Probabilistic Choice Models for Psychological Scaling*

Florian Wickelmaier[1], Christian Schmid[2]
[1] Department of Acoustics, Aalborg University, DK-9220 Aalborg East, Denmark, Email: fw@acoustics.dk
[2] Institut für Experimentelle Psychologie, Universität Regensburg, D-93040 Regensburg, Germany

Introduction

Paired comparisons require nothing more from a subject than a choice between two stimuli with respect to a specified criterion. The subsequent analysis of such data by probabilistic choice models (PCMs) yields further advantages: ratio-scale measures of the stimuli can be derived, standard statistical theory can be employed for estimation and testing, and a theory of human decision making is incorporated in the scaling procedure.

Probabilistic choice models

A psychologically motivated PCM, called the elimination-by-aspects (EBA) model, is due to Tversky [1]. According to EBA a subject chooses one stimulus over another because of a certain aspect that belongs to this stimulus, but not to the other one. I. e. only the distinguishing aspects of the alternatives determine the decision. Let $x, y, z, \ldots$ denote the alternatives or stimuli under study, and let $x' = \{\alpha, \beta, \gamma, \ldots\}$ be the set of aspects that characterize the alternative x. Then according to EBA the probability of choosing x over y equals

$$P(x; y) = \frac{\sum_{\alpha \in x' \setminus y'} u(\alpha)}{\sum_{\alpha \in x' \setminus y'} u(\alpha) + \sum_{\beta \in y' \setminus x'} u(\beta)}, \tag{1}$$

where $x' \setminus y'$ is the set of aspects characterizing alternative x, but not alternative y. Note that the EBA model distinguishes between the scale values of the stimuli and the values of their aspects, $u(\alpha), u(\beta), \ldots$, which are the model parameters: the scale values are defined as the sum of the respective parameters and provide ratio-scale measures of the stimuli along the specified criterion.

The EBA model includes the Bradley-Terry-Luce (BTL) model [2] as a special case (if there is only one parameter per stimulus), and the preference tree model [3] which requires that the aspects are hierarchically structured. Unlike BTL, EBA can model stimulus similarity.

Parameter estimation

Point estimation The data are arranged in a square matrix having as many rows as stimuli. Each element of the matrix denotes the number of times the row stimulus has been chosen over the column stimulus. In order to obtain maximum likelihood estimates (MLEs) of the model parameters, the likelihood function of the model has to be specified. It takes the binomial form

$$L = \prod_{i<j} \pi_{ij}^{N_{ij}} (1 - \pi_{ij})^{N_{ji}}, \tag{2}$$

where i and j are the row and column indices, respectively, of the data matrix and N_{ij} is the ijth element. In the EBA model the probabilities π_{ij} are computed by equation (1). The MLEs, $\hat{u}(\alpha), \hat{u}(\beta), \ldots$ are the values that maximize equation (2); they are determined by numerical optimization.

Interval estimation Confidence intervals for the MLEs in the EBA model are calculated similarly as described in [2]: The Hessian matrix of the log-likelihood function is defined as the square matrix of second partial derivatives with respect to the model parameters:

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 \log L}{\partial u_1^2} & \cdots & \frac{\partial^2 \log L}{\partial u_1 \partial u_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \log L}{\partial u_k \partial u_1} & \cdots & \frac{\partial^2 \log L}{\partial u_k^2} \end{bmatrix}, \tag{3}$$

where k is the number of parameters. Plugging the vector of MLEs $\hat{\mathbf{u}}$ into equation (3) allows one to construct the matrix $\mathbf{C}$, which is the inverse of the negative Hessian augmented by a column and a row vector of ones, and a zero in the bottom right corner. The first k rows and columns of $\mathbf{C}$ form the estimated covariance matrix of $\hat{\mathbf{u}}$. The variances, and thus the standard errors, of the MLEs can be estimated from the main diagonal of the covariance matrix. The $100(1-\alpha)\%$ confidence intervals are obtained by

$$\hat{\mathbf{u}} \pm z_{1-\alpha/2} \sqrt{\mathbf{Diag}(\widehat{\mathbf{cov}}(\hat{\mathbf{u}}))}, \tag{4}$$

where z is the quantile of the standard normal distribution.

Hypothesis testing

Goodness of fit To check the goodness of fit of the EBA model it is convenient to compare it to the saturated binomial model that fits the data perfectly. In the saturated model the π_{ij} are estimated by $N_{ij}/(N_{ij} + N_{ji})$, thus it has $\binom{n}{2}$ free parameters, with n the number of stimuli. In the BTL model, the number of free parameters reduces to $n-1$. Every additional parameter, e. g. for a branch in a preference tree, has to be added, so in general the EBA model has $n-1+c$ free parameters. The likelihood ratio of the two models yields the test statistic. The expression

$$\chi_1^2 = -2 \log \left[\frac{L_{EBA}}{L_{SAT}} \right] = 2(\log L_{SAT} - \log L_{EBA}) \tag{5}$$

*This work was supported by the Centercontract on Sound Quality, Aalborg University.

is approximately χ^2-distributed with $\binom{n}{2} - (n-1+c)$ degrees of freedom. The EBA model can be rejected if the χ_1^2 exceeds the conventional critical value.

