 Probabilistic choice models

Overview

Probabilistic choice models
Perceived health risk of drugs
Within-pair order effects
Sound quality evaluation

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

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:**

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

- \(P_{xy}\): 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

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

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

**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

\[
P_{xy} \geq 0.5, P_{yz} \geq 0.5 \Rightarrow P_{xz} \geq \max\{P_{xy}, P_{yz}\}
\]

Elimination by aspects (EBA): model equations

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

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

\(x' \setminus y'\): 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:**

\(x' = \{\alpha, \beta, \zeta\}, y' = \{\gamma, \delta, \varepsilon, \zeta\}\) \(\Rightarrow\ P_{xy} = \frac{u(\alpha) + u(\beta)}{u(\alpha) + u(\beta) + u(\gamma) + u(\delta) + u(\varepsilon)}\)

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

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>17</td>
<td></td>
</tr>
<tr>
<td>Her</td>
<td>44</td>
<td>48</td>
<td>47</td>
<td>0</td>
<td>44</td>
<td></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:

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

Counting the number of transitivity violations

<table>
<thead>
<tr>
<th></th>
<th>violations</th>
<th>error.ratio</th>
<th>mean.dev</th>
<th>max.dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>weak</td>
<td>0</td>
<td>0.00</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>moderate</td>
<td>1</td>
<td>0.05</td>
<td>0.0417</td>
<td>0.0417</td>
</tr>
<tr>
<td>strong</td>
<td>5</td>
<td>0.25</td>
<td>0.0625</td>
<td>0.1458</td>
</tr>
</tbody>
</table>

---

Number of Tests: 20

BTL model

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:

