Ranking models in conjoint analysis

Kar Yin Lam   Alex J. Koning   Philip Hans Franses

Econometric Institute Report EI 2010-51

Abstract

In this paper we consider the estimation of probabilistic ranking models in the context of conjoint experiments. By using approximate rather than exact ranking probabilities, we do not need to compute high-dimensional integrals. We extend the approximation technique proposed by Henery (1981) in the Thurstone-Mosteller-Daniels model for any Thurstone order statistics model and we show that our approach allows for a unified approach. Moreover, our approach also allows for the analysis of any partial ranking. Partial rankings are essential in practical conjoint analysis to collect data efficiently to relieve respondents' task burden.

Contents

1 Introduction

Although conjoint analysis originated more than forty years ago, see Luce and Tukey (1964); Kruskal (1965), it continues to attract active interest as a research field, see Green et al. (2001); Bradlow (2005). According to Hauser and Rao (2004), full-profile analysis remains the most common form of conjoint analysis. Full-profile conjoint has the advantage that the respondent evaluates each profile holistically and in the context of all attributes. In full-profile conjoint experiments, each respondent evaluates and ranks a (sub)set of stimuli, each stimulus defined as a specific combination of attributes levels.

As there is only a finite number of possible rankings, the rankings have a discrete distribution. In principle, standard methods for analysing discrete data apply here, see Marden (1995, p. 140). However, probability models for rankings become very complex, as the computation of each ranking probability usually requires high-dimensional integration when the number of stimuli becomes large.

Earlier approaches for analysing conjoint experiments thus avoid the use of probability models for rankings by resorting to multi-dimensional scaling techniques to derive respondent preferences (see e.g. Green et al. (2001)).

Recently there is renewed interest in the modeling and estimation of ranking models, see Maydeu-Olivares (1999); Maydeu-Olivares and Böckenholt (2005); Böckenholt (2006); Maydeu-Olivares and Hernández (2007). In such an approach typically rankings are transformed to (possibly intransitive) paired comparisons and analyzed by formulating a Thurstonian model for paired comparisons as a structural equation model with binary indicators. Hence, the computation of ranking probabilities is replaced by the computation of probabilities of binary outcomes. However, in full-profile conjoint the focus of a respondent's decision is on the acceptability of a stimulus' attributes, rather than differences between stimuli. Moreover, it is also an inefficient way to gather preference information as paired comparisons only indicate which stimulus is preferred rather than the strength of preference.

In this paper we consider the estimation of probabilistic ranking models in the context of full-profile conjoint experiments. We reduce the complexity of probabilistic ranking models considerably by using approximate rather than exact ranking probabilities. In the literature an abundance of ranking models is available, see Critchlow et al. (1991, Section 3) and also Marden (1995, Chapter 5), but we concentrate on Thurstone order statistics models.

Henery (1981) used a simpler model to approximate the Thurstone-Mosteller-Daniels model. We show that any Thurstone order statistics model may be approximated by such a "Henery model". This allows for a unified approach.

Moreover, our approach also allows for the analysis of partial rankings.

Partial rankings are essential in practical conjoint analysis to collect data efficiently to relieve respondents' task burden. A specific partial ranking which is gaining in popularity is Best-Worst ranking. In this specific case, respondents are instructed to select only the best and the worst stimulus. Partial rankings may also be dictated by the experimental design of the conjoint study.

The structure of the paper is as follows. In Section 2 we discuss ranking models and the approximation of ranking probabilities. In Section 3, we adapt general ranking models to conjoint experiments by introducing a linear model which allows for modeling the dependence of the rankings on the stimulus characteristics. In Section 4 we illustrate our methodology and compare our estimation with analysis results in Maydeu-Olivares and Böckenholt (2005). Finally, in Section 5 we conclude with suggestions for further research. Most technical issues are discussed in the appendices.

2 Ranking models

2.1 Preliminaries

In this section, we consider a single respondent who lists all stimuli, $1, 2, \ldots, C$, in order of preference, with the most preferred stimulus listed first. For each stimulus c in $\{1, 2, \ldots, C\}$, we define the rank $\pi_{(c)}$ of c as the position of c within this ordering. For example, $\pi_{(3)} = 7$ indicates that stimulus 3 is listed in 7^{th} place in order of preference. We shall refer to $\pi = \left(\pi_{(1)}, \pi_{(2)}, \ldots, \pi_{(C)}\right)$ as a full ranking.

Observe that in a fullranking for each rank r there exists exactly one stimulus c such that $\pi_{(c)} = r$. We shall denote this stimulus by $\pi^{-1}_{(r)}$. For example, $\pi^{-1}_{(7)} = 3$ denotes that stimulus 3 is listed in 7^{th} place in order of preference. Remark that we now may express the ordering as $\pi^{-1} = \left(\pi^{-1}_{(1)}, \pi^{-1}_{(2)}, \ldots, \pi^{-1}_{(C)}\right)$.

We assume that for each respondent the probability p_π of actually obtaining π as full ranking depends on a C-dimensional linear predictor vector $\eta = (\eta_1, \eta_2, \ldots, \eta_C)^t$; that is,

$$p_\pi = p\left(\pi \mid \eta\right). \tag{1}$$

A so-called ranking model specifies the exact nature of the dependence of p_π on η.

