Analyzing paired-comparison data in R using probabilistic choice models

Florian Wickelmaier

The R User Conference, August 12-14, 2008
Overview

1. Probabilistic choice models
2. Survey: perceived health risk of drugs
3. Conclusions
Probabilistic choice models

Goal: Scaling of psychological attributes

Procedure:

Participants are not asked to provide a numerical judgment (e.g., on a rating scale), but their behavior in a choice situation is observed. Scaling follows from modeling the data.

- Psychological theory of decision making
- Easy task for participants: pairwise comparison between alternatives, avoiding “scale usage heterogeneity”
- Measurement-theoretical foundation: testable conditions for numerical representation, unique scale level
Probabilistic choice models: applications

Main areas of application: consumer research, opinion surveys, sensory evaluation, psychophysical scaling

- Decision between insurance packages (McGuire & Davison, 1991, \( N = 14000 \))
- Political choice (Tversky & Sattath, 1979)
- Ranking of universities (Dittrich et al., 1998)
- Experimental perception research:
  - Measurement of pain (Matthews & Morris, 1995)
  - Taste, food quality (Bradley & Terry, 1952; Lukas, 1991; Duineveld et al., 1999)
  - Facial attractiveness (Bäuml, 1994)
  - Unpleasantness of environmental sounds (Ellermeier et al., 2004; Zimmer et al., 2004)
  - Sound quality of reproduction systems (Choisel & Wickelmaier, 2007)
Choice models (1): Bradley-Terry-Luce (BTL) model

Choice of an alternative \((x, y, \ldots)\) is probabilistic and depends on the weight (strength) of the alternative \((u(x), u(y), \ldots)\)

**BTL model equations:**

\[
P_{xy} = \frac{u(x)}{u(x) + u(y)} = \frac{1}{1 + \frac{k \cdot u(y)}{k \cdot u(x)}}
\]

- \(P_{xy}\): probability of choosing alternative \(x\) over \(y\) in a paired comparison
- \(u(\cdot)\): ratio scale of the stimuli
- BTL model very parsimonious: only \(n - 1\) free parameters, \(n = \) number of stimuli
- BTL imposes strong restrictions on the choice probabilities
Independence of irrelevant alternatives (IIA)

Choice between two options is independent of the context provided by the choice set

\[
\frac{P(x, \{x, y\})}{P(y, \{x, y\})} = \frac{P(x, \{x, y, z\})}{P(y, \{x, y, z\})}
\]

**Problem:** similarity between groups of stimuli may cause IIA to fail (Debreu, 1960; Rumelhart & Greeno, 1971; Zimmer et al., 2004; Choisel & Wickelmaier, 2007)

Consequence of IIA: strong stochastic transitivity

\[P_{xy} \geq 0.5, P_{yz} \geq 0.5 \Rightarrow P_{xz} \geq \max\{P_{xy}, P_{yz}\}\]
Choice models (2): “Elimination by aspects” (EBA) (Tversky, 1972)

Alternatives (stimuli) are characterized by various features (aspects)

Choice is based on a hidden (sequential) elimination process:

- Aspects are chosen with a probability proportional to their weight (strength)
- Stimuli without the desired aspects are eliminated from the set of alternatives, until only one stimulus remains

- Only the discriminating aspects influence the decision

→ EBA model does not require context independence (IIA)
→ Bradley-Terry-Luce (BTL) model is a special case of EBA
Elimination by aspects (EBA): model equations

Stimuli $x, y, \ldots$ characterized by a set of aspects $x', y', \ldots$

Probability of choosing $x$ over $y$:

$$P_{xy} = \frac{\sum_{\alpha \in x' \setminus y'} u(\alpha)}{\sum_{\alpha \in x' \setminus y'} u(\alpha) + \sum_{\beta \in y' \setminus x'} u(\beta)}$$

