UNIVERSITY OF TWENTE.

On the develoment of a Bayesian CAT

Bernard Veldkamp
RCEC, University of Twente

General plan for CAT development

1. Develop item bank
2. Design CAT algorithm
3. Pretesting
4. Operational CAT
5. Monitoring

Bayesian CAT

- Bayesian procedures are being applied for:
  - Item bank design
  - the various steps of CAT
    - Item selection
    - Ability estimation
    - Exposure control

Bayes rule

posterior ~ prior * likelihood

Priors in Bayesian CAT

- Generally, a standard normal N(0,1) prior is defined for the ability parameter.
- Especially useful for short CATs.

Empirical priors

- Collateral information
- Reduces number of observations needed

Item pool development

- Major drawback of CAT: costs of item pool development.
- Number of items needed: 12 times test length.
- Large calibration samples

Item pool development

- Empirical Bayesian priors can be applied to reduce the number of candidates needed for pre-testing the items.
  - Cloning
  - Task models
  - Regression trees

Example: number series

1 2 4 8 ?

a) 12

b) 16

c) 20

Item properties

- Radicals
  - {addition, subtraction, multiplication, division}
  - {level 1, level 2}
- Incidentals
  - Starting value
  - Numbers involved in various steps.

Example

Radical:

- Multiplication

Incidentals:

- Starting value: 1
- Number value of multiplication: 2

Item pool development

- Empirical Bayesian priors can be applied to reduce the number of candidates needed for pre-testing the items.
  - Cloning
  - Task models
  - Regression trees

Veldkamp, B. P., & Matteucci, M. (2013). Bayesian computerized adaptive testing. *Revista Ensaio: Avaliação e Política Públicas em Educação, 21,* 57-82.

Priors about person parameter

- Empirical information about the candidtes
  - Earlier test administrations
  - Information coming from related constructs

Example: HADS

- Hospital Anxiety and Depression Scale
- Administered four times
- Power priors

$$\text{posterior} \sim \text{prior} * (\text{history})^{\alpha} * \text{likelihood}$$

EVERYBODY CONVINCED?

Eric Bradlow:

"THAT IS ALL VERY NICE, BUT IN TRULY BAYESIAN CAT THE ITEM PARAMETERS HAVE DISTRIBUTIONS INSTEAD OF FIXED VALUES".

The problem:

- Item banking
- Estimated item parameters
- Item parameters are assumed to be fixed

The problem:
How to deal with uncertainties in the item parameters?

That problem is not new!

Hambleton & Jones (1994)

- The problem is big
- Item selection capitalizes on chance
- The amount of information in a test might be overestimated by more than 100%.

Olea et al (2012)

- It is a problem.
- However, it can be managed with exposure control.

Research Question:

Can the problem be solved?