A common issue in the analysis of rankings is the handling of ties. A tie means that the same rank is assigned to multiple stimuli. Ties may occur due to the respondent's inability to differentiate between two or more stimuli. Ties may also occur due to requirements imposed by the research design. It has been widely recognized that respondents may find it difficult to compare too many choice options. This can be solved by asking respondents to rank only a subset of stimuli. For instance, in Best-Worst ranking, a respondent is instructed to select only the best and the worst stimulus; thus, all the other stimuli are tied.

Another common issue in the analysis of rankings is the handling of missings. A missing means that none rank is assigned to a stimulus. Missings may occur due to requirements imposed by the research design. For example to alleviate respondents task complexity, respondents are shown only a subset of all stimuli. Missings differ from ties, in that a missing could have been assigned any rank r in $\{1, 2, \ldots, C\}$.

As a ranking which contains ties or missings should be considered as a partial ordering of the stimuli rather than a full ordering, we shall refer to it as a partial ranking. Observe that for each partial ranking ϖ there exists a set $\mathcal{S}_\varpi$ of all full rankings which do not contradict the partial ordering implied by ϖ. Thus, we may assign the probability

$$p_\varpi = p(\varpi \mid \eta) = \sum_{\pi \in \mathcal{S}_\varpi} p(\pi \mid \eta) \tag{2}$$

to the partial ranking ϖ.

2.2 Thurstone order statistics models

In Thurstone order statistics models, see Thurstone (1927); Critchlow et al. (1991); Luce (1994); Böckenholt (2006), it is assumed that the rank of stimulus c among the stimuli $1, 2 \ldots, C$ is in fact equal to the rank of a random variable Y_c among the random variables $Y_1, Y_2, \ldots, Y_C$. Here, $Y_1, Y_2, \ldots, Y_C$ are random variables having some joint continuous distribution. It follows that

$$p_\pi = P(Y_{c_1} < Y_{c_2} < \ldots < Y_{c_C}) \tag{3}$$

for an ordering $\pi^{-1} = (c_1, c_2, \ldots, c_C)$. Observe that

$$p_\pi = \int_{-\infty}^{\infty} \int_{y_{c_1}}^{\infty} \int_{y_{c_2}}^{\infty} \cdots \int_{y_{c_{C-1}}}^{\infty} f(y_{c_1}, y_{c_2}, \ldots, y_{c_C}) \, dy_{c_C} \cdots dy_{c_2} dy_{c_1}, \tag{4}$$

where $f(y_{c_1}, y_{c_2}, \ldots, y_{c_C})$ denotes the joint density of $Y_{c_1}, Y_{c_2}, \ldots, Y_{c_C}$.

Thurstone order statistics models often assume that $Y_1, Y_2, \ldots, Y_C$ are independent random variables with distributions from the same family with density $g(Y, \eta)$. Under this assumption, $Y_{c_1}, Y_{c_2}, \ldots, Y_{c_C}$ have joint density

$$f\left(y_{c_1}, y_{c_2}, \ldots, y_{c_C}\right) = g\left(y_{c_1}; \eta_{c_1}\right) g\left(y_{c_2}; \eta_{c_2}\right) \cdots g\left(y_{c_C}; \eta_{c_C}\right), \tag{5}$$

where the parameters $\eta_{c_1}, \ldots, \eta_{c_C}$ are allowed to vary. Combining (4) and (5) yields

$$p_\pi = \int_{-\infty}^{\infty} \int_{y_{c_1}}^{\infty} \cdots \int_{y_{c_{C-1}}}^{\infty} g\left(y_{c_1}; \eta_{c_1}\right) g\left(y_{c_2}; \eta_{c_2}\right) \cdots g\left(y_{c_C}; \eta_{c_C}\right) dy_{c_C} \cdots dy_{c_2} dy_{c_1}. \tag{6}$$

In the special case that η a location parameter is, that is, $g(Y, \eta)$ takes the form $g(Y - \eta)$, the model is referred to as a Thurstone model, see Critchlow et al. (1991). Well-known Thurstone models are the Thurstone-Mosteller-Daniels model (see Mosteller (1951); Daniels (1950)), and the Luce model (see Luce (1959)). In the Thurstone-Mosteller-Daniels model the density $g(z)$ is a standard normal density, and in the Luce model $g(z)$ is a Gumbel density.

2.3 Approximate probabilities

The multiple integral on the right-hand side of (4) is usually evaluated by means of numerical integration. Thus, this approach is not feasible when the number of stimuli becomes large.

Henery (1981) approximates p_π in the Thurstone-Mosteller-Daniels model by means of a first order Taylor expansion around $\eta_1 = \eta_2 = \ldots = \eta_C = \eta_0$, where η_0 is any value. Below, we extend Henery's approach to any model in which $Y_1, Y_2, \ldots, Y_C$ have joint density of the form (5).

