OPTIC

Optimal Calibration of Questions in Computerized Achievement Tests

Optimal item calibration with the R-package optical

Frank Miller, Dept. of Computer and Information Science, Linköping University

IRT workshop, Twente, November 29, 2024

Joint work with Mahmood Ul Hassan, Karolinska Institutet, Stockholm

Funded by the Swedish Research Council, Grant 2019-02706

LINKÖPING UNIVERSITY

Content

- Item calibration method
- Example 1 (small illustrative example)
- Example 2 (realistic example)

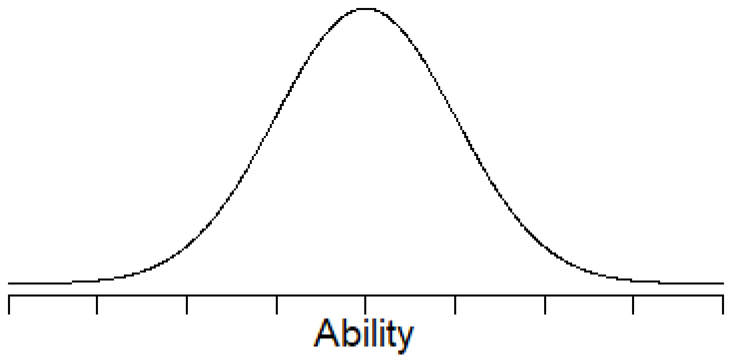
Item calibration method

- A set of new items is given, answers modeled by IRT model (e.g., 2PL/3PL)

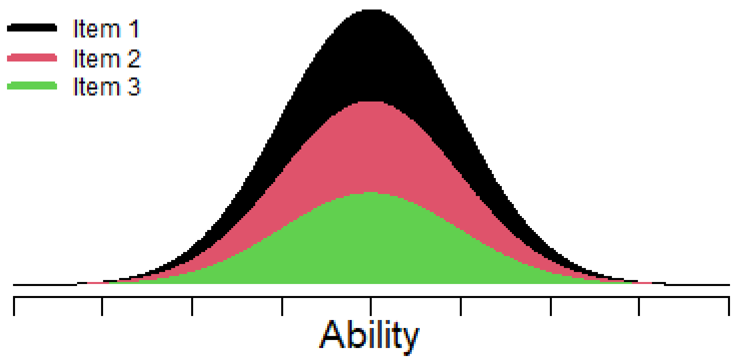
- Item parameters should be estimated in a calibration study, either
  - **Adding** some of the new items **to an operational test**, or
  - **Conducting a specific calibration test**

- Each examinee receives a subset of new items
  - Optimal design: **allocate items based on ability** of examinee such that **precision of parameter estimates is optimized**

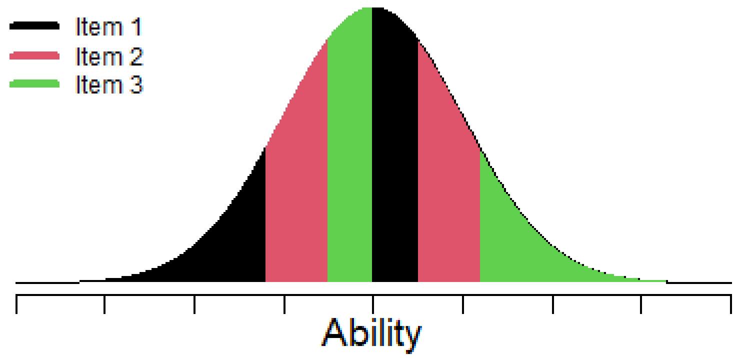
Item calibration method



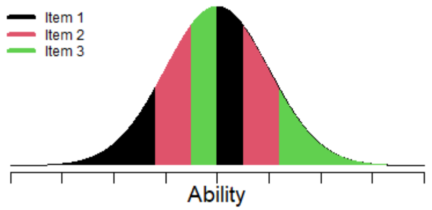
- Population of examinees: standard normal abilities
- Toy-example: Three new items, we can assign one item to each examinee
- Random design



