Multi-Attribute Probabilistic Choice Models

Florian Wickelmaier

Eberhard Karls Universität Tübingen

Workshop on Psychometric Computing
Vienna, January 23, 2009

Overview

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 $(x, y, \dots)$ is probabilistic and depends on the weight (strength) of the alternative $(u(x), u(y), \dots)$

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)

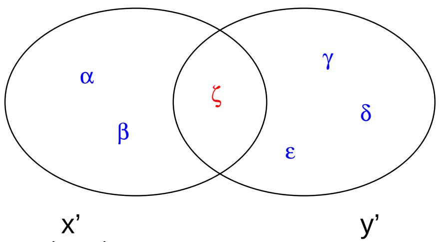
Choice is based on a hidden (sequential) elimination process:

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

The eba package

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

`strans`	Counting stochastic transitivity violations
`eba`	Fitting and testing EBA models
`summary, anova` `plot, residuals`	Extractor functions
`group.test`	Comparing samples of subjects
`eba.order`	Testing within-pair order effects

- Manual
  Wickelmaier, F. & Schmid, C. (2004). A Matlab function to estimate choice-model parameters from paired-comparison data. *Behavior Research Methods, Instruments, & Computers*, *36*, 29–40.

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)

	Alc	Tob	Can	Ecs	Her	Coc
Alc	0	28	35	10	4	7
Tob	20	0	18	2	0	3
Can	13	30	0	3	1	0
Ecs	38	46	45	0	1	17
Her	44	48	47	47	0	44
Coc	41	45	48	31	4	0

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         0.00    0.0000   0.0000
moderate           1         0.05    0.0417   0.0417
strong             5         0.25    0.0625   0.1458
---
Number of Tests:   20
```


BTL model

Fitting a BTL model using the `eba()` function

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

Obtaining summary statistics and model tests

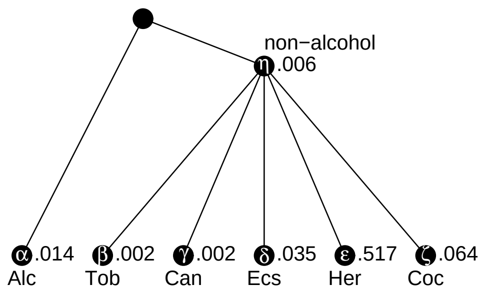
```
summary(btl)

...
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

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



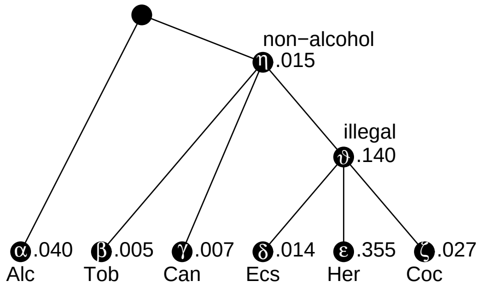

```
A1 <- list(c(1), c(2,7), c(3,7), c(4,7), c(5,7), c(6,7))
eba1 <- eba(dat, A1)
```

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

EBA model with two additional aspects – EBA2

Model structure

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




```
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.

Model selection

Nested models can be compared using likelihood ratio tests.

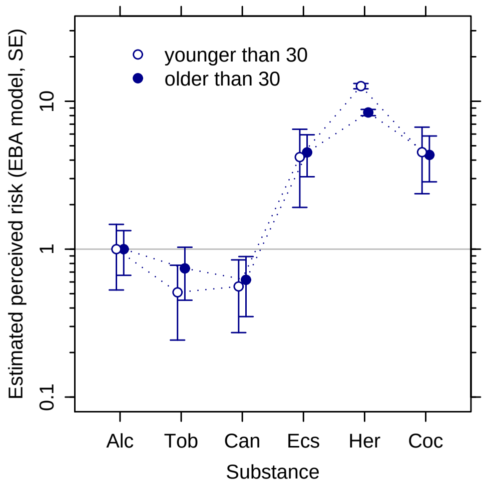
```
anova(btl, eba1, eba2)
 Model Resid. df Resid. Dev   Test Df LR stat.   Pr(Chi)
1  btl        10   24.94225          NA       NA        NA
2 eba1         9   17.54611 1 vs 2  1 7.396143 0.006536
3 eba2         8   11.45401 2 vs 3  1 6.092099 0.013579
```

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.

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

```
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.

