Open-source CAT software: R packages and Concerto

David Magis

Department of Psychology, KU Leuven, Belgium
Department of Education, University of Liège, Belgium

Joint work with

Juan Ramon Barrada
Universidad Zaragoza, Teruel, Spain

ISOQOL-NL, Utrecht, March 12, 2014
Outline:

1. CAT software
2. R packages *catR* and *catIrt*
3. Comparisons
4. A demo of *catR*
5. R package *MAT*
6. Platform *Concerto*
7. Final comments
Outline:

1. CAT software
2. R packages catR and catIrt
3. Comparisons
4. A demo of catR
5. R package MAT
6. Platform Concerto
7. Final comments
1. CAT software
1. CAT software

Three main categories of CAT software:

1. Commercial software:
   - CATSim (Assessment Systems Corporation, 2012)
   - Adaptest (https://adaptest.vpgcentral.com/)
   - Assessment Center (https://www.assessmentcenter.net/)
   - ...

2. Open-source software (mostly based on R):
   - Firestar (Choi, 2009)
   - catIrt (Nydick, 2013)
   - catR (Magis & Raiche, 2012)
   - MAT (Choi, 2011)
   - ...

1. CAT software

Three main categories of CAT software:

3. Open-source web-based platform:
   - **Concerto** (Kosinski & Rust, 2011)

Commercial software designed for specific assessment purposes
R packages mostly designed for simulation studies
Web-based platform built for real online CAT assessments
1. CAT software

Purpose of this talk:

- to briefly present several CAT-related software
- to focus on their goals and options

Software to be presented (in detail): \textit{catR} and \textit{catIrt}

R package \textit{MAT} and web-based platform \textit{Concerto} briefly outlined
Outline:

1. CAT software
2. R packages *catR* and *catIrt*
3. Comparisons
4. A demo of *catR*
5. R package *MAT*
6. Platform *Concerto*
7. Final comments
2. **R packages catR and catIrt**

R packages to generate response patterns under CAT scenarios

**Advantages:**
- Flexibility
- Absolute control of all CAT steps (selecting first items, next item selection, proficiency estimation, stopping rules, ...)
- Easy to generate lots of response patterns

**Drawbacks:**
- Some knowledge of R is necessary
- Limited to R

⇒ Most useful for intensive simulation studies
2. R packages *catR* and *catIrt*
2. R packages \textit{catR} and \textit{catIrt}

Currently available versions (on CRAN):

- \textit{catR}: 2.6 (released: December 3rd, 2013)
- \textit{catIrt}: 0.4-1 (released: May 31st, 2013)

Beta version of \textit{catR} 3.0 (released: last week)

Not yet available on CRAN

Holds recent improvements
Outline:

1. CAT software
2. R packages *catR* and *catIrt*
3. Comparisons
4. A demo of *catR*
5. R package *MAT*
6. Platform *Concerto*
7. Final comments
3. Comparisons

Comparisons of catR (2.6), catR (3.0) and catIrt

Comparison criteria:

- IRT models
- Proficiency estimators
- First item(s) selection
- Next item selection
- Stopping rules
- Miscellaneous topics
### 3. Comparisons

**IRT models:**

<table>
<thead>
<tr>
<th>Model</th>
<th>catR 2.6</th>
<th>catR 3.0</th>
<th>catIrt</th>
</tr>
</thead>
<tbody>
<tr>
<td>1PL</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>2PL</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>3PL</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>4PL</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRM</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>M-GRM</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCM</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>GPCM</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RSM</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NRM</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. Comparisons

Proficiency estimators:

- All packages can estimate proficiencies with
  - Maximum likelihood (ML)
  - Bayes modal or Maximum a posteriori (MAP)
  - Expected a posteriori (EAP)
  - Weighted likelihood (WL)

- Prior distributions for MAP and EAP:

<table>
<thead>
<tr>
<th>Model</th>
<th>catR 2.6</th>
<th>catR 3.0</th>
<th>catIrt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Uniform</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Jeffreys</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>
3. Comparisons

First item(s) selection:

- \textit{catR} and \textit{catIrt} can select one or several items to start the CAT
- Selection rules:
  - at random
  - selected item(s) by the user (\textit{catR} only)
  - with respect to starting proficiency levels
  - with respect to optimal item selection
- Optimal selection based on maximum information (\textit{catR}) or more sophisticated versions (\textit{catIrt})
3. Comparisons

Next item selection:

<table>
<thead>
<tr>
<th>Model</th>
<th>catR 2.6</th>
<th>catR 3.0</th>
<th>catIrt</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFI</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>MLWI</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>MPWI</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>MEI</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MEPV</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

- **MFI**: Maximum Fisher information
- **MLWI**: Maximum likelihood weighted information
- **MPWI**: Maximum posterior weighted information
- **MEI**: Maximum expected information
- **MEPV**: Minimum expected posterior variance
## 3. Comparisons

Next item selection:

