Applied experimental psychology and psychophysics

PROBABILISTIC CHOICE MODELS

- BTL MODELS ON UNPLEASANT SOUNDS -

26. Oktober 2017

Group members:

Andreas Kornmaaler Hansen

Lucca Julie Nellemann

Emil Bonnerup

Juliane Nilsson

Sara Nielsen

Group 782
Electronic Systems, Aalborg University
17gr782@es.aau.dk

Introduction

The purpose of this article is to analyse data obtained from **Ellermeier2004** using the probabilistic choice model Bradley-Terry-Luce (BTL). More specifically to check for transitivity in the data which includes: Weak Stochastic Transitivity (WST), Mean Stochastic Transitivity (MST), and Strong Stochastic Transitivity (SST). The dataset is analysed using the Matlab function *fOptiPt.m* developed by Florian Wickelmaier and Christian Schmid, (**Wickelmaier2004**). See Table 0.1 for the absolute frequencies used in a cumulative preference matrix. The scores denote how many participants voted the sound in question as being the most unpleasant.

Sounds	No.	1	2	3	4	5	6	7	8	9	10
Truck	1	0	9	16	45	56	5	29	6	24	33
Brake	2	51	0	34	58	58	13	46	30	39	50
Train	3	44	26	0	55	57	9	48	37	38	55
Water	4	15	2	5	0	38	2	17	6	6	20
Boat	5	4	2	3	22	0	3	6	3	3	12
Jackhammer	6	55	47	51	58	57	0	58	53	55	57
Mower	7	31	14	12	43	54	2	0	16	17	41
Crash	8	54	30	23	54	57	7	44	0	40	52
Mixer	9	36	21	22	54	57	5	43	20	0	43
Vent	10	27	10	5	40	48	3	19	8	17	0

Table 0.1. The cumulative preference matrix of the unpleasant sounds from **Ellermeier2004** The absolute frequency denotes how many participants judged the sound as being the most unpleasant.

The probabilistic method used is called the BTL model. It is a model which checks for transitivity in data containing pairwise comparisons and is able to predict the outcome of the comparison. See Equation 1.

$$P_{ab} = \frac{v(a)}{v(a) + v(b)} \tag{1}$$

P_{ab} is the probability of sound a being rated more unpleasant than sound b. $v(a)$ and $v(b)$ denotes the scale values of the sounds a and b.

In order to get estimates of the scale values, one has to maximise the likelihood of the data seen in Table 0.1. This is done given the model in Equation 2:

$$L(D|\Theta_{model}) = \prod_{i<j} p_{ij}^{n_{ij}} \cdot (1 - p_{ij})^{n - n_{ij}} \tag{2}$$

Where n is the number of comparisons, which is equal to the amount of test subjects in this case (n=60) and n_{ij} is the frequency of how many times the sound in row i was rated as being more unpleasant over the sound in column j. P_{ij} is the preference probability and is estimated using Equation 1.

Method

In order to know whether or not the BTL model can be applied to the data shown in Table 0.1, it is first necessary to check for transitivity in the data. In other words how reliable and consistent the data is. The cumulative preference matrix is a pooled data matrix meaning it contains data from all the participants deemed sufficiently consistent. This is typically done using a $\chi^2 - test$. Now a problem could arise because it is unknown whether the pooled subjects have shown an opposite decision behaviour eg. chosen pleasant where others chose unpleasant or not. If that is the case, then the preference matrix becomes inconsistent. Before it is possible to do the transitivity check, it is necessary to calculate the probability of a stimulus being rated as unpleasant. See Equation 3.

$$p = \frac{freq}{n} \tag{3}$$

Here *freq* is the frequency that a sound has been rated unpleasant (the pooled preference matrix, Table 0.1) and n is the number of test subjects. Thereafter it is needed to check every combination of the stimuli for transitivity violations. There are no set rules but a rough estimate of when the probabilistic choice models holds is shown in Table 0.2:

Transitivity violations	Expected
None or few SST violations	BTL may fit
Some SST violations and few MST violations	Preference-tree might fit (Not BTL)
Considerable SST and MST but few WST violations	Possible to rank order (no model will fit)

Table 0.2. General estimates of the outcome of probabilistic choice models considering the amount of weak (WST), moderate (MST) and strong (SST) transitivity violations.