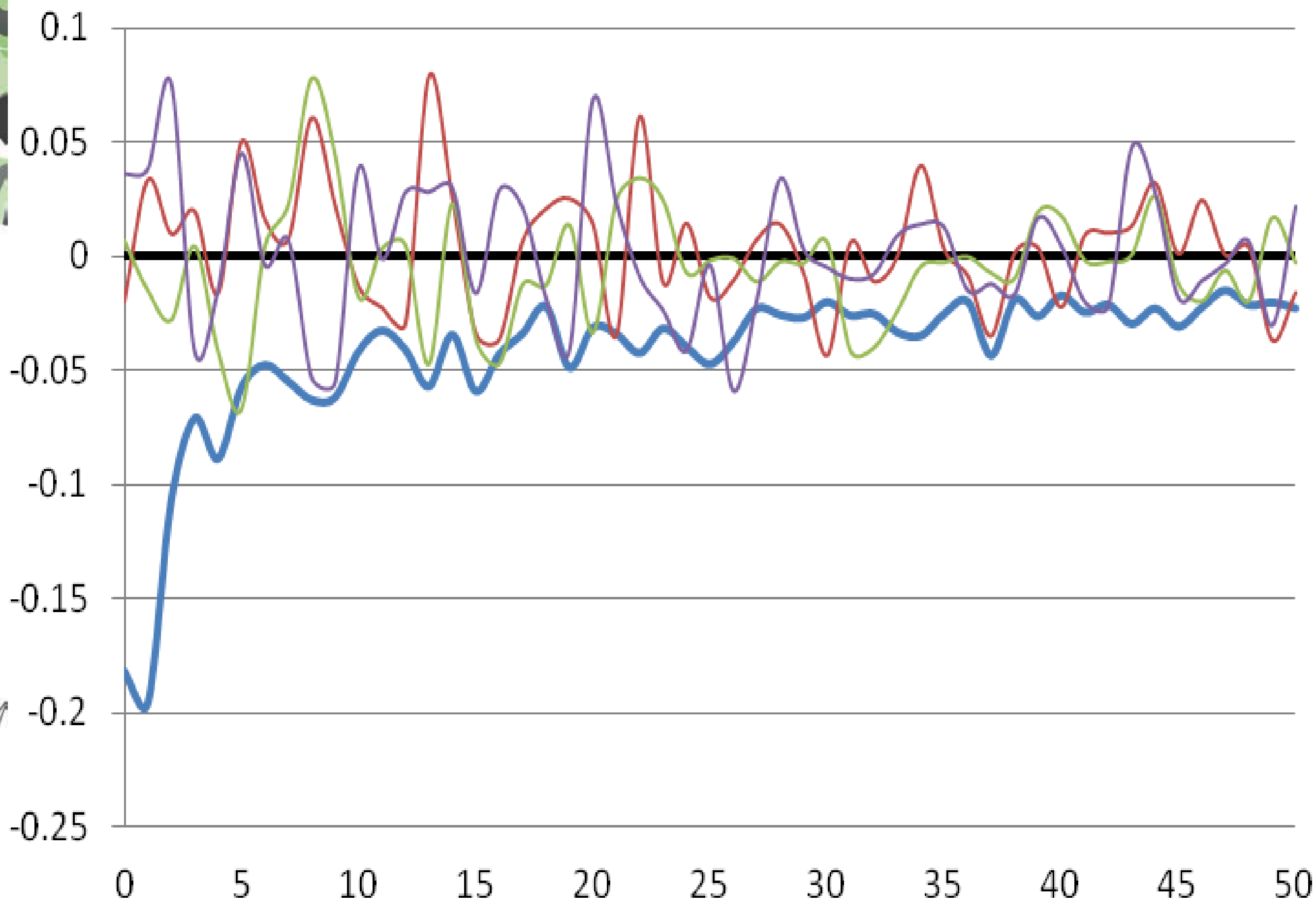
Robust optimization

- Test Assembly can be seen as an optimization problem with uncertainty in the parameters:
  - Soyster (1973)
  - Bertsimas & Sim (2003)
  - De Jong, Steenkamp, & Veldkamp (2009)
  - Veldkamp (2012)

Idea: Apply robust optimization to CAT

Robust item pool

- Expected Response Functions (Lewis,1985)
- Expected Fisher Information
- Measurement error is normally distributed
- Assume that it hits where it hurts most



0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
-0.25
0
5
10
15
20
25
30
35
40
45
50