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 ϑ accounts for order of presentation

Model equations:

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

- $P_{xy|x}$: 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 ϑ accounts for order of presentation
– Context independence of choice is not required

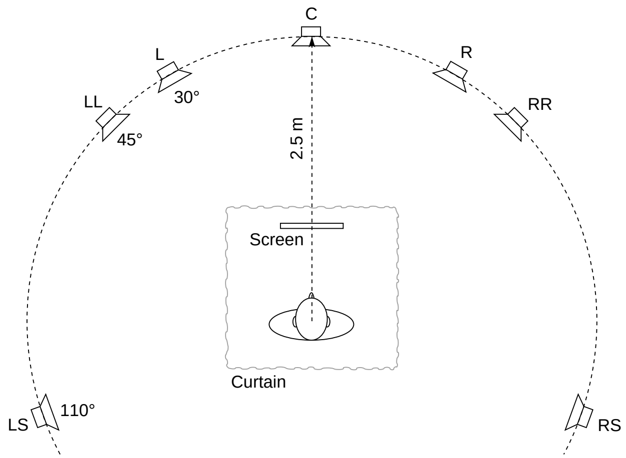
Model equations:

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

- $\vartheta_{xy} > 1$: 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)

Perceptual evaluation of multichannel sound

(Choisel & Wickelmaier, 2007, JASA)

Subjects: 39 selected listeners (27 male, 12 female)

Procedure:

- 2IFC (all possible paired comparisons among 8 audio formats)
- within-pair order counterbalanced
- repeated for four musical excerpts (2 × classic, 2 × pop)

Task 1: Select the sound that is more . . . wide, elevated, spacious, enveloping, far ahead, bright, clear, natural

Task 2: Select the sound that you prefer (measured 2×)

Envelopment: "A sound is enveloping when it wraps around you. A very enveloping sound will give you the impression of being immersed in it, while a nonenveloping one will give you the impression of being outside of it."

Ordered paired-comparison data

Row stimulus first

	mo	ph	st	ws	ma	u1	u2	or
mo	–	6	0	2	1	2	1	0
ph	14	–	1	2	2	3	3	1
st	19	19	–	7	0	8	10	2
ws	18	18	13	–	6	9	10	5
ma	19	17	19	14	–	12	14	5
u1	17	17	12	11	8	–	13	2
u2	19	16	9	10	5	7	–	7
or	19	19	18	14	14	18	12	–

Column stimulus first

	mo	ph	st	ws	ma	u1	u2	or
mo	–	4	2	0	2	3	1	3
ph	15	–	0	0	6	3	6	2
st	18	19	–	7	9	8	13	7
ws	19	19	12	–	9	11	11	9
ma	17	14	11	10	–	14	19	13
u1	17	16	11	8	5	–	13	4
u2	18	14	7	8	1	6	–	7
or	17	17	12	11	7	15	13	–

– When *st* was presented first, nobody chose it over *ma*

– When *st* was presented second, 9 subjects chose it over *ma*

Descriptive statistics

```
strans(ord1 + ord2)
           violations  error.ratio  mean.dev  max.dev
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
```

- Many violations of strong stochastic transitivity
- BTL model inadequate?