Consider three stimuli: a, b, and c. If it is observed that $p_{ab} \geq 0.05$ and $p_{ab} \geq 0.05$, in other words if the probability of a being rated as more unpleasant compared to b and b more unpleasant compared to c. Then Weak Stochastic Transitivity (WST) holds when $p_{ac} \geq 0.05$. It is considered a WST violation if $p_{ac} < 0.05$. The same logic applies to MST and SST. If p_{ac} is bigger than or equal to the minimum value in p_{ab} and p_{bc}, MST holds ($p_{ac} \geq min(p_{ab}; p_{bc})$. It is a violation if not. For SST, p_{ac} is compared to the maximum value of both p_{ab} and p_{bc}. If p_{ac} is not bigger or equal to these values, it counts as an SST violation ($p_{ac} < max(p_{ab}; p_{bc}.)$

Luckily, with the help of Matlab it is not needed to perform every single comparison by hand. A loop was written which identified the number of WST, MST and SST violations respectively. See Figure 0.1.

```
WST = 0;
MST = 0;
SST = 0;
Count = 0;

for a = 1:10;
    for b = 1:10;
        for c = 1:10;
            if (a~=b)&&(b~=c)&&(a~=c); %avoids comparing the same value (0)
                if p(a,b)>=0.5 && p(b,c) >=0.5;
                    Count = Count+1;
                    if p(a,c)<0.5; %checks for WST violations
                    WST = WST+1; %counts number of violations
                    end
                        if p(a,c) < min(p(a,b),p(b,c)); %checks for MST violations
                        MST = MST+1; %counts number of violations
                        end
                            if p(a,c) < max(p(a,b),p(b,c)); %checks for SST violations
                                SST = SST+1;%counts number of violations
                            end
                end
            end
        end
    end
end
```

Figure 0.1. Loop used to calculate WST, MST and SST violations.

The same goes for the likelihood estimation. The *fOptiPt.m* Matlab function has been developed for this purpose. The function requires two mandatory input, M and A, where M is the paired comparison matrix shown in Table 0.1, and A is a cell array with length corresponding to the number of stimuli, which is 10. Further there is an optional input, s, which denotes the starting values for the estimation routine. The search algorithm starts at $\frac{1}{k}$ for each parameter value, where k is the number of parameters, if s is not specified. In this case s is not specified nor is it used.

Results

The probability of choosing one stimuli over another is calculated for each comparison. The results for the probabilities are shown in Table 0.3.

Sounds	No.	1	2	3	4	5	6	7	8	9	10
Truck	1	0	0.15	0.27	0.75	0.93	0.08	0.48	0.10	0.40	0.55
Brake	2	0.85	0	0.57	0.97	0.97	0.22	0.77	0.50	0.65	0.83
Train	3	0.73	0.43	0	0.92	0.95	0.15	0.80	0.62	0.63	0.92
Water	4	0.25	0.03	0.08	0	0.63	0.03	0.28	0.10	0.10	0.33
Boat	5	0.07	0.03	0.05	0.37	0	0.05	0.10	0.05	0.05	0.20
Jackhammer	6	0.92	0.78	0.85	0.97	0.95	0	0.97	0.88	0.92	0.95
Mower	7	0.52	0.23	0.20	0.72	0.90	0.03	0	0.27	0.28	0.68
Crash	8	0.90	0.50	0.38	0.90	0.95	0.12	0.73	0	0.67	0.87
Mixer	9	0.60	0.35	0.37	0.90	0.95	0.08	0.72	0.33	0	0.71
Vent	10	0.45	0.17	0.08	0.67	0.80	0.05	0.32	0.13	0.28	0

Table 0.3. The probability that one stimulus is chosen over another stimulus.

To find out how many times the stochastic transitivities is violated, each of the three types is investigated and the number of times there is a violation is counted. The results are shown in Table 0.4.

Stochastic transitivity	Violations
WST	2
MST	3
SST	25

Table 0.4. Results for the number of violations of the three stochastic transitivities.

To check how good a fit the BTL-model is, a Chi-square test is conducted and the results are shown in Table 0.5.

χ^2	Df	p-value
38.1358	36	0.3725

Table 0.5. Results from the Chi-square test.