As there are $C!$ possible full rankings, the average full ranking probability is

$$p_* = \frac{1}{C!}. \tag{7}$$

Note that $p_\pi = p_*$ if $\eta_1 = \eta_2 = \ldots = \eta_C = \eta_0$.

Introduce

$$\phi_0(y) = \left.\frac{\partial \ln g(y; \eta)}{\partial \eta}\right|_{\eta=\eta_0} = \frac{1}{g(y; \eta_0)} \left.\frac{\partial g(y; \eta)}{\partial \eta}\right|_{\eta=\eta_0}. \tag{8}$$

We shall refer to $\phi_0(y)$ as the score function.

The score function is well-known in mathematical statistics, especially in likelihood theory and the theory of rank tests. (In particular, our definition (8) corresponds with Equation (I.2.4.4) in Hájek and Šidák (1967).) An important property of the score function is that the expected value is zero, that is

$$\mathcal{E}\phi_0(Y) = \int \left.\frac{\partial g(y;\eta)}{\partial \eta}\right|_{\eta=\eta_0} dy = 0. \tag{9}$$

Denote the expected score of the r^{th} order statistic $Y_{r:C}$ by

$$q_{r:C} = \mathcal{E}_0\phi_0(Y_{r:C}), \tag{10}$$

where $\mathcal{E}_0$ denotes the expectation under the condition that $\eta_1 = \eta_2 = \ldots = \eta_C = \eta_0$. When this condition holds, we show in Appendix A.1 that a first order Taylor expansion in $(\eta_0, \eta_0, \cdots, \eta_0)$ yields

$$p_\pi \approx p_* + p_* \sum_{r=1}^{C} q_{r:C}(\eta_{c_r} - \eta_0) = p_* \left(1 + \sum_{r=1}^{C} q_{r:C}\eta_{c_r}\right) \tag{11}$$

for $\pi = (c_1, c_2, \ldots, c_C)$.

For a given full ranking π, let $\mathbf{q}_\pi$ denote the C-dimensional vector containing the $\pi_{(c)}^{th}$ expected score $q_{\pi_{(c)}:C}$ as c^{th} element. For example, for $\pi_{(3)} = 7$, the 3^{th} element of vector $\mathbf{q}_\pi$ is the 7^{th} expected score $q_{7:C}$.

As $\mathbf{q}_\pi$ is central in deriving an approximation to the probability p_π, we shall refer to $\mathbf{q}_\pi$ as the expected score vector belonging to π. We may now write (11) as

$$p_\pi \approx p_* \left(1 + \mathbf{q}_\pi^t \eta\right). \tag{12}$$

The right hand side of (12) is not necessarily positive and hence, it does not necessarily define a valid probability model. However, when all η_c's are sufficient close to η_0

$$1 + \mathbf{q}_\pi^t \eta \approx \exp\left\{\mathbf{q}_\pi^t \eta\right\}, \tag{13}$$

and we may approximate p_π by

$$p_\pi \approx \frac{\exp\{\mathbf{q}_\pi^t \eta\}}{\sum_{\pi'} \exp\{\mathbf{q}_{\pi'}^t \eta\}}. \tag{14}$$

The probabilities on the right hand side of (14) are all positive and add up to one, and thus define a probability model with respect to the rankings.

Above, we have shown that expected scores allow the approximation of any Thurstone model. For example, one may show that in the Thurstone-Mosteller-Daniels model $q_{r:C}$ coincides with a normal score; that is,

$$q_{r:C} = \mathcal{E}Y_{r:C}, \tag{15}$$

where $Y_{r:C}$ denotes the r^{th} order statistic correponding to the sample $Y_1, Y_2, \ldots, Y_C$ drawn from a standard normal distribution. In Harter (1961) all normal scores for $C = 400$ are given.

Another Thurstone model is the Luce model, also known as Plackett's first order model, and one may show that in this model

$$q_{r:C} = 1 - \sum_{r'=r}^{C} \frac{1}{r'}. \tag{16}$$

3 A linear model for conjoint experiments

3.1 Incorporating attribute values

In the previous section, we have seen how a ranking model translates the predictor vector η into a probability distribution on rankings. In this section, we focus on the question how the attributes of the stimuli influence η. Assume that the stimuli are adequately described by means of M attributes. Each attribute takes a limited number of values, which we call levels. Every stimulus may be viewed as a specific combination of levels of the attributes. Let x_{cm} denote the value attribute m takes for stimulus c.

In order to be able to perform a statistical analysis of conjoint experimental data, we have to specify the construction of the predictor vectors. We assume that $\eta = (\eta_1, \eta_2, \ldots, \eta_C)^t$ is given by

$$\eta_c = \beta_1 x_{c1} + \beta_2 x_{c2} + \beta_M x_{cM} = \sum_{m=1}^{M} \beta_m x_{cm}, \tag{17}$$

where $\beta_1, \beta_2, \ldots, \beta_M$ are unknown coefficients. We may write $\eta = \mathbf{X}\beta$, where β is the M dimensional coefficient vector $(\beta_1, \beta_2, \ldots, \beta_M)^t$ and $\mathbf{X}$ is the $C \times M$ matrix which contains the value x_{cm} in its (c, m) location. We shall refer to $\mathbf{X}$ as the plan matrix of the conjoint experiment.