Davidson-Beaver (DB) model

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

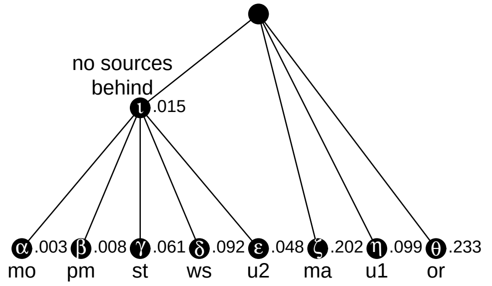
```
dabe <- eba.order(ord1, ord2)
summary(dabe)
...
Order effects (H0: parameter = 1):
       Estimate Std. Error z value Pr(>|z|)
order   1.35513    0.10271   3.458 0.000545 ***

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

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))
ebao <- eba.order(ord1, ord2, A1)
```

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

EBA model with order effect

Comparing models

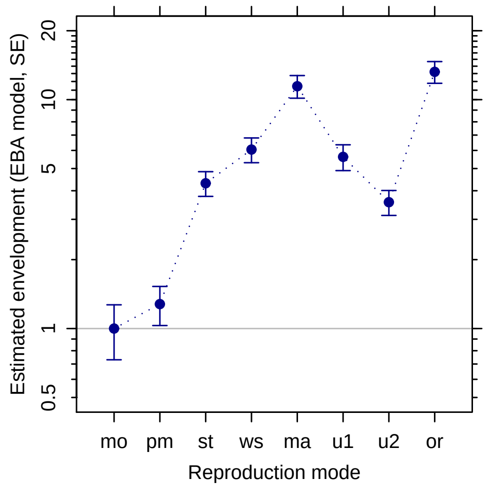
```
anova(dabe, ebao)
  Model Resid. df Resid. Dev   Test Df LR stat.  Pr(Chi)
1  dabe        48   76.40717          NA       NA       NA
2  ebao        47   63.37553 1 vs 2  1 13.03164 0.000306
```

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.

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, $\vartheta = 1$)
- EBA order-effect model allows for measuring the magnitude of such biases where context independence (IIA) of judgments does not hold

Thank you for your attention

florian.wickelmaier@uni-tuebingen.de

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

References

Choisel, S. & Wickelmaier, F. (2006). Extraction of auditory features and elicitation of attributes for the assessment of multichannel reproduced sound. *Journal of the Audio Engineering Society*, *54*, 815–826.

Choisel, S. & Wickelmaier, F. (2007). Evaluation of multichannel reproduced sound: scaling auditory attributes underlying listener preference. *Journal of the Acoustical Society of America*, *121*, 388–400.

Debreu, G. (1960). Review of R. D. Luce's Individual choice behavior: A theoretical analysis. *American Economic Review*, *50*, 186–188.

Rumelhart, D. L. & Greeno, J. G. (1971). Similarity between stimuli: An experimental test of the Luce and Restle choice models. *Journal of Mathematical Psychology*, *8*, 370–381.

Tversky, A. (1972). Elimination by aspects: a theory of choice. *Psychological Review*, *79*, 281–299.

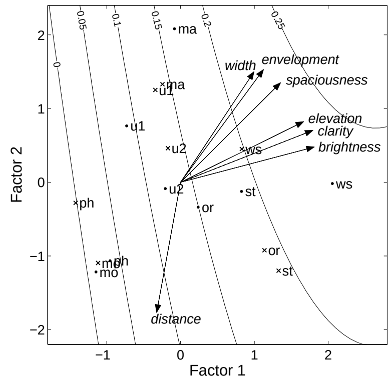
Wickelmaier, F. & Schmid, C. (2004). A Matlab function to estimate choice model parameters from paired-comparison data. *Behavior Research Methods, Instruments, & Computers*, *36*, 29–40.

Zimmer, K., Ellermeier, W., & Schmid, C. (2004). Using probabilistic choice models to investigate auditory unpleasantness. *Acta Acustica united with Acustica*, *90*, 1019–1028.

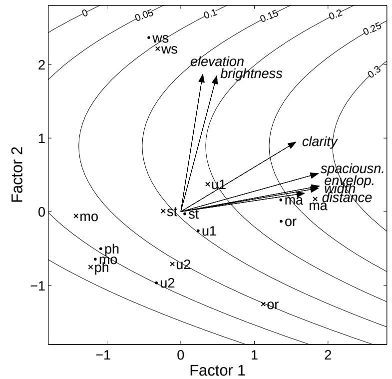
Predicting preference from specific auditory attibutes

(Choisel & Wickelmaier, 2007, JASA)

Equal-preference contours for eight audio formats



Classical music



Pop music