Cognitive and Hierarchical Bayesian Models for Subjective Preference Data

Christophe Micheyl[1], Jumana Harianawala[2], Brent Edwards[1], Tao Zhang[2]

[1]*Starkey Hearing Research Center, Berkeley, CA* [2]*Starkey Hearing Technologies, Eden Prairie, MN*

Motivation

Preference data are a key source of information in the subjective evaluation of hearing aids.

Unfortunately, the analysis of preference data is complicated by:

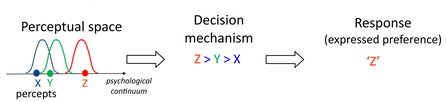
- ordinal nature of the data (A>B)
  → linear models such as ANOVA are inadequate; ordinal models are needed;
- data incompleteness (e.g., subjects only give their top preference)
  → calls for latent-variable models, which can handle partial (or missing) data.

Here, we illustrate an approach for aggregating and analyzing partial preference data using a Thurstonian model in a hierarchical Bayesian framework.

Traditional approaches to preference-data analysis

- **Wins-count models**
  Count the number of times a given preference judgment was expressed, then analyze counts using parametric (e.g., Poisson or binomial) or non-parametric statistics (David, 1988).
  Limitations: ignores unexpressed preferences; difficult to adapt to complex experimental designs (e.g., mixed designs or designs involving more than two response alternatives).

- **Ordinal-regression or Thurstonian models** (Thurstone, 1927)



See: Marden (1995) Agresti (2002) for reviews

Special case: Plackett-Luce model (a Thurstonian model with Gumbel instead of Gaussian distributions)
Advantages: flexible; long history in psychology, stats; many applications (e.g., marketing)

- **Permutation models**
  Permutation group: A>B>C; B>A>C; B>C>A; ...
  Based on probability distribution on permutation group (e.g., Mallows, 1957).
  Limitations: does not lend itself as easily to modeling of underlying psychological processes.

Real-world example

Setup:
18 hearing-impaired participants compared three different sound-processing algorithms (M1, M2, M3) on five aspects: Noise; Artifacts; Fluctuations; Speech clarity; Overall quality.

Example data (answers from participants):
"M1 is the noisiest"
"M1 and M2 are best for sound quality"
"I cannot tell the difference"

Question:
Do subjects prefer one algorithm over the others?

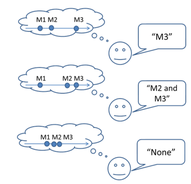
Hurdles for statistical analysis:

- Unusual response set (subjects expressed 0, 1, or 2 preferences, or greatest dislikes)
- Partial rankings (subjects only had to report their top preference/least-liked choice)
- Mixed (within- and across-subjects) design

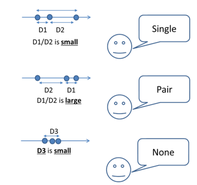
Our approach: combine a Thurstonian model with a hierarchical Bayesian model

Thurstonian model of preference judgments

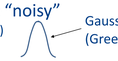
Preferences reflect ordering of percepts on psychological continuum.



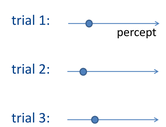
Distances between percepts determine which preferences are expressed.



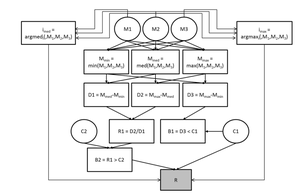
Percepts are "noisy" (Thurstone, 1927)



Gaussian noise (Green & Swets, 1966)

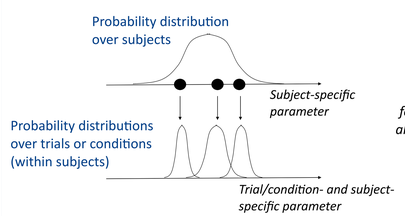


Model schematic



Hierarchical Bayesian model

A-priori exchangeability assumption over subjects implies a hierarchical model (Jackman, 2009):