Analysis

Based on the results presented in the previous section it is possible to analyze how well a fit the BTL-model is.

According to the total number of possible violations namely 129, it seems that the BTL-model is a good fit, especially according to the low number of WST violations of which there where only two. Because the BTL-model sufficiently describe the data it is possible to view the original ordinal data as ratio data where each of the 10 sounds have their own scale value. Furthermore the resultant p-value of 0.3725, found with the Chi-square test, confirms the goodness of the fit because the p-value is above the significant level set at 0.1.

To signify what the analysis shows, the scale values for each sound and their respective confidence intervals are plotted in Figure 0.2.

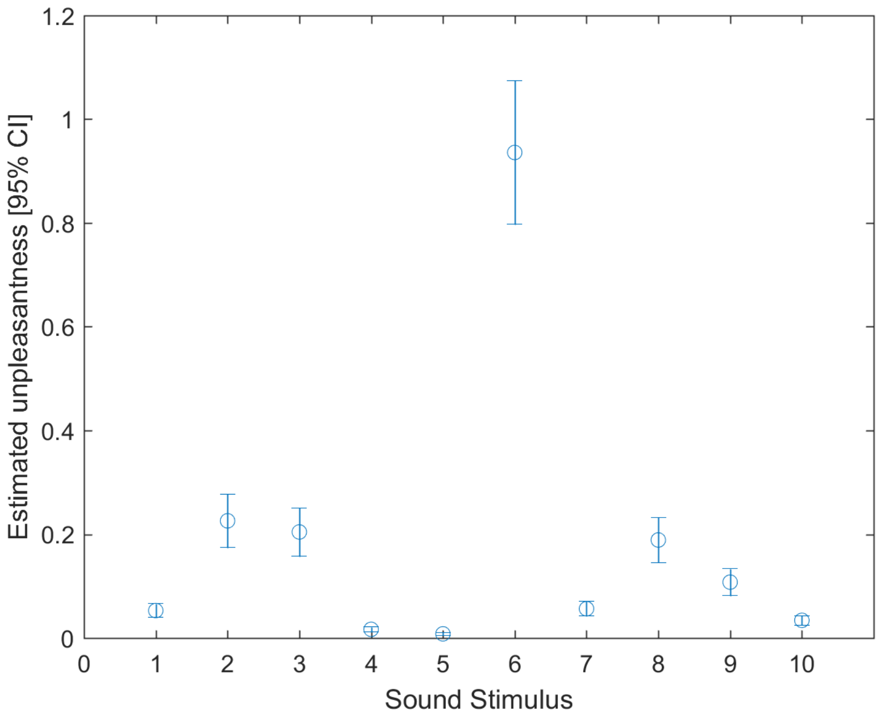

Figure 0.2. Scale values and 95 % confidence intervals for each of the 10 sounds.

From Figure 0.2 it is clear that the scale value for sound number six is significantly different from all other values, i.e. the confidence intervals for sound number six does not overlap with the remaining nine confidence intervals. Based on Figure 0.2, there are some indications that the remaining nine sounds are split into two groups. The first group, which is the group with the highest ratings except from sound six, consists of sound number two, three, eight, and nine. These four sound numbers or scale values are not significantly different from each other, but they are significantly different from both sound six and the other group. The other group consists of sound number one, four, five, seven, and ten which got the lowest rankings, thus are the sounds which are less unpleasant compared to the others.

Discussion

The number of SST violations was a bit high, namely 25 out of 129 possible violations. This account for 16.2% out of the possible violations. There does not exist a definitive rule which defines the acceptable amount of SST violations before the BTL-model does not fit. However, the number of WST and MST violations was low, two and three respectively. The low number of WST violations is contributing to the opinion that the BTL-model fits well with the data. If this is the case, according to **Zimmer2003** the test subjects compared the unpleasantness of the ten sounds according to the same criterions.

Conclusion

The data from **Ellermeier2004** were analyzed by using the BTL-model. Based on both the results namely the number of WST violations and the Chi-square test, presented in section (*Results*), and the analysis, presented in section (*Analysis*), it can be concluded that the BTL-model sufficiently describe the collected data. According to Figure 0.2 it is obvious that the jackhammer is the absolute most unpleasant sound of the ten otherwise unpleasant sounds.