Improved design
(Item 1 = easy, 2 = middle, 3 = difficult)



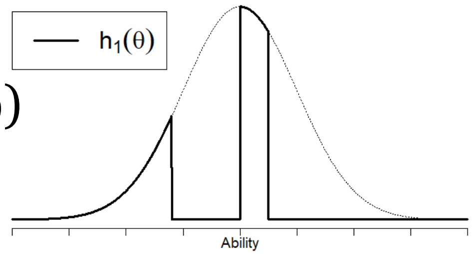
Item calibration method



- Covariance of the estimate for item parameter vector $\beta_i = (a_i, b_i, \ldots)$ is inversely proportional to **information**:

$$M_i = \int p_i(\theta)\big(1 - p_i(\theta)\big)\left(\frac{\partial \eta_i(\theta)}{\partial \beta_i}\right)\left(\frac{\partial \eta_i(\theta)}{\partial \beta_i}\right)^T h_i(\theta)d\theta$$

where $p_i(\theta)$ is the probability of correct answer, $\eta_i(\theta) = \log[p_i(\theta)/\{1 - p_i(\theta)\}]$ and $\boldsymbol{h_i(\theta)}$ **sub-population allocated to item** $\boldsymbol{i}$ (Ul Hassan and Miller, 2019)



- Optimality criteria
  - **D-optimality**: Maximize determinant of information matrix, $\det(M) = \prod \det(M_i)$
  - A-optimality: Minimize average variance for parameters, $\mathrm{trace}(M^{-1}) = \sum \mathrm{trace}(M_i^{-1})$
  - I-optimality: Minimize integrated precision of Item Characteristic Curve estimates

Item calibration method – Requirements

- We need to **know the examinees' abilities**
  - from a separate earlier achievement test, or
  - from earlier operational part of the test

- We need some **prior guesses for item parameters** based on
  - earlier pretest with smaller number of examinees, or
  - expert judgement

R-package **optical**

- Method implemented in R-package **optical** available on CRAN
  - New version will be published there soon
- Call optimization in package optical:

```
R> install.packages("optical")
R> library("optical")
```

optical: Optimal Item Calibration

The restricted optimal design method is implemented to optimally allocate a set of items that require calibration to a group of examinees. The optimization process is based on the method described in detail by Ul Hassan and Miller in their works published in (2019) <doi:10.1177/0146621618824854> and (2021) <doi:10.1016/j.csda.2021.107177>. To use the method, preliminary item characteristics must be provided as input. These characteristics can either be expert guesses or based on previous calibration with a small number of examinees. The item characteristics should be described in the form of parameters for an Item Response Theory (IRT) model. These models can include the Rasch model, the 2-parameter logistic model, the 3-parameter logistic model, or a mixture of these models. The output consists of a set of rules for each item that determine which examinees should be assigned to each item. The efficiency or gain achieved through the optimal design is quantified by comparing it to a random allocation. This comparison allows for an assessment of how much improvement or advantage is gained by using the optimal design approach. This work was supported by the Swedish Research Council (Vetenskapsrådet) Grant 2019-02706.

Version:	1.7.1
Depends:	R (≥ 4.1.0)
Imports:	stats
Published:	2023-05-25
DOI:	10.32614/CRAN.package.optical
Author:	Mahmood Ul Hassan [aut, cre], Frank Miller [aut]
Maintainer:	Mahmood Ul Hassan <scenic555 at gmail.com>
BugReports:	https://github.com/scenic555/optical/issues
License:	GPL (≥ 3)
URL:	https://scenic555.github.io/optical/, https://github.com/scenic555/optical
NeedsCompilation:	no
Language:	en-GB
Materials:	README
CRAN checks:	optical results

Documentation:

Reference manual:	optical.pdf

Example 1

- Assume three items to be calibrated. 2PL models used for them and prior guess for parameters is:
  Discrimination 1.6 for all items, difficulty = -2, 0.5, 2, respectively

```
R> a <- c(1.6, 1.6, 1.6); b <- c(-2, 0.5, 2)
R> ip <- cbind(a, b)
```