$x' \setminus y'$: aspects belonging to $x$, but not to $y$

$u(\cdot)$: ratio scale of the aspects

Scale value of $x$ equals the sum of the characterizing aspect values

**Example:**

$x' = \{\alpha, \beta, \zeta\}, \ y' = \{\gamma, \delta, \varepsilon, \zeta\} \implies P_{xy} = \frac{u(\alpha)+u(\beta)}{u(\alpha)+u(\beta)+u(\gamma)+u(\delta)+u(\varepsilon)}$
The eba package

- Provides functionality for fitting and testing probabilistic choice models: Bradley-Terry-Luce, elimination by aspects, preference tree, Thurstone-Mosteller

- Key functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>strans</td>
<td>Counting stochastic transitivity violations</td>
</tr>
<tr>
<td>eba</td>
<td>Fitting and testing EBA models</td>
</tr>
<tr>
<td>summary, anova</td>
<td>Extractor functions</td>
</tr>
<tr>
<td>plot, residuals</td>
<td></td>
</tr>
<tr>
<td>group.test</td>
<td>Comparing samples of subjects</td>
</tr>
<tr>
<td>eba.order</td>
<td>Testing within-pair order effects</td>
</tr>
</tbody>
</table>

- Manual

Survey: perceived health risk of drugs

- $N = 192$ stratified by sex and age, 48 in each subgroup
- Task: Which of the two drugs do you judge to be more dangerous for your health?
- Drugs
  - Alcohol
  - Tobacco
  - Cannabis
  - Ecstasy
  - Heroine
  - Cocaine
- Each participant did all $6 \cdot 5/2 = 15$ pairwise comparisons.
- Analyses performed separately in the four subgroups
Descriptive statistics

Aggregate judgments (male participants, younger than 30)

<table>
<thead>
<tr>
<th></th>
<th>Alc</th>
<th>Tob</th>
<th>Can</th>
<th>Ecs</th>
<th>Her</th>
<th>Coc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alc</td>
<td>0</td>
<td>28</td>
<td>35</td>
<td>10</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Tob</td>
<td>20</td>
<td>0</td>
<td>18</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Can</td>
<td>13</td>
<td>30</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Ecs</td>
<td>38</td>
<td>46</td>
<td>45</td>
<td>0</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>Her</td>
<td>44</td>
<td>48</td>
<td>47</td>
<td>47</td>
<td>0</td>
<td>44</td>
</tr>
<tr>
<td>Coc</td>
<td>41</td>
<td>45</td>
<td>48</td>
<td>31</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

Probability of choosing \( x \) over \( y \):

\[
\hat{P}_{xy} = \frac{N_x}{N_x + N_y}
\]

Example:

\[
\hat{P}_{Alc, Tob} = \frac{28}{28 + 20} = 0.58
\]

Counting the number of transitivity violations

```
strans(dat)
violations error.ratio mean.dev max.dev
weak       0  0.0000  0.0000  0.0000
moderate   1  0.0500  0.0417  0.0417
strong     5  0.2500  0.0625  0.1458
---
Number of Tests: 20
```
Fitting a BTL model using the `eba()` function

```r
btl <- eba(dat)
```

Obtaining summary statistics and model tests

```r
summary(btl)
```

...  

Model tests:

|       | Df1 | Df2 | logLik1 | logLik2 | Deviance | Pr(>|Chi|)  |
|-------|-----|-----|---------|---------|----------|-----------|
| EBA   | 5   | 15  | -34.09  | -21.62  | 24.94    | 0.00546 **|
| Effect| 0   | 5   | -284.57 | -34.09  | 500.97   | < 2e-16 ***|
| Imbalance | 1 | 15  | -42.84  | -42.84  | 0.00     | 1.00000   |

AIC: 78.181  
Pearson Chi2: 28.09  

The BTL model does not describe the data adequately  

g^2(10) = 24.94, p < .001.
EBA model with one additional aspect – EBA1

Model structure

\[ A_1 = \{\{\alpha\}, \{\beta, \eta\}, \{\gamma, \eta\}, \{\delta, \eta\}, \{\varepsilon, \eta\}, \{\zeta, \eta\}\} \]

\[ A1 \leftarrow \text{list} (c(1), c(2,7), c(3,7), c(4,7), c(5,7), c(6,7)) \]
\[ \text{ebal1} \leftarrow \text{eba} (\text{dat}, A1) \]

