Introduction
Computerized adaptive tests
Simulations and examples
Conclusion

Lesson 11: Computerized Adaptive Testing

Patrícia Martinková

Department of Statistical Modelling
Institute of Computer Science, Czech Academy of Sciences

Institute for Research and Development of Education
Faculty of Education, Charles University, Prague

NMST570, December 18, 2018

Outline

1 Introduction

2 Computerized adaptive tests

3 Simulations and examples

4 Conclusion

Motivation

- Student abilities may differ remarkably
  - Some students may become bored by too easy items
  - Some students may become stressed by too hard items

- Selection of proper items may save time

Different types of tests

- Linear tests
  - most popular
  - test takers take all items
- Linear Computer-based tests (CBT)
  - linear test administered on computer
  - allow for taking track of response time for each item
- Computerized adaptive tests (CAT)
  - gets estimate after each item
  - guides selection of subsequent items
- Multistage test (MST)
  - compromise between Linear and CAT,
  - similar to CAT, but by groups of items

Computerized adaptive tests: components

1 Initialization
- administer initial item(s)
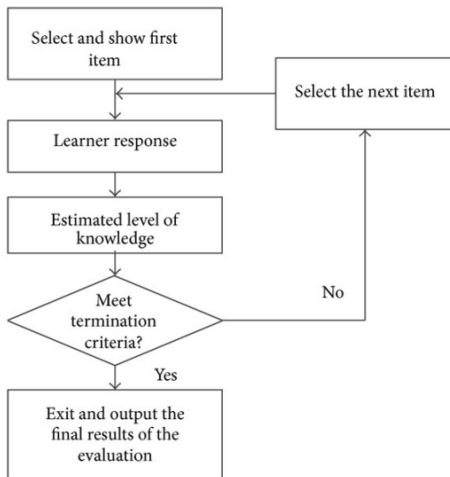- estimate student ability

2 Testing cycle
- Select item and gather response
- Re-estimate student ability
- Check termination criteria

3 Output the final ability estimate / final classification

Further possibilities:
- Exposure rate control
- Content balancing

CAT flow chart

Item Bank

- Based on test specifications
- Sufficient number of items in each content category
- Item quality reviewed by content experts
- Newly written items pretested
- Reviewed items calibrated with CTT or IRT (parameters a, b, c, d)
- Qualified items selected to item bank
- Item bank re-evaluated for size, specifications, content balance

Item properties described by IRT model

- Selected IRT model
  - Rasch, 1PL, 2PL, 3PL, 4PL, (G)PCM, GRM, NRM,...
- Functional form (here for 2PL IRT model)

$$P(Y_{ij} = 1|\theta_i, a_j, b_j) = \frac{e^{a_j(\theta_i - b_j)}}{1 + e^{a_j(\theta_i - b_j)}}$$

a_j discrimination, b_j difficulty
- Item parameter estimates
  - JML, CML, MML, Bayesian methods
  - Item Characteristic Curve (ICC)
  - Item Information Curve (IIC)
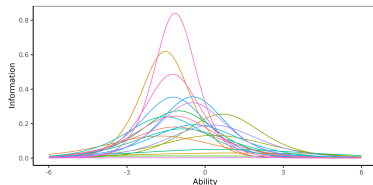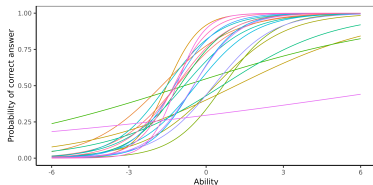
Initialization

- Initial estimate of student ability
    - Mean ability ($\hat{\theta}_0 = 0$)
    - Based on a single item
    - Based on a pretest

- Item selection for initial estimate of ability / pretest
    - Item(s) with mean difficulty (b close to 0)
    - Item(s) with difficulty closest to ability estimated by pretest
    - Random item(s)

Ability Estimation

- Maximum likelihood (MLE)
- Bayes modal estimator, maximal a posteriori estimator (MAP)
- Expected a posteriori estimator (EAP)

Introduction
○○

Computerized adaptive tests
○○○○○○●○○

Simulations and examples
○○○

Conclusion
○○○

Item Selection

- Sequential (non-adaptive)
- Urry's criterion: difficulty closest to current ability estimate
- Fisher Information
    - maximum Fisher information (MFI)
    - maximum likelihood weighted information criterion (MLWI)
    - maximum posterior weighted criterion (MPWI)
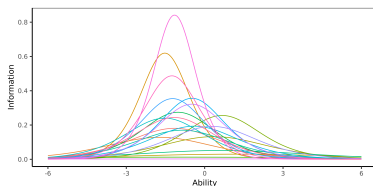
- Kullback-Leibler information criteria
- Maximum expected information criterion (MEI)

Other Issues

- Exposure rate control
  - randomesque
  - 5-4-3-2-1 method
  - etc.

- Content balancing

Stopping rule

- Test length reached
- Test time reached
- Precision criterion on ability estimate
- Classification criterion: threshold met specification for ability confidence interval

Introduction
00

Computerized adaptive tests
000000000

Simulations and examples
●00

Conclusion
000

Post-hoc analysis of real data

- Having real data from linear test
- Checking how would the results be in CAT design
- Comparison of different settings

Simulation study

- Simulating ability
- Setting item parameters (from real test or based on a model)
- Comparison of different settings
- Correlation with true ability, bias, length,...

CAT Examples and software

- `catIRT`
- `catR` + Concerto
- `mirtCAT`

Conclusion

- CAT is an up-do-date testing method!
  - Good precision using much less items
  - Increased security
  - Faster score reporting
  - Less stress and less boredom for students

Thank you for your attention!

`www.cs.cas.cz/martinkova`

References

- Magis D, Yan D, von Davier AA. (2017). *Computerized Adaptive and Multistage Testing with R.* Springer.

- Magis D, & Raichle G. (2012). Random Generation of Response Patterns under Computerized Adaptive Testing with the R Package catR. *Journal of Statistical Software*, 48, pp. 1-31. https://www.jstatsoft.org/article/view/v048i08

- Chalmers P. (2016). Generating Adaptive and Non-Adaptive Test Interfaces for Multidimensional Item Response Theory Applications. *Journal of Statistical Software*, 71, pp. 1-38. https://www.jstatsoft.org/article/view/v071i05

- Chalmers P. (2017-08-11). Wiki mirtCAT https://github.com/philchalmers/mirtCAT/wiki