Item	a	b
1	1.6	-2
2	1.6	0.5
3	1.6	2

- Call optimization function **`optical()`**:

```
R> yyy <- optical(ip)
R> yyy
```

Example 1

- Output:

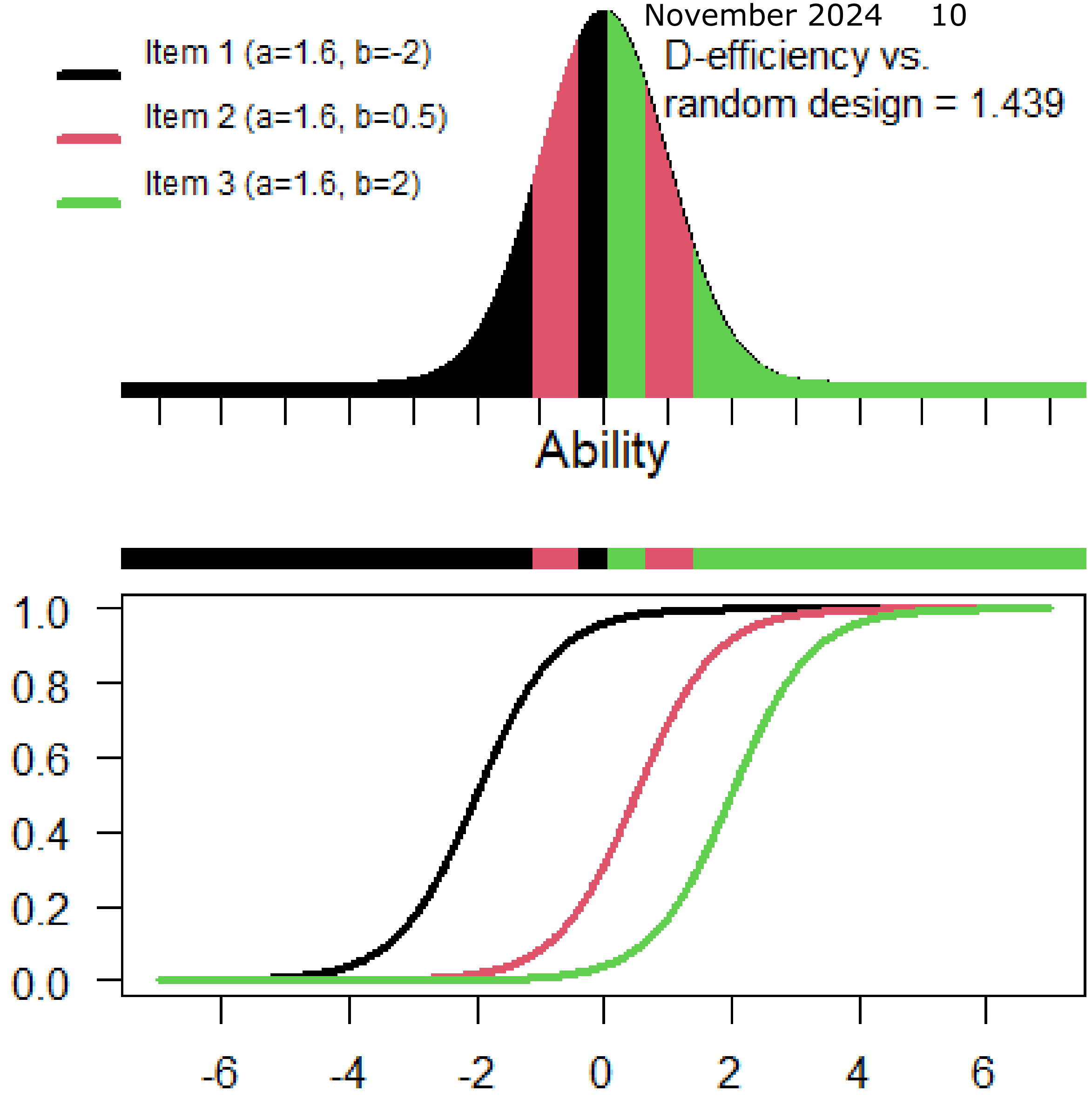
```
Block1
====================================================================
Table of interval boundaries for D-optimal design with items and
probabilities (expected proportion of examinees in this interval)
====================================================================
     Lower   Upper Item Probability
1     -Inf -1.0957    1     0.13662
2  -1.0957 -0.3965    2     0.20925
3  -0.3965  0.0705    1     0.18223
4   0.0705  0.6431    3     0.21183
5   0.6431  1.4099    2     0.18077
6   1.4099     Inf    3     0.07929
```