<table>
<thead>
<tr>
<th>Model</th>
<th>catR 2.6</th>
<th>catR 3.0</th>
<th>catIrt</th>
</tr>
</thead>
<tbody>
<tr>
<td>KL</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>KLP</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Progressive</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Proportional</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At random</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

- **KL**: Kullback-Leibler information
- **KLP**: Posterior Kullback-Leibler information
3. Comparisons

Stopping rules:

• Three kinds of stopping rules:
  – **Length**: stop after $n$ items administered
  – **Precision**: stop when SEM of proficiency small enough
  – **Classification**: stop when classification is possible

• Classification rules:
  – Confidence interval method (Kingsbury & Weiss, 1983) in all packages
  – Sequential Probability Ratio Test (SPRT) (Eggen, 1999) only in *catIrt*
  – Generalized Likelihood Ratio (GLR) (Thompson, 2009) only in *catIrt*
3. Comparisons

Miscellaneous topics:

• All software allow for item exposure control:
  – Sympson-Hetter method (Sympson & Hetter, 1985) in \textit{catIrt}
  – Randomesque method (Kingsbury & Zara, 1989) in \textit{catR}
  – Progressive and proportional methods (for next item selection) in \textit{catR 3.0}

• \textbf{Content balancing} (Kingsbury & Zara, 1989) is allowed in \textit{catR}

• Several response patterns can be generated with one run in \textit{catR 3.0} and \textit{catIrt}

• Graphical output available
Outline:

1. CAT software
2. R packages *catR* and *catIrt*
3. Comparisons
4. A demo of *catR*
5. R package *MAT*
6. Platform *Concerto*
7. Final comments
4. A demo of *catR*

Small demo of the logic and functioning of *catR 3.0*

CAT design:

- Creation of a 200-item bank calibrated under Partial Credit Model (and at most 5 response categories per item)
- Generation of one CAT response pattern for true ability level $\theta = 0$
- 3 initial items selected randomly in the bank
- Next item selection by maximum Fisher information
- Interim ability estimation by weighted likelihood
- Stopping rule: stop when 10 items are administered
- Final ability estimation by maximum likelihood
4. A demo of \textit{catR}

\textbf{R code:}

\begin{verbatim}
R> bank <- genPolyMatrix(nrItems = 200, nrCat = 5, 
                        model = "PCM")
R> head(bank)

          deltaj1  deltaj2  deltaj3  deltaj4
1    -0.6260   0.1844     NA     NA
2    -0.8360  1.5950  0.3299 -0.8199
3     0.4867   0.7381  0.5759     NA
4    -0.3048  1.5115  0.3895 -0.6208
5    -2.2146   1.1249     NA     NA
6    -0.0445 -0.0156  0.9437     NA
\end{verbatim}
4. A demo of *catR*

R code:

```r
R> start <- list(nrItems = 3, seed = 1)
R> test <- list(method = "WL", itemSelect = "MFI")
R> stop <- list(rule = "length", thr = 10)
R> final <- list(method = "ML")

R> res <- randomCAT(trueTheta = 0, itemBank = bank,
                     model = "PCM", start = start,
                     test = test, stop = stop,
                     final = final)
```
4. A demo of \textit{catR}

Output:

Random generation of a CAT response pattern

Item bank calibrated under Partial Credit Model

True ability level: 0

Starting parameters:
  Number of early items: 3
  Early items selection: Random selection in item bank
  Items administered: 54, 75 and 114

Adaptive test parameters:
  Next item selection method: maximum Fisher information
  Provisional ability estimator: Weighted likelihood estimator
4. A demo of *catR*

Output:

Stopping rule:
   Stopping criterion: length of test
   Maximum test length: 10 items

Item exposure control:
   Method: ’randomesque’
   Number of ’randomesque’ items: 1

Content balancing control:
   No control for content balancing
4. A demo of *catR*

Output:

Adaptive test details:

<table>
<thead>
<tr>
<th>Nr</th>
<th>Item</th>
<th>Resp.</th>
<th>Est.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>54</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>1</td>
<td>75</td>
<td>0</td>
<td>0.264</td>
<td>0.332</td>
</tr>
<tr>
<td>2</td>
<td>114</td>
<td>1</td>
<td>0.111</td>
<td>0.279</td>
</tr>
<tr>
<td>3</td>
<td>159</td>
<td>0</td>
<td>-0.066</td>
<td>0.266</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>0</td>
<td>-0.027</td>
<td>0.232</td>
</tr>
<tr>
<td>5</td>
<td>59</td>
<td>3</td>
<td>-0.012</td>
<td>0.208</td>
</tr>
<tr>
<td>6</td>
<td>110</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nr</th>
<th>Item</th>
<th>Resp.</th>
<th>Est.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>175</td>
<td>2</td>
<td>0.027</td>
<td>0.19</td>
</tr>
<tr>
<td>9</td>
<td>34</td>
<td>1</td>
<td>0.029</td>
<td>0.179</td>
</tr>
<tr>
<td>10</td>
<td>140</td>
<td>3</td>
<td>0.082</td>
<td>0.168</td>
</tr>
</tbody>
</table>
4. A demo of *catR*