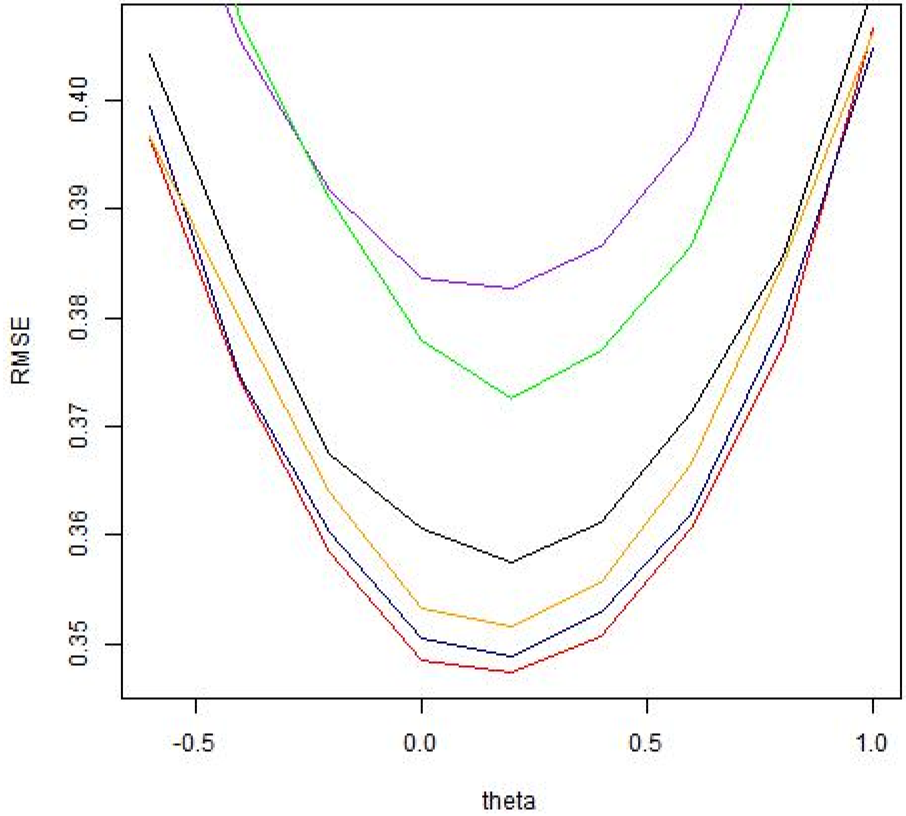
Simulation study

- Item bank consists of 300 items
  - replications of the same parent
  - (a,b,c) = (1.4,0.0,0.2)
- Discrimination parameter a ~ N(1.4,0.05)
- Difficulty parameter b ~ N(0,0.1)
- Guessing parameter c ~ N(0.2,0.02)
- Test length = 20 items
- CatR package in R

Six methods

- No measurement error
- No control
- Soyster's method
- Exposure Control
- Robust item pool
- Robust item pool + EC

Results



Conclusions

Simulation study:

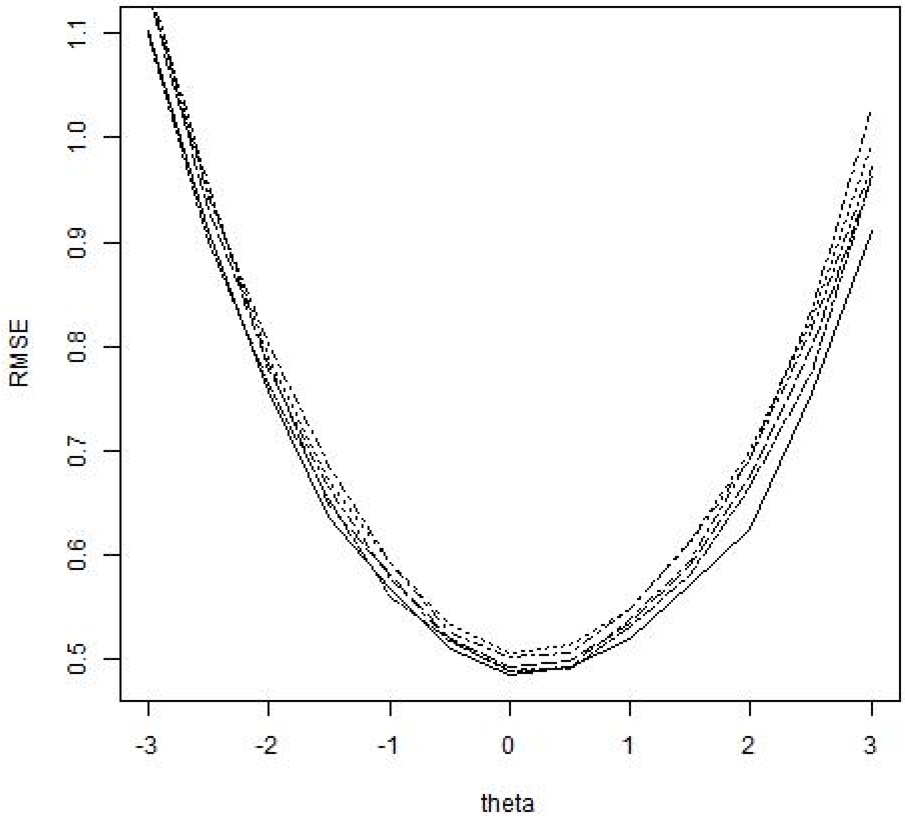
- Problem exists
- EC does not solve the problem
- Soyster is far too conservative
- Robust item pools might do the trick

How about operational CAT?

Real data set

- Item bank consists of 306 items
  - Calibrated with a 3PLM
  - Range of parameters
  - Range of uncertainties
  - Test length = 20 items
- catR package in R

Results



Results

- In line with Olea et al (2012)
- Uncertainty only seems to matter for a couple of items.

Conclusions

- Real data
- Differences are smaller
- Problem is bigger for small tests

Conclusions

- Robust item pools are useful in dealing with uncertainty in the item parameters
- Robust ATA and Exposure Control can be added to further improve the results
- Impact seems to be limited

Further research

- Multi-dimensional CAT
- Quality of Life
- Item bank calibration