- Assign examinees with ability >0.0705 and <0.6431 to Item 3

Example 1

```
R> drawdesign(yyy, layout=5)
```

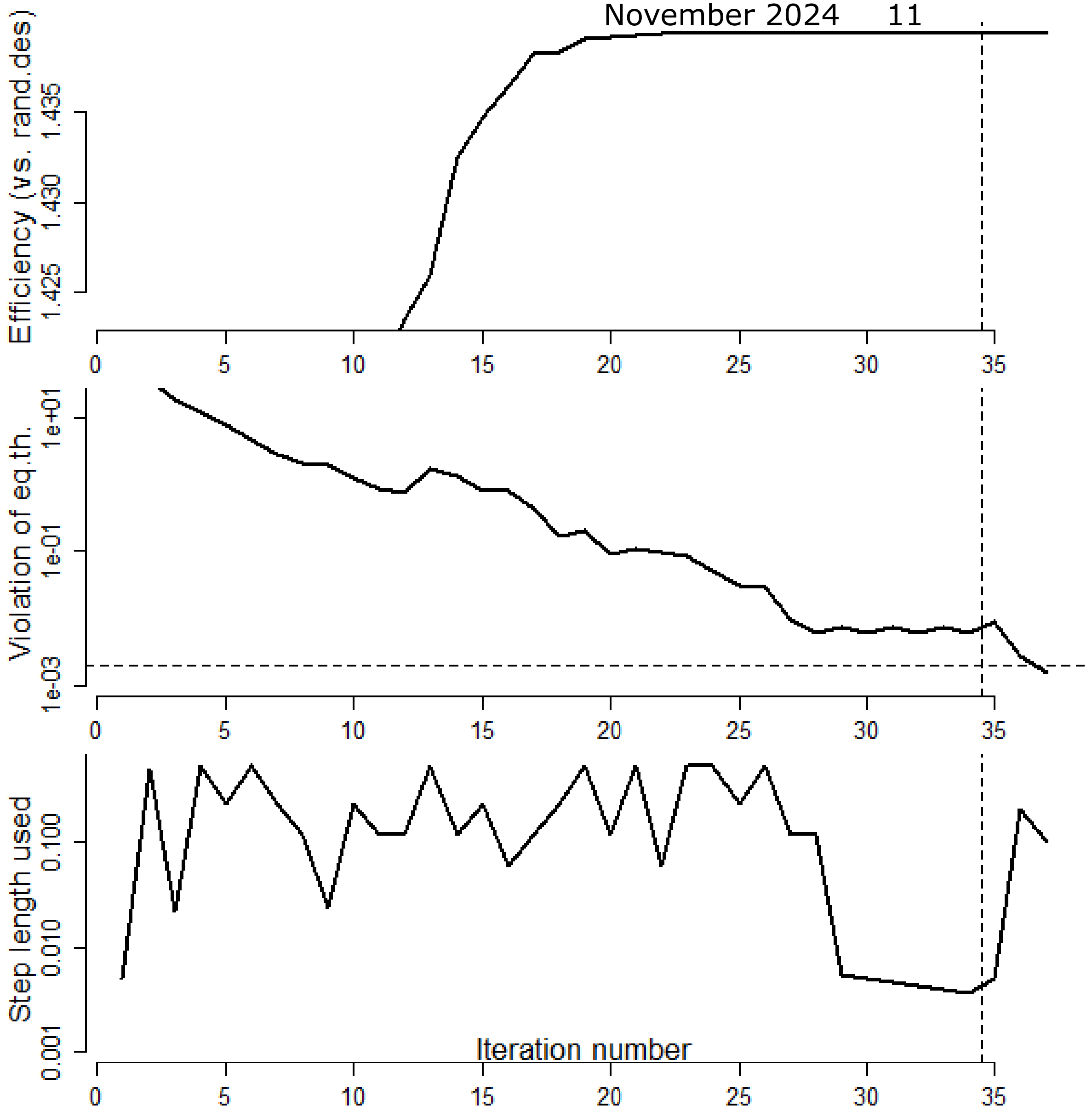
- Visual result:
  - Top: Recommended item allocation (D-optimal design)
  - Middle line: Proof of optimality if coincides with color on top (for convergence check)
  - Bottom: Item characteristic curves (for prior assumptions)
- D-efficiency = 1.439:
  - A random allocation would need 44% more examinees for same precision of parameters