Non-alcohol drugs share a feature that affects decision when comparing them with alcohol.
EBA model with two additional aspects – EBA2

Model structure

\[ A_2 = \{\{\alpha\}, \{\beta, \eta\}, \{\gamma, \eta\}, \{\delta, \eta, \vartheta\}, \{\varepsilon, \eta, \vartheta\}, \{\zeta, \eta, \vartheta\}\} \]

\[
\begin{align*}
\alpha & = .040 \\
\beta & = .005 \\
\gamma & = .007 \\
\delta & = .014 \\
\varepsilon & = .355 \\
\zeta & = .027 \\
\eta & = .015 \\
\vartheta & = .140
\end{align*}
\]

\[ A_2 \leftarrow \text{list}(c(1), c(2, 7), c(3, 7), c(4, 7, 8), c(5, 7, 8), c(6, 7, 8)) \]
\[ \text{eba2} \leftarrow \text{eba}(\text{dat}, A_2) \]

Three of the non-alcohol drugs share a feature that comes into play only when comparing them with the other drugs.
Model selection

Nested models can be compared using likelihood ratio tests.

\[
\text{anova}(btl, \ eba1, \ eba2)
\]

<table>
<thead>
<tr>
<th>Model</th>
<th>Resid. df</th>
<th>Resid. Dev</th>
<th>Test Df</th>
<th>LR stat.</th>
<th>Pr(Chi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>btl</td>
<td>10</td>
<td>24.94225</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>eba1</td>
<td>9</td>
<td>17.54611</td>
<td>1 vs 2</td>
<td>7.396143</td>
<td>0.006536</td>
</tr>
<tr>
<td>eba2</td>
<td>8</td>
<td>11.45401</td>
<td>2 vs 3</td>
<td>6.092099</td>
<td>0.013579</td>
</tr>
</tbody>
</table>

Non-nested models may be selected based on information criteria.

\[
\text{AIC}(btl, \ eba1, \ eba2)
\]

<table>
<thead>
<tr>
<th>df</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>btl</td>
<td>5</td>
</tr>
<tr>
<td>eba1</td>
<td>6</td>
</tr>
<tr>
<td>eba2</td>
<td>7</td>
</tr>
</tbody>
</table>

Conclusion: The elimination-by-aspects model with two extra parameters (eba2) fits the data best.
Scales derived from EBA model

• Younger males judge heroine to be about 13 times as dangerous as alcohol.
• Older males judge heroine to be only about 8 times as dangerous as alcohol.
Comparing subsamples

Is the same scaling valid in several groups?

Comparing male participants younger and older than 30 years

```r
males <- array(c(young, old), c(6,6,2))

group.test(males, A2)
```

|               | Df1 | Df2 | logLik1 | logLik2 | Deviance | Pr(|Chi|)    |
|---------------|-----|-----|---------|---------|----------|------------|
| EBA.g         | 14  | 30  | -60.49  | -48.94  | 23.09    | 0.111307   |
| Group         | 7   | 14  | -74.08  | -60.49  | 27.18    | 0.000309 ***|
| Effect        | 0   | 7   | -490.56 | -74.08  | 832.96   | < 2e-16 ***|
| Imbalance     | 1   | 30  | -85.69  | -85.69  | 0.00     | 1.000000   |

The scales of perceived health risk are significantly different ($G^2(7) = 27.18, p = .0003$) in the two groups.
Conclusions

- Pronounced differences between drugs w.r.t. perceived health risk
- Differences between male/female and younger/older participants
- Bradley-Terry-Luce model not valid in the male samples
- Elimination-by-aspects model with two additional parameters fits the data
- Elimination-by-aspects modeling is now easy to do using eba()
Thank you for your attention

florian.wickelmaier@uni-tuebingen.de

The ‘eba’ package http://CRAN.r-project.org
References


Predicting preference from specific auditory attributes
(Choisel & Wickelmaier, 2007, JASA)

Equal-preference contours for eight audio formats

Classical music

Pop music