We may rewrite (14) as

$$p_\pi \approx \frac{\exp\{\mathbf{q}_\pi^t \mathbf{X}\beta\}}{\sum_{\pi'} \exp\{\mathbf{q}_{\pi'}^t \mathbf{X}\beta\}}. \tag{18}$$

Observe that $\mathbf{q}_\pi^t \mathbf{X}$ is in fact a weighted average of the columns of the plan matrix, where the weights are completely determined by the preferences in π. Combining (2) and (18) now yields

$$p_\varpi \approx \frac{\sum_{\pi \in \mathcal{S}_\varpi} \exp\{\mathbf{q}_\pi^t \mathbf{X}\beta\}}{\sum_{\pi'} \exp\{\mathbf{q}_{\pi'}^t \mathbf{X}\beta\}}. \tag{19}$$

Note that (18) is actually a special case of (19). Although in the next section we shall focus on partial rankings, full rankings are implicitly covered as well.

3.2 Approximate log-likelihood

In principle, the maximum likelihood estimator $\hat{\beta}$ of the parameter vector $\beta = (\beta_1, \beta_2, \ldots, \beta_M)^t$ may be obtained via maximization of the log-likelihood in the Thurstone model. As mentioned earlier, computing p_{π_j} requires the numerical evaluation of the C-dimensional integral (4) and hence, is not feasible when the number of stimuli becomes large.

Fortunately, we may approximate p_{π_j} by (19) and thus we may estimate β by maximizing the corresponding approximate log-likelihood

$$\ln \tilde{L}(\beta) = \sum_{j=1}^{J} \ln \left(\sum_{\pi \in \mathcal{S}_{\varpi_j}} \exp \left\{ \mathbf{q}_{\pi}^t \mathbf{X} \beta \right\} \right) - J \ln \left(\sum_{\pi'} \exp \left\{ \mathbf{q}_{\pi'}^t \mathbf{X} \beta \right\} \right), \quad (20)$$

where the rankings $\varpi_1, \varpi_2, \ldots, \varpi_J$ are independently obtained from J different respondents. Note that the standard likelihood theory applies as we have shown that (19) is a probability model itself.

In particular, in case of full rankings we have that the log-likelihood (20) simplifies to

$$\ln \tilde{L}(\beta) = \sum_{j=1}^{J} \left(\mathbf{q}_{\pi_j}^t \mathbf{X} \beta \right) - J \ln \left(\sum_{\pi'} \exp \left\{ \mathbf{q}_{\pi'}^t \mathbf{X} \beta \right\} \right). \quad (21)$$

Standard iterative methods for finding an estimator $\hat{\beta}$ maximizing the log-likelihood (20) require the first order derivatives of (20) with respect to β, and possibly the second order derivatives as well. The computation of these are given in Appendix A.2.

4 Illustration

We will illustrate our model with two data sets. First, we will compare our estimates to estimates obtained in Maydeu-Olivares and Böckenholt (2005). We analyze their career preference data set. Next, we will illustrate our methodology by incorporating attributes and examine how these attributes influences stimulus preferences.

We should first make a general remark: the smaller its rank, the more preferred a stimulus is. Hence, when interpreting the estimation results, we should always take into account that our preference measure is inversely related to preference. Hence, a positive coefficient indicates that higher levels lead to a higher, i.e. worse, ranking. Consequently, a negative coefficient indicates that higher levels lead to a lower, i.e. more preferred, ranking. The strength of preference is reflected in the absolute value of the coefficient. In addition, a positive coefficient value does not necessarily mean that the respective attribute is rejected, but that it is less preferable than the reference level.

4.1 Comparison data set

In Maydeu-Olivares and Böckenholt (2005) career preferences among undergraduate psychology students from a Spanish university were investigated. A sample of 57 psychology students were asked to rank their preferences for four broad psychology career areas: academia, clinical, educational and industrial. In Table 1 the estimated coefficients are reported. We have set career area industrial as reference level. Note that in this illustration we have only one attribute (career area) and this attribute has four levels (academia, clinical, educational and industrial).

Career area	Coefficient	Stand. Err.	p-value
Academic	1.110	0.276	0.000
Clinical	-1.106	0.269	0.000
Educational	-0.336	0.243	0.166

Table 1: Estimated coefficients career ranking data

Our estimates differ slightly from Maydeu-Olivares and Böckenholt (2005), but the conclusions are the same. Remark that the estimated coefficients are inversely related to preference. The estimated coefficients for career area clinical is negative, which means that the clinical career area is more preferred than the reference career area industrial. As the estimated coefficient of educational is not significantly different from the reference level, we can not conclude that educational is more preferred than industrial. The least preferred career area is academic, as this coefficient is positive which leads to higher ranks and hence worse ranking.

4.2 Data set incorporating attributes

In real conjoint experiments, we are often interested in how stimulus' characteristics influence preferences. We have collected data concerning winter sports holidays, where each winter sports holiday is described by a number of attributes: country levels, period of holiday, duration of holiday and size of the ski area.