Example 1

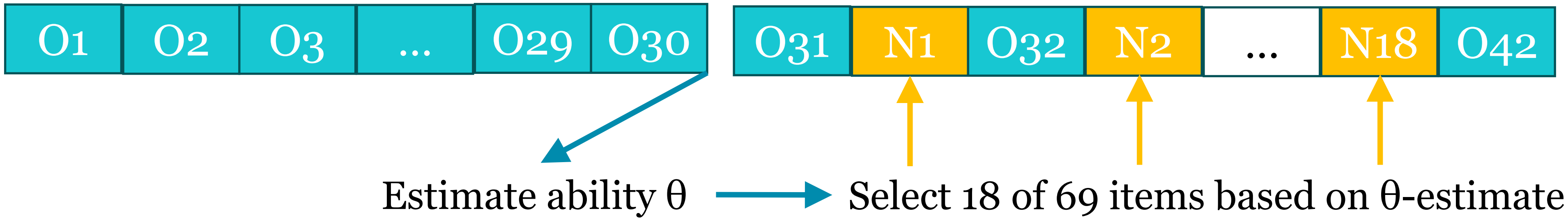
```
R> convergenceplot(yyy)
```

- Monitor convergence of optimization algorithm:
  - Top panel shows that efficiency of computed design plateaus after 20 iterations
  - Middle panel shows that gap to optimal design converges to 0
  - Lower panel shows a step size used (no special behaviour, here)

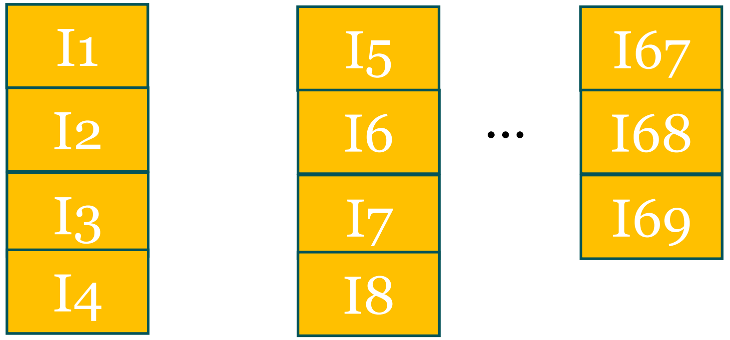


Example 2 – test situation

- **Operational test with 42 items** (O1, …, O42); can add **18 new items** (N1, …)
- We can estimate the examinee's ability based on first 30 items
- Test:



- **69 new items** (I1, …, I69) to be calibrated:
  60 2PL and nine 3PL items
- 15 "blocks" with four 2PL items, and
  3 blocks with three 3PL items each
- Each examinee receives one item of each block