Stimulus equality A test whether there is any difference between the stimuli with respect to the given criterion is provided by a comparison to the *null model* where all the π_{ij} are fixed to 0.5. The test statistic is given by

$$\chi_0^2 = 2(\log L_{EBA} - \log L_0), \tag{6}$$

where L_0 denotes the likelihood of the null model. It has as many degrees of freedom as there are free parameters in the EBA model. A significant result indicates that at least two stimuli are different from each other.

Test of a single parameter Each model parameter can be tested whether it is significantly different from zero using the test statistic

$$T = \hat{u}(\alpha)/\sqrt{\hat{\sigma}_\alpha^2}, \tag{7}$$

where $\hat{\sigma}_\alpha^2$ is an element in the main diagonal of $\widehat{\mathbf{cov}}(\hat{\mathbf{u}})$, and T is approximately standard normal. Non-significant parameters might hint at a mis-specification of the model.

Model selection

Nested models If the parameter space Ω' of one model is a proper subset of the other model's parameter space, Ω, the two models are nested. The restricted model *EBA′* and the unrestricted model *EBA* can be tested against each other by

$$\chi_2^2 = 2(\log L_{EBA} - \log L_{EBA'}), \tag{8}$$

which has as many degrees of freedom as the difference between the number of parameters in *EBA* and in *EBA′*. The restricted model can be rejected if the likelihood ratio test is significant.

Non-nested models It is not in every case that two different EBA models are nested. For non-nested models, it is common practice to employ so-called information criteria as a tool for model selection, which take into account both the likelihood of the model and the number of free parameters. For the EBA model, Akaike's information criterion (AIC) is defined as

$$\mathrm{AIC} = -2\log L_{EBA} + 2(n-1+c). \tag{9}$$

When comparing two models using the AIC, the model with the smaller AIC should be selected.

Scaling auditory unpleasantness

The following section illustrates the methods discussed so far (see [4] for a full description of the experiment).

Method 74 normal-hearing subjects were presented with all 66 pairs of twelve binaurally-recorded environmental sounds. On each trial, the task was to choose the more unpleasant sound. The results were pooled over subjects in order to yield the aggregate paired-comparison matrix.

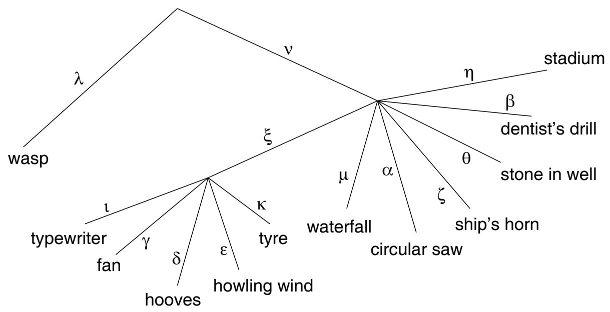

Figure 1: Schematic structure of a preference tree which represents the unpleasantness of twelve environmental sounds.

Results A 14-parameter preference tree (one parameter per sound plus two extra parameters) was found to account for the data; its structure is depicted in Figure 1.

Employing the goodness-of-fit test in (5) reveals that the model fits the data well [$\chi_1^2(53) = 58.47$, $p = .282$]. The hypothesis that all sounds are equal in unpleasantness can be clearly rejected on the basis of (6) [$\chi_0^2(13) = 2612.65$, $p < .001$]. The two branch parameters ν and ξ are according to (7) significantly different from zero [$T_\nu = 2.23$, $p = .026$; $T_\xi = 3.10$, $p = .002$]. In order to compare the preference tree to the simpler BTL model, the test in (8) can be employed, since the models are nested. The BTL model, however, fits the data significantly worse [$\chi_2^2(2) = 25.72$, $p < .001$]. The AIC amounts to 336.18 for the preference tree and to 357.90 for the BTL model, which argues for the selection of the preference tree.

A ratio-scale measure of the unpleasantness of each stimulus is obtained by summing over the values of the characterizing aspects, e. g. the unpleasantness of the *fan* is determined by $u(\nu) + u(\xi) + u(\gamma)$.

Concluding remarks

PCMs should be considered as a powerful alternative to traditional direct scaling procedures. Recently, a computer program was made available [5], which meets the need for software for fitting and testing these models.

References

[1] A. Tversky: Elimination by aspects: A theory of choice. Psych. Rev. **79** (1972) 281-299.

[2] R. A. Bradley: Paired comparisons: some basic procedures and examples. – In: Handbook of Statistics. Vol. 4 P. R. Krishnaiah, P. K. Sen (eds.). Elsevier, Amsterdam, 1984, 299-326.

[3] A. Tversky, S. Sattath: Preference trees. Psych. Rev. **86** (1979) 542-573.

[4] K. Zimmer, W. Ellermeier, C. Schmid: Using probabilistic choice models to investigate auditory unpleasantness. (under review).

[5] F. Wickelmaier, C. Schmid: A Matlab function to estimate choice-model parameters from paired-comparison data. Behav. Res. Meth. Instr. Comp. **36** (2004) 29–40.