As each attribute can take a limited number of values, every holiday may be viewed as a specific combination of levels of the attributes. The corresponding levels for country are: France, Austria, Italy and Andorra. The attributes period and duration have both two levels: January or February and eight days or ten days, respectively. Attribute ski area has three levels.

As our approach is not limited to full ranking data, as in the previous illustration, we demonstrate here that we are also able to handle partial ranking data. Each of $J = 169$ respondents are asked to indicate the most preferred alternative and the least preferred alternative.

We use dummy variables to code the attribute levels in the planmatrix $\mathbf{X}$ which was introduced in Subsection 3.1. As each stimulus belongs to exactly one factor level, the rows of the extended plan matrix sum to some constant. Combined with (17) this suggests that adding the same constant to each of the coefficients $\beta_1, \ldots, \beta_M$ amounts to adding this same contant to each of the predictors $\eta_1, \ldots, \eta_C$. However, as we have already seen in (11), it follows from (26) that this does not affect the approximate ranking probabilities. In short, there is an unwanted indeterminacy present in the model.

This indeterminacy may be resolved in different ways. We opt for setting the first level as reference level, that is, adding a constant $-\beta_1$ to each of the coefficients in the model, so as to make the first coefficient equal to zero. In effect, we use the matrix $\mathbf{X}$ obtained by removing the first level for each attribute from the extended plan matrix. The consequence is that we should now interpret each of the remaining coefficients relative to the omitted levels of the corresponding attributes. In some applications we have to make an arbitrary choice, as a natural candidate for the reference category is missing.

As our approach allows for a unified approach, that is, we can estimate any Thurstone order statistic model, we will give the estimated results in the Thurstone-Mosteller-Daniels model as well as the results estimated in the Luce model. Note that we approximate any Thurstone order statistic model by a simpler model. More precisely, when we are discussing the Thurstone-Mosteller-Daniels model or the Luce model (or any other Thurstone model), we mean an approximate model.

4.2.1 Results Thurstone-Mosteller-Daniels model

Table 2 shows the estimated coefficients. The reference levels are: France, January, eight-days holiday, small ski area. France is the most preferred country to spent the winter sports holidays according to the respondents,

Attribute	Variable	Coefficient	Stand. Err.	p-value
Country	Austria	1.619	0.188	0.000
	Italy	0.918	0.239	0.000
	Andorra	0.047	0.237	0.843
Period	February	-0.087	0.171	0.611
Duration	ten-days	0.573	0.175	0.001
Ski area	average	-0.660	0.172	0.000
	large	-1.941	0.258	0.000

Table 2: Estimated coefficients

as all other country level coefficients are positive. This does not necessarily mean that the respective attribute is rejected, but that it is less preferable than the reference level. The strength of preference is reflected in the absolute value of the coefficient, and thus Austria is the least preferred winter sports holiday country compared to France. Notice, that coefficient of level Andorra is not significant and there is no difference in preference between France and Andorra. In addition, the estimated coefficient for period is also not significant and hence, it makes no difference whether the holiday takes place in January or February. Respondents prefer an eight days holiday over a ten-days holiday. Ski area size has a positive effect on the ranking of the alternative as the absolute value of the coefficient increases as ski size increases.

Testing the independence model versus the Thurstone-Mosteller-Daniels model yields a chi-squared test statistic of 1134.373 with corresponding p-value of zero. Hence, at the 5% significance level, the independence model is clearly rejected in favor of the Thurstone-Mosteller-Daniels model.

As leaving out each attribute from the original model, gives us a nested model, we can examine the effect of each attribute by means of likelihood ratio tests. The likelihood ratio tests results are given in Table 3. One can observe in this table that attributes period and duration of holiday have no significant effect on the preference of the respondents as leaving each of these attribute out of the model does not lead to a significant better model. On the other side, including the attributes country and ski size area do lead to a significant better model and thus, these two attribute influences respondents' preferences significantly.

Attribute	Chi-square	df	p-value
country	114.259	3	0.000
period holiday	0.258	1	0.611
duration holiday	11.083	1	0.001
size ski area	65.402	2	0.000

Table 3: Likelihood ratio tests

4.2.2 Results Luce model

The Thurstone-Mosteller- Daniels model places the same emphasis on the lower ranks as on the higher ranks. In the Luce model greater emphasis is placed on the highly preferred, that is low-ranked, stimuli. In Figure 1 the expected scores for a sample of size 8 are plotted. Note the symmetry around zero for expected scores belonging to the Thurstone-Mosteller-Daniels model. In contrast, expected scores for highly preferred stimuli receive more emphasis in the Luce model.

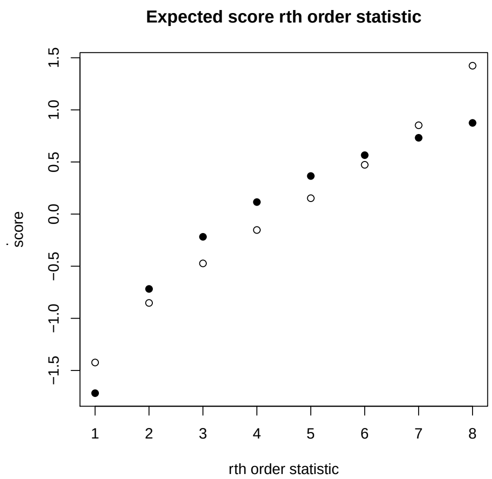

Figure 1: Expected score of r^{th} order statistic drawn from a sample of size 8. The open dots are Thurstone-Mosteller-Daniels scores (15) and the solid dots are Luce scores (16).