Example 2

- Preliminary item parameter guesses available (see Miller and Fackle-Fornius, 2024), stored in a 69x3-matrix **`ip`**
- Variable **`bid`** identifies blocks

ip	a_i	b_i	c_i	bid
[1,]	1.348	0.454	NA	1
[2,]	1.501	0.983	NA	1
[3,]	2.178	0.332	NA	1
[4,]	1.444	-0.905	NA	1
[5,]	2.097	-0.072	NA	2
[6,]	0.822	0.720	NA	2
[7,]	1.511	-0.232	NA	2
[8,]	1.375	1.245	NA	2
...				...
[60,]	0.662	-2.291	NA	15
[61,]	0.744	1.289	0.001	16
[62,]	2.541	-0.775	0.274	16
[63,]	1.883	1.035	0.246	16
[64,]	0.753	0.721	0.001	17
[65,]	1.925	-0.280	0.001	17
[66,]	2.337	0.807	0.292	17
[67,]	4.972	0.657	0.426	18
[68,]	0.147	14.163	0.001	18
[69,]	1.005	-1.610	0.003	18

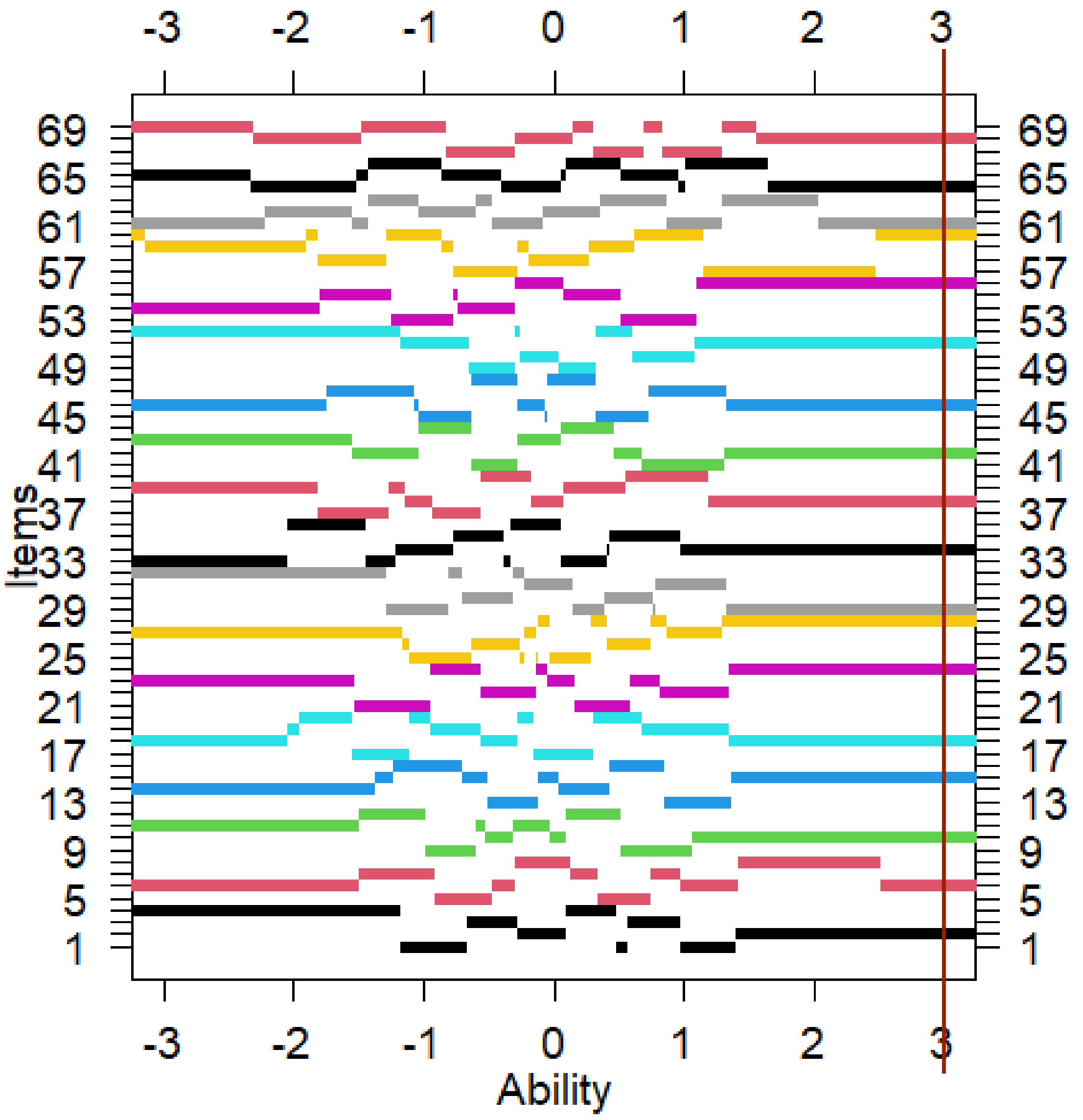
Example 2

```
R> yyy <- optical(ip, bid=bid)
R> drawdesign_allitems(yyy, linew=5,
                       ablim=3, colvec=bid)
```