Output:

Final results:
- Length of adaptive test: 10 items
- Final ability estimator: Maximum likelihood estimator
- Final range of ability values: [-4,4]
- Final ability estimate (SE): 0.079 (0.173)
- 95% confidence interval: [-0.259,0.418]

Output was not captured!
4. A demo of `catR`
Outline:

1. CAT software
2. R packages \textit{catR} and \textit{catIrt}
3. Comparisons
4. A demo of \textit{catR}
5. R package \textit{MAT}
6. Platform \textit{Concerto}
7. Final comments
5. **R package MAT**

Packages *catR* and *catIrt* make use of usual (dichotomous and polytomous) IRT models.

Package *MAT* (for *Multidimensional Adaptive Testing*; Choi, 2011) incorporates the multidimensional 3-parameter logistic model (Reckase, 2009).

Promising for administering multidimensional CATs.

(Apparently) not yet fully operational:

- Currently less options than *catR* and *catIrt*.
- Item response matrix to be provided ⇒ post-hoc simulations?

⇒ To be followed up.
Outline:

1. CAT software
2. R packages *catR* and *catIrt*
3. Comparisons
4. A demo of *catR*
5. R package *MAT*
6. Platform *Concerto*
7. Final comments
6. Platform *Concerto*

*catR*, *catIrt* and *MAT* mostly designed to generate response patterns with given calibrated item bank

Item bank must be pre-calibrated and provided only in terms of item parameters

No formal item administration with these software

⇒ Need for software for “real” CAT administrations
6. Platform *Concerto*

*Concerto* is a web-based platform for CAT building and administration.

Open-source platform with *catR* package as underworking routine.

*catR* used for proficiency level estimation, first item(s) and next item selection.

*Concerto* provides html menus and webpages to

- build items and create item banks
- incorporate additional item material (e.g. pictures, videos)
- specify the CAT options

CAT can be delivered through a variety of web browsers and on mobile devices (smartphones, tablets, ...).
6. Platform *Concerto*

**Open-source Online R-based Adaptive Testing Platform**

Concerto is an open-source testing platform that allows users to create various online assessments, from simple surveys to complex IRT-based adaptive tests.

Concerto is free for academic and commercial use!

**Demos**

1. Basic Adaptive Demo Test
2. Music Taste Questionnaire by Arielle Bonneville-Roussy
3. 100-item Personality Test by Koen de Couck
4. 20-item Personality Test by Luning Sun.

**Why use Concerto?**

- **Open source** - You can both use Concerto and host your test on our server free of charge. The platform's open-source components are continuously updated and improved by our growing community of users.
- **Computerised Adaptive Testing** - You can use the powerful R engine to apply virtually any Item Response Theory and Computerised Adaptive Testing model.
6. Platform *Concerto*

*Concerto* is built in two parts:

- **Concerto web panel**: online application
  - to prepare **R-based scripts** containing test logic,
  - that provides rich text editor,
  - interacting with **databases** for storing test items and their results,
  - that exports prepared **tests** to the Test Runner

- **Test Runner**: responsible for executing prepared tests and interacting with the user:
  - keeps the test state for the duration of the session, enabling resuming the test (in case of communication failure),
  - to be deployed on multiple machines to serve large numbers of participants
6. **Platform *Concerto***
6. Platform *Concerto*
6. **Platform** *Concerto*

![Table query form](image)
Outline:

1. CAT software
2. R packages *catR* and *catIrt*
3. Comparisons
4. A demo of *catR*
5. R package *MAT*
6. Platform *Concerto*
7. Final comments
7. Final comments

Recent CAT software mostly based on R development

- as packages (*catR* and *catIrt*)
- as front end-user interfaces with back-end R routines such as *Firestar* (Choi, 2009)
- as web-based platforms for CAT delivery with back-end R routines (*Concerto*)

Both *catR* and *Concerto* are under development

Further beneficial updates of *Concerto*: inclusion of user interfaces to simplify item bank and test developments

ERC starting grant application related to *Concerto* updates (soon) submitted
7. Final comments

\textit{catR} and \textit{catIrt} available (freely) from CRAN:

http://cran.r-project.org

\textit{Concerto} can be accessed from

http://www.psychometrics.cam.ac.uk/page/338/concerto-testing-platform.htm
References


References


References

To get the slides:

http://hdl.handle.net/2268/163595

To get the beta version of catR 3.0:

david.magis@ppw.kuleuven.be

or

david.magis@ulg.ac.be
To get the slides:

http://hdl.handle.net/2268/163595

To get the beta version of *catR 3.0*:

david.magis@ppw.kuleuven.be

or

david.magis@ulg.ac.be

THANK YOU!
To get the slides:

http://hdl.handle.net/2268/163595

To get the beta version of \textit{catR 3.0}:

david.magis@ppw.kuleuven.be

or

david.magis@ulg.ac.be

BEDANKT!