Table 4 shows the estimated coefficients. The estimates are just slightly different and the conclusion remains the same as in the Thurstone-Mosteller-

Attribute	Variable	Coefficient	Stand. Err.	p-value
Country	Austria	1.403	0.188	0.000
	Italy	0.767	0.330	0.020
	Andorra	-0.152	0.269	0.571
Period	February	-0.181	0.219	0.408
Duration	ten-days	0.646	0.199	0.001
Ski area	average	-0.855	0.228	0.000
	large	-1.871	0.311	0.000

Table 4: Estimated coefficients in the Luce model

Daniels model. Note that though the estimated coefficient of Andorra has the opposite sign but this is however not significant.

The independence model is clearly rejected in favor of the Luce model as the chi-squared test statistic is 1152.119 with corresponding p-value of zero. Also in the Luce model we test the importance of each attribute by means of likelihood ratio tests, see Table 5. We draw again the same conclusion: attributes country and ski size area lead to a significant better model.

Attribute	Chi-square	df	p-value
country	104.645	3	0.000
period holiday	0.691	1	0.708
duration holiday	11.885	1	0.003
size ski area	45.003	2	0.000

Table 5: Likelihood ratio tests in the Luce model

5 Conclusion

Rankings are a simple tool to measure preferences. Metric measurements such as rating and matching may be less reliable due to respondents limited ability to accurately report degrees of preferences (Ben-Akiva et al. (1992)).

New efficient data collecting methods, such as Best-Worst ranking, become more popular nowadays in practical conjoint analysis. It has been well known that task difficulty increases substantially with the number of stimuli to be ranked. Partial rankings reduce task complexity for respondents. This requires new methods to analyze these partial rankings data. We have shown that our model is also able to handle any partial rankings, not limited to Best-Worst rankings.

In this paper we have introduced approximate ranking models in the context of conjoint experiments. We have shown that by computing approximate rather than exact ranking probabilities reduces the complexity considerably. We extend the approximation technique proposed by Henery (1981) for any Thurstone order statistics model and our approach allows a unified approach. We have shown how we could incorporate attribute values as is usual in conjoint experiments to estimate the effect of attribute levels on respondents' choice.

In recent marketing literature the respondents' heterogeneity is an important topic in analyzing respondents' preference behavior. In further research it would be interesting to incorporate respondent's heterogeneity in our model.

Another interesting extension of the model for further research amounts adding nuisance parameters to embed in larger family. Recall that the Luce model may be viewed as a Thurstone model derived from the Gumbel distribution, which is a special case of a generalized extreme value cumulative distribution.

References

Moshe Ben-Akiva, Takayuki Morikawa, and Fumiaki Shiroishi. Analysis of the reliability of preference ranking data. *Journal of Business Research*, 24:149–164, 1992.

Ulf Böckenholt. Thurstonian-based analyses: Past, present, and future utilities. *Psychometrika*, 71(4):615–629, 2006. 10.1007/s11336-006-1598-5.

Eric T. Bradlow. Current issues and a wish list for conjoint analysis. *Applied Stochastic Models in Business and Industry*, 21:319–323, 2005. doi: 10.1002/asmb.559.

Douglas E. Critchlow, Michael A. Fligner, and Joseph S. Verducci. Probability models on rankings. *Journal of Mathematical Psychology*, 35(3): 294–318, 1991. doi: 10.1016/0022-2496(91)90050-4.

H. E. Daniels. Rank correlation and population models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 12(2):171–191, 1950. URL http://www.jstor.org/stable/2983980.

Paul E. Green, Abba M. Krieger, and Yoram Wind. Thirty years of conjoint analysis: Reflections and prospects. *Interfaces*, 31(3(2)):56–73, May 2001. URL `http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=5674961&site=ehost-live`.

Jaroslav Hájek and Zbyněk Šidák. *Theory of rank tests*. Academic Press, New York, 1967.

H. Leon Harter. Expected values of normal order statistics. *Biometrika*, 48 (1/2):151–165, 1961.

John R. Hauser and Vithala R. Rao. Conjoint analysis, related modeling, and applications. In Paul Edgar Green and Yoram J. Wind, editors, *Marketing research and modeling: progress and prospects : a tribute to Paul E. Green*, International series in quantitative marketing, pages 141–168. Kluwer Academic Publishers, Dordrecht, 2004. ISBN 1-402-07596-0.

R. J. Henery. Permutation probabilities as models for horse races. *Journal of the Royal Statistical Society. Series B (Methodological)*, 43(1):86–91, 1981. URL `http://www.jstor.org/stable/2985154`.

J. B. Kruskal. Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 27(2):251–263, 1965. URL `http://www.jstor.org/stable/2984194`.