- Blocks are shown in same color here

```
R> calitems(yyy, abil=3)
[1]  2  6 10 15 18 24 28 29 34
38 42 46 51 56 60 61 64 68
```

- E.g., an examinee with ability=3 receives calibration item 2, 6, …, 68



References

- Miller F, Fackle-Fornius E (2024). Parallel optimal calibration of mixed-format items for achievement tests. *Psychometrika*, 89, 903–928.
- Ul Hassan M, Miller F (2023). optical: Optimal Item Calibration. R package. https://cran.r-project.org/web/packages/optical/optical.pdf
- Ul Hassan M, Miller F (2019). Optimal item calibration for computerized achievement tests. *Psychometrika*, 84, 1101-1128.

Contact

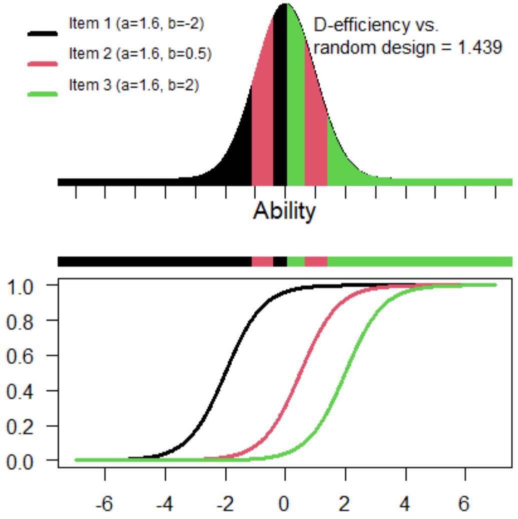
- frank.miller@liu.se

Appendix

Example 1

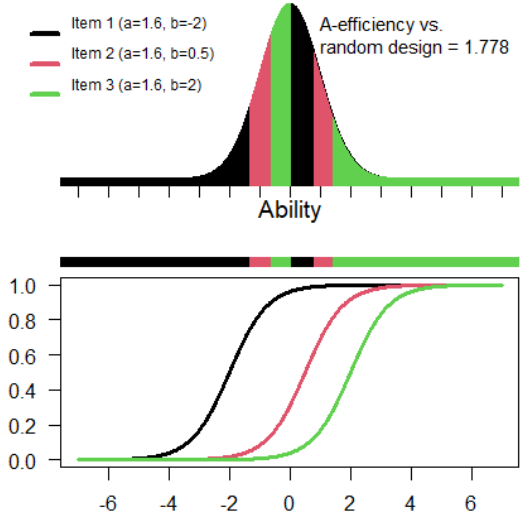
- D-optimal design:

Lower	Upper	Item	Probability
-Inf	-1.0957	1	0.13662
-1.0957	-0.3965	2	0.20925
-0.3965	0.0705	1	0.18223
0.0705	0.6431	3	0.21183
0.6431	1.4099	2	0.18077
1.4099	Inf	3	0.07929