Sources of variances are structured using a general linear model, with main effects and interactions (Gelman and Hill, 2007):

$$\mu_{mjq} = \mu_m + \mu_{mj} + \mu_{mq} + \varepsilon_{jmq}$$

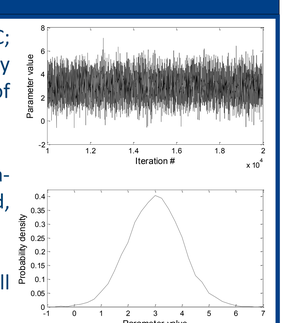
mean percept for given subject, algorithm, and aspect — *main effect of algorithm* — *algorithm-subject interaction* — *algorithm-aspect interaction* — *error*

Estimation of model parameters

Numerical methods (Markov-chain Monte Carlo, MCMC; see: Bolstad, 2009) are used to infer posterior probability distributions of model parameters, i.e., distributions of model parameters conditioned on the data.

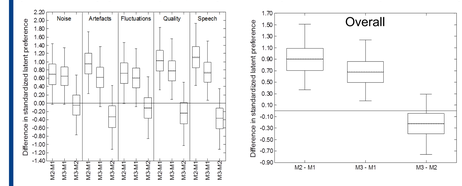
One advantage of this approach is that all missing observations are automatically 'imputed', i.e., they are estimated, taking into account all other variables in the model.

This is done via 'marginalization', i.e., integrating over all unknown parameters.



Results

To determine whether participants showed a statistically significant preference for one algorithm, we examined posterior distributions of pairwise differences in latent (inferred) percepts, separately for each question (Noise, Artifacts, etc...), and overall.



95% posterior confidence intervals that do not span 0 are indicative of statistically significant preferences.

Overall, and for some aspects (Artifacts, Quality, Speech), participants preferred M2 (and/or) M3 to M1. A (non-significant) tendency for participants to prefer M2 to M3 is also apparent, and might be confirmed with more data.

Conclusions

Advantageous features of the approach:

- Based on an explicit model of preference judgments.
- Modeling assumptions are apparent, and testable.
- Like ordinal-regression, it is well-suited to ordinal data (or rankings).
- Handles incomplete (partial) preference data, as well as missing data.
- Can be adapted to suit specific/unusual experimental designs.
- Uses a hierarchical structure for data aggregation within and across subjects.

Caveats:

- Implementation can be tricky (requires experience with Bayesian modeling techniques)
- Modeling assumptions (e.g., Gaussian noise) can be difficult to check.

Avenues for further research include:

- Apply the approach on other datasets.
- Check modeling assumptions (e.g., Gaussian distributions).
- Compare models (e.g., Thurstonian vs Mallows).

References

- Agresti, A. (2002). Categorical Data Analysis. Wiley: Hoboken.
- Bolstad WM (2009) Understanding Computational Bayesian Statistics. Wiley, Hoboken.
- David HA (1988) The Method of Paired Comparisons. Oxford University Press, New York.
- Gelman A, Hill J (2007) Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, New York.
- Green DM, Swets JA (1966) Signal Detection Theory and Psychophysics. Wiley, New York.
- Jackman S (2009) Bayesian Analysis for the Social Sciences. Wiley, Hoboken.
- Yu PLH (2000) Order-statistics models for ranking data. Psychometrika 65, 281-299.
- Luce RD (1959) Individual choice behavior: A theoretical analysis. Wiley, New York.
- Mallows CL (1957) Non-null ranking models. Biometrika, 44, 114-130.
- Marden JI (1995) Analyzing and modeling rank data. Chapman & Hall, London.
- Thurstone LL (1927) A Law of comparative judgment. Psych Rev 34 (4), 273–286.
- Yellot J (1680) Generalized Thurstone models for ranking: Equivalence and reversibility. J Math Psychol 22, 48-69.