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

Non-alcohol drugs share a feature that affects decision when comparing them with alcohol.

```r
A1 <- list(c(1), c(2,7), c(3,7), c(4,7), c(5,7), c(6,7))
eba1 <- eba(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

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

\[ \eta, \vartheta \]

A2 <- list (c(1) ,c(2,7) ,c(3,7) ,c(4,7,8) ,c(5,7,8) ,c(6,7,8))
eba2 <- eba (dat , A2)

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

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.

Model selection

Nested models can be compared using likelihood ratio tests.

anova(btl , eba1 , eba2)

Model Resid. df Resid. Dev Test Df LR stat. Pr(Chi)
1 btl 10 24.94225 NA NA NA
2 eba1 9 17.54611 1 vs 2 1 7.396143 0.006536
3 eba2 8 11.45401 2 vs 3 1 6.092099 0.013579

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

AIC(btl , eba1 , eba2)

df AIC
btl 5 78.18143
eba1 6 72.78528
eba2 7 68.69318

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

Comparing subsamples

Is the same scaling valid in several groups?

Comparing male participants younger and older than 30 years

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 \((G^2(7) = 27.18, \rho = .0003)\) in the two groups.
Summary

• 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()

Modeling order effects: Motivation

• Paired-comparison scaling has advantages over direct scaling procedures
  • Only qualitative (binary) judgments required
  • Consistency (transitivity) of judgments may be evaluated
• In paired-comparison experiments, stimuli are often presented sequentially
• How can a potential bias for one presentation interval be quantified?

Order effect: Davidson-Beaver (DB) model

Generalization of BTL model:
– Multiplicative parameter $\vartheta$ accounts for order of presentation

Model equations:

$$P_{xy|x} = \frac{u(x)}{u(x) + \vartheta_{xy} \cdot u(y)}, \quad P_{xy|y} = \frac{\vartheta_{xy} \cdot u(x)}{\vartheta_{xy} \cdot u(x) + u(y)}$$

• $P_{xy|x}$: probability of choosing alternative $x$ over $y$ given $x$ presented first
• $\vartheta_{xy} > 1$: advantage for the second stimulus
• $\vartheta_{xy} < 1$: advantage for the first stimulus
• Special case: $\vartheta_{xy} = \vartheta$ for all pairs of stimuli

EBA model with order effect

Generalization of Davidson-Beaver model:
– Multiplicative parameter $\vartheta$ accounts for order of presentation
– Context independence of choice is not required

Model equations:

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

• $\vartheta_{xy} > 1$: advantage for the second stimulus
• $\vartheta_{xy} < 1$: advantage for the first stimulus
• Special case: $\vartheta_{xy} = \vartheta$ for all pairs of stimuli
Application: Perceptual evaluation of multichannel sound
(Choisel & Wickelmaier, 2006, JAES)

8 audio formats:
Mono (mo)
Phantom mono (ph)
Stereo (st)
Wide stereo (ws)
Matrix upmixing (ma)
Dolby Prologic II (u*)
DTS Neo:6 (u*)
Original 5.0 (or)

Ordered paired-comparison data

Row stimulus first

<table>
<thead>
<tr>
<th></th>
<th>mo</th>
<th>ph</th>
<th>st</th>
<th>ws</th>
<th>ma</th>
<th>u1</th>
<th>u2</th>
<th>or</th>
</tr>
</thead>
<tbody>
<tr>
<td>mo</td>
<td>–</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ph</td>
<td>14</td>
<td>–</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>st</td>
<td>19</td>
<td>19</td>
<td>–</td>
<td>7</td>
<td>0</td>
<td>8</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>ws</td>
<td>18</td>
<td>18</td>
<td>13</td>
<td>–</td>
<td>6</td>
<td>9</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>ma</td>
<td>19</td>
<td>19</td>
<td>14</td>
<td>12</td>
<td>14</td>
<td>5</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>u1</td>
<td>17</td>
<td>17</td>
<td>12</td>
<td>11</td>
<td>8</td>
<td>13</td>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td>u2</td>
<td>19</td>
<td>19</td>
<td>14</td>
<td>18</td>
<td>14</td>
<td>12</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Column stimulus first

<table>
<thead>
<tr>
<th></th>
<th>mo</th>
<th>ph</th>
<th>st</th>
<th>ws</th>
<th>ma</th>
<th>u1</th>
<th>u2</th>
<th>or</th>
</tr>
</thead>
<tbody>
<tr>
<td>mo</td>
<td>–</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ph</td>
<td>15</td>
<td>–</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>st</td>
<td>18</td>
<td>19</td>
<td>–</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>ws</td>
<td>19</td>
<td>19</td>
<td>12</td>
<td>–</td>
<td>9</td>
<td>11</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>ma</td>
<td>17</td>
<td>17</td>
<td>11</td>
<td>10</td>
<td>–</td>
<td>14</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>u1</td>
<td>17</td>
<td>16</td>
<td>11</td>
<td>8</td>
<td>13</td>
<td>5</td>
<td>–</td>
<td>4</td>
</tr>
<tr>
<td>u2</td>
<td>18</td>
<td>14</td>
<td>7</td>
<td>8</td>
<td>1</td>
<td>6</td>
<td>–</td>
<td>7</td>
</tr>
<tr>
<td>or</td>
<td>19</td>
<td>19</td>
<td>14</td>
<td>18</td>
<td>12</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

– When st was presented first, nobody chose it over ma
– When st was presented second, 9 subjects chose it over ma

Descriptive statistics

\[ \text{strans(} \text{ord1} + \text{ord2} \) \]

violations   error_ratio mean_dev max_dev
\begin{tabular}{lcccc}
weak & 0 & 0.0000 & 0.0000 & 0.0000 \\
moderate & 2 & 0.0357 & 0.0385 & 0.0513 \\
strong & 23 & 0.4107 & 0.0803 & 0.2051 \\
--- & & & & \\
Number of Tests: & 56 & & & \\
\end{tabular}

• Many violations of strong stochastic transitivity
• BTL model inadequate?
Davidson-Beaver (DB) model

Fitting a DB model using the `eba.order()` function

dabe <- eba.order(ord1, ord2)
summary(dabe)

...  
Order effects (H0: parameter = 1):
   Estimate Std. Error z value Pr(>|z|)  
order 1.35513 0.10271 3.468 0.000545 ***

Model tests:
   Df1 Df2 logLik1 logLik2 Deviance Pr(>| Chi |)
EBA.order 8 56 -112.4 -74.2 76.407 0.00564 **
Order 7 8 -120.6 -112.4 16.370 5.21e-05 ***
Effect 1 8 -328.3 -112.4 431.775 < 2e-16 ***

AIC: 240.80  
Pearson Chi2: 66.65

Pronounced order effect, but DB model does not describe the data adequately ($G^2(48) = 76.41, p = .006$)

EBA model with order effect

Comparing models

anova(dabe, ebao)

   Model Resid. df Resid. Dev Test Df LR stat. Pr(Chi)
1 dabe 48 76.40717 NA NA NA
2 ebao 47 63.37553 1 vs 2 1 13.03164 0.000306

EBA order-effect model fits better than the DB model.

summary(ebao)

   Order effects (H0: parameter = 1):
      Estimate Std. Error z value Pr(>|z|)  
order 1.36147 0.10336 3.497 0.000470 ***

When two equally enveloping sounds are compared, the second one is chosen 36% more often than the first one.

EBA model with order effect

Model structure

$A_1 = \{\{\alpha, \iota\}, \{\beta, \iota\}, \{\gamma, \iota\}, \{\delta, \iota\}, \{\varepsilon\}, \{\zeta\}, \{\eta, \iota\}, \{\theta\}\}$

A1 <- list(c(1, 9), c(2, 9), c(3, 9), c(4, 9), c(5), c(6), c(7, 9), c(8))

Hypothesis: envelopment judged differently, depending on whether or not there are distinct sources (instruments) in surround channels

Scale derived from EBA order-effect model

- Original five-channel recording about 13 times as enveloping as mono downmix
- Commercially available upmix algorithms not more enveloping than stereo
Summary

- Pronounced order effects in the paired-comparison judgments
- For seven out of nine auditory attributes (including preference), biases favored the second choice interval
  Exceptions: distance (first interval), brightness (no order effect, $\theta = 1$)
- EBA order-effect model allows for measuring the magnitude of such biases where context independence (IIA) of judgments does not hold

References


Thank you for your attention

florian.wickelmaier@uni-tuebingen.de

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

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

Equal-preference contours for eight audio formats

Classical music

Pop music