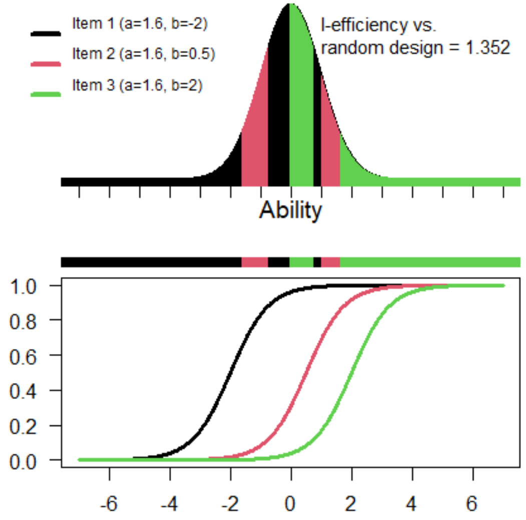
A-optimal design:

Lower	Upper	Item	Probability
-Inf	-1.32050	1	0.09333
-1.32050	-0.61840	2	0.17482
-0.61840	0.04881	3	0.25131
0.04881	0.80365	1	0.26974
0.80365	1.41950	2	0.13292
1.41950	Inf	3	0.07788



I-optimal design:

Lower	Upper	Item	Probability
-Inf	-1.61350	1	0.05332
-1.61350	-0.73876	2	0.17671
-0.73876	-0.05008	1	0.25000
-0.05008	0.74543	3	0.29196
0.74543	0.99977	1	0.06930
0.99977	1.62850	2	0.10700
1.62850	Inf	3	0.05171



optical function

- Usage

```
optical(
  ip,
  bid = NULL,
  oc = "D",
  uncert = FALSE,
  ipop,
  imf = c(0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.45),
  maxiter = rep(300, 6),
  eps = rep(0.002, 6),
  nnn = c(0, 50, 50, 200, 200, 200),
  nsp = c(0.001, 1e-04, 1e-04, 1e-05, 1e-05, 1e-05),
  sss = 0.001,
  falpha = 1.08,
  sdr = TRUE,
  ig = 3,
  ex = 0,
  integ = TRUE,
  show_progress = 1
)
```

Example 2 – design efficiency

- Efficiency of proposed optimal design versus random design:

```
efficiency(yyy)
```

- General function call:

```
efficiency(
  yyy,
  uncert = FALSE,
  ipop,
  oc = "D",
  L = NULL,
  items = FALSE,
  integ = TRUE
)
```

```
$block1
[1] 1.221323

$block2
[1] 1.18604

$block3
[1] 1.285204
…

$block13
[1] 1.458904
…

$block15
[1] 1.165814
…
```

Result is efficiencies by blocks, smallest efficiency +16.6%, largest efficiency +45.9% versus random design.

Example 2 – item-efficiency

- Efficiency of proposed optimal design versus random design for all items:

```
> efficiency(yyy, items=TRUE)
$block1
              item 1   item 2   item 3   item 4       tot
crit_od     5.976140 6.118099 5.764875 5.776016 23.635131
crit_rand   6.178970 6.474586 6.171032 6.410024 25.234611
eff         1.106736 1.195116 1.225168 1.373008  1.221323

$block2
              item 5   item 6   item 7   item 8      tot
crit_od     5.653314 6.414538 5.805036 6.226356 24.09924
crit_rand   6.096397 6.588512 6.090166 6.689132 25.46421
eff         1.247999 1.090882 1.153228 1.260348  1.18604

$block3
```

References

- Miller F, Fackle-Fornius E (2024). Parallel optimal calibration of mixed-format items for achievement tests. *Psychometrika*, 89, 903–928.
- Ul Hassan M, Miller F (2021). An exchange algorithm for optimal calibration of items in computerized achievement tests. Computational *Statistics and Data Analysis*, 157: 107177.
- Ul Hassan M, Miller F (2025). optical: An R Package for Optimal Calibration of Items in Achievement Tests. Manuscript.
- Ul Hassan M, Miller F (2023). optical: Optimal Item Calibration. R package. https://cran.r-project.org/web/packages/optical/optical.pdf
- Ul Hassan M, Miller F (2019). Optimal item calibration for computerized achievement tests. *Psychometrika*, 84, 1101-1128.