R. D. Luce. *Individual choice behavior, a theoretical analysis*. Wiley, New York, 1959.

R. D. Luce. Thurstone and sensory scaling: then and now. *Psychological Review*, 101(2):271–277, 1994.

R. Duncan Luce and John W. Tukey. Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1(1):1–27, 1964. doi: 10.1016/0022-2496(64)90015-X.

John I. Marden. *Analyzing and modeling rank data*, volume 64 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1995. ISBN 0-412-99521-2.

Albert Maydeu-Olivares. Thurstonian modeling of ranking data via mean and covariance structure analysis. *Psychometrika*, 64(3):325–340, 1999. doi: 10.1007/BF02294299.

Albert Maydeu-Olivares and Ulf Böckenholt. Structural equation modeling of paired-comparison and ranking data. *Psychological Methods*, 10(3):285–304, Sep 2005.

Alberto Maydeu-Olivares and Adolfo Hernández. Identification and small sample estimation of thurstone's unrestricted model for paired comparisons data. *Multivariate Behavioral Research*, 42(2):323–347, 2007.

Frederick Mosteller. Remarks on the method of paired comparisons: I. the least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, 16(1):3–9, 1951. doi: 10.1007/BF02313422.

L. L. Thurstone. A law of comparative judgment. *Psychological Review*, 34 (4):273–286, Jul 1927. doi: 10.1037/h0070288.

A Appendix

A.1 Approximate probability (11)

Define

$$\phi(y;\eta) = \frac{\partial \ln g(y;\eta)}{\partial \eta} = \frac{1}{g(y;\eta)} \frac{\partial g(y;\eta)}{\partial \eta}, \tag{22}$$

and remark that

$$\frac{\partial g(y;\eta)}{\partial \eta} = \phi(y;\eta) g(y;\eta).$$

Let $g_0(y)$ and $\phi_0(y)$ denote $g(y;\eta_0)$ and $\phi(y;\eta_0)$, respectively. Recall that $\phi_0(y)$ is introduced in (8) as the score function.

As

$$\begin{aligned}
\frac{\partial p_\pi}{\partial \eta_{c_r}} &= \int_{-\infty}^{\infty} \int_{y_{c_1}}^{\infty} \cdots \int_{y_{c_{C-1}}}^{\infty} g\left(y_{c_1};\eta_{c_1}\right) g\left(y_{c_2};\eta_{c_2}\right) \\
&\quad \cdots g\left(y_{c_{r-1}};\eta_{c_{r-1}}\right) \frac{\partial g(y_{c_r};\eta_{c_r})}{\partial \eta_{c_r}} g\left(y_{c_{r+1}};\eta_{c_{r+1}}\right) \\
&\quad \cdots g\left(y_{c_C};\eta_{c_C}\right) dy_{c_C} \cdots dy_{c_2} dy_{c_1} \\
&= \int_{-\infty}^{\infty} \int_{y_{c_1}}^{\infty} \cdots \int_{y_{c_{C-1}}}^{\infty} g\left(y_{c_1};\eta_{c_1}\right) g\left(y_{c_2};\eta_{c_2}\right) \\
&\quad \cdots g\left(y_{c_{r-1}};\eta_{c_{r-1}}\right) \phi\left(y_{c_r};\eta_{c_r}\right) g\left(y_{c_r};\eta_{c_r}\right) g\left(y_{c_{r+1}};\eta_{c_{r+1}}\right) \\
&\quad \cdots g\left(y_{c_C};\eta_{c_C}\right) dy_{c_C} \cdots dy_{c_2} dy_{c_1},
\end{aligned} \tag{23}$$

we obtain

$$\left.\frac{\partial p_\pi}{\partial \eta_{c_r}}\right|_{\eta_1=\eta_2=\ldots=\eta_C=\eta_0} = \int_{-\infty}^{\infty}\int_{y_{1:C}}^{\infty}\cdots\int_{y_{C-1:C}}^{\infty}\phi_0\left(y_{r:C}\right)$$
$$g_0\left(y_{1:C}\right)g_0\left(y_{2:C}\right)\cdots g_0\left(y_{C:C}\right)dy_{C:C}\cdots dy_{2:C}dy_{1:C}. \tag{24}$$

Let $Y_{1:C} < Y_{2:C} < \cdots < Y_{C:C}$ be the order statistics of a random sample of size C from a density g_0. Recall that the joint density of $Y_{1:C}, Y_{2:C}, \cdots, Y_{C:C}$ equals

$$C!g_0\left(y_{1:C}\right)g_0\left(y_{2:C}\right)\cdots g_0\left(y_{C:C}\right)$$

for $y_{1:C} < y_{2:C} < \cdots < y_{C:C}$. It follows that

$$\left.\frac{\partial p_\pi}{\partial \eta_{c_r}}\right|_{\eta_1=\eta_2=\ldots=\eta_C=\eta_0} = \frac{q_{r:C}}{C!} = p_* q_{r:C}, \quad \text{with} \quad q_{r:C} = \mathcal{E}\phi_0\left(Y_{r:C}\right), \tag{25}$$

and p_* given by (7). Since $\sum_{r=1}^{C}\phi_0\left(Y_{r:C}\right)$ coincides with $\sum_{r=1}^{C}\phi_0\left(Y_r\right)$, (9) implies

$$\sum_{r=1}^{C} q_{r:C} = \sum_{r=1}^{C}\mathcal{E}\phi_0\left(Y_{r:C}\right) = \mathcal{E}\sum_{r=1}^{C}\phi_0\left(Y_{r:C}\right) = \mathcal{E}\sum_{r=1}^{C}\phi_0\left(Y_r\right) = \sum_{r=1}^{C}\mathcal{E}\phi_0\left(Y_r\right) = 0. \tag{26}$$

