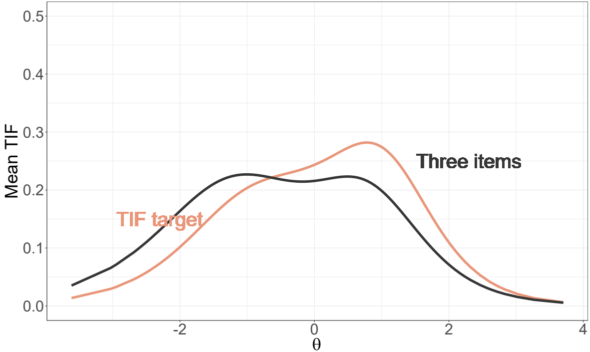
$$\theta_{target} := \arg\max |\mathbf{TIF}^* - \mathbf{TIF}^1|$$



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

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

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



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

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

Brute Force Procedure – BFP

Item bank B

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

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

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

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

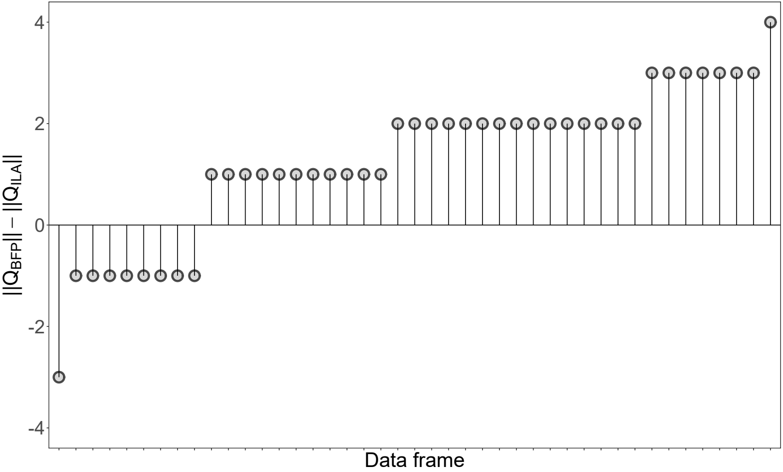
100 data frames:

1. Generate an item bank $||B|| = 6$ 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*

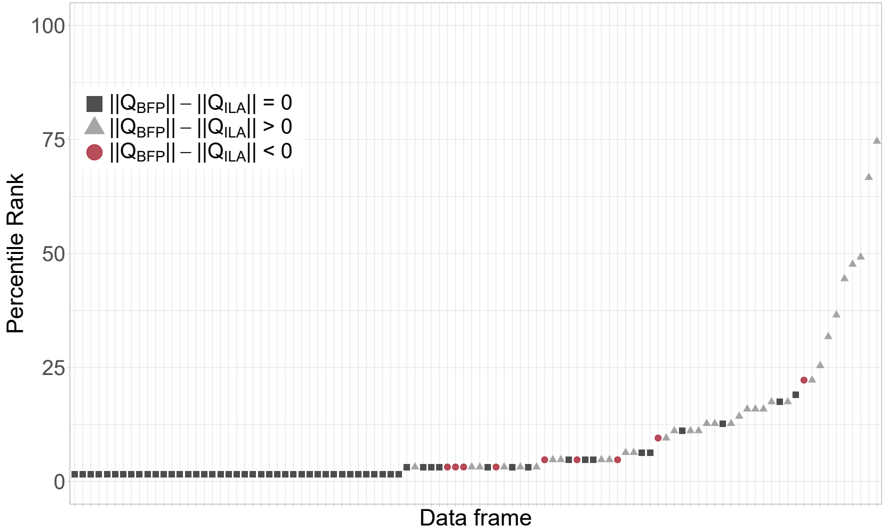
Comparison:

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

$||Q_{\text{BFP}}|| - ||Q_{\text{ILA}}|| = 0$ in 57% of cases

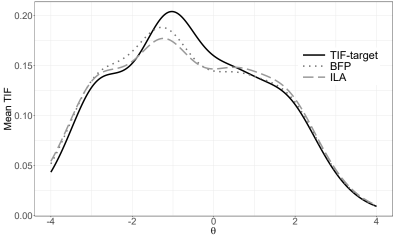


Distance

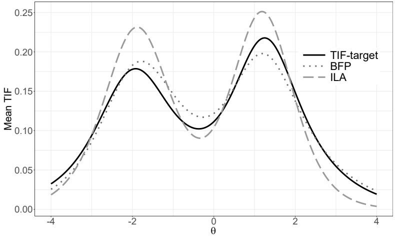

100
75
50
25
0
Percentile Rank
Data frame
||QBFP|| – ||QILA|| = 0
||QBFP|| – ||QILA|| > 0
||QBFP|| – ||QILA|| < 0

TIF comparison

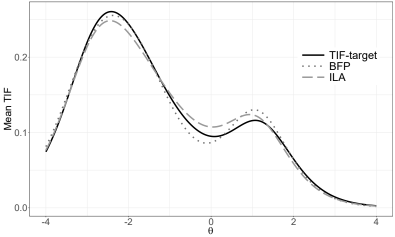
$||Q_{\text{BFP}}|| = ||Q_{\text{ILA}}||$, $Q_{\text{BFP}} \neq Q_{\text{ILA}}$, $RP = 3.17$



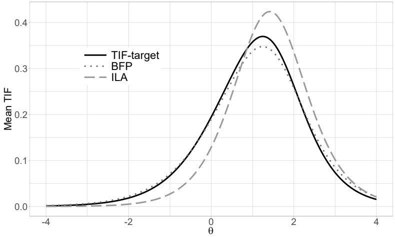
$||Q_{\text{BFP}}|| > ||Q_{\text{ILA}}||$, $RP = 4.76$



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



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



- IRT models and the information they provide are indeed useful for developing STFs
- There's not a one-fits-all solution, and all depends on:
  - The distribution of the latent trait (e.g., normal vs. skewed)
  - The aim of the assessment
  - External constraints

Open questions:

- Is it convenient?
- Where should we stop?
- Does it work on real data....?