When all η_c's are close to η_0, a first order Taylor expansion in $(\eta_0, \eta_0, \cdots, \eta_0)$ yields

$$p_\pi \approx p_* + p_*\sum_{r=1}^{C} q_{r:C}\left(\eta_{c_r} - \eta_0\right) = p_*\left(1 + \sum_{r=1}^{C} q_{r:C}\eta_{c_r}\right) \tag{27}$$

for $\pi = (c_1, c_2, \ldots, c_C)$. The equality follows from 26.

A.2 First and second order derivatives of (20)

Standard iterative methods for finding an estimator $\hat{\beta}$ maximizing the log-likelihood (20) require the first order derivatives of (20) with respect to β, and possibly the second order derivatives as well. Write p_ϖ as $\sum_{\pi\in\mathcal{S}_\varpi} s_\pi / \sum_{\pi'} s_{\pi'}$ with $s_\pi = \exp\{\mathbf{q}_\pi^t\mathbf{X}\beta\}$. As $(\partial/\partial\beta)s_\pi = s_\pi\mathbf{X}^t\mathbf{q}_\pi$, it follows that

$$\begin{aligned}
\frac{\partial p_\varpi}{\partial\beta} &= \frac{\partial}{\partial\beta}\frac{\sum_{\pi\in\mathcal{S}_\varpi} s_\pi}{\sum_{\pi'} s_{\pi'}} \\
&= \frac{\sum_{\pi\in\mathcal{S}_\varpi} s_\pi\mathbf{X}^t\mathbf{q}_\pi}{\sum_{\pi'} s_{\pi'}} - \frac{\sum_{\pi\in\mathcal{S}_\varpi} s_\pi}{\sum_{\pi'} s_{\pi'}}\cdot\frac{\sum_{\pi} s_\pi\mathbf{X}^t\mathbf{q}_\pi}{\sum_{\pi'} s_{\pi'}} \\
&= \mathbf{X}^t\left(\sum_{\pi\in\mathcal{S}_\varpi} p_\pi\mathbf{q}_\pi - p_\varpi\sum_{\pi'} p_{\pi'}\mathbf{q}_{\pi'}\right),
\end{aligned} \tag{28}$$

and in particular,

$$\frac{\partial p_\pi}{\partial \beta} = \mathbf{X}^t p_\pi \left(\mathbf{q}_\pi - \sum_{\pi'} p_{\pi'} \mathbf{q}_{\pi'} \right). \tag{29}$$

For any set $\mathcal{S}$ of rankings, (29) yields

$$\begin{aligned}
\frac{\partial}{\partial \beta} \frac{\sum_{\pi\in\mathcal{S}} p_\pi \mathbf{q}_\pi^t}{\sum_{\pi\in\mathcal{S}} p_\pi} &= \frac{\sum_{\pi\in\mathcal{S}} \frac{\partial p_\pi}{\partial\beta} \mathbf{q}_\pi^t}{\sum_{\pi\in\mathcal{S}} p_\pi} - \frac{\sum_{\pi\in\mathcal{S}} \frac{\partial p_\pi}{\partial\beta}}{\sum_{\pi\in\mathcal{S}} p_\pi} \cdot \frac{\sum_{\pi\in\mathcal{S}} p_\pi \mathbf{q}_\pi^t}{\sum_{\pi\in\mathcal{S}} p_\pi} \\
&= \frac{\sum_{\pi\in\mathcal{S}} \mathbf{X}^t p_\pi \left(\mathbf{q}_\pi - \sum_{\pi'} p_{\pi'}\mathbf{q}_{\pi'}\right) \mathbf{q}_\pi^t}{\sum_{\pi\in\mathcal{S}} p_\pi} \\
&\quad - \frac{\sum_{\pi\in\mathcal{S}} \mathbf{X}^t p_\pi \left(\mathbf{q}_\pi - \sum_{\pi'} p_{\pi'}\mathbf{q}_{\pi'}\right)}{\sum_{\pi\in\mathcal{S}} p_\pi} \cdot \frac{\sum_{\pi\in\mathcal{S}} p_\pi \mathbf{q}_\pi^t}{\sum_{\pi\in\mathcal{S}} p_\pi} \\
&= \mathbf{X}^t \left(\frac{\sum_{\pi\in\mathcal{S}} p_\pi \mathbf{q}_\pi \mathbf{q}_\pi^t}{\sum_{\pi\in\mathcal{S}} p_\pi} - \frac{\sum_{\pi\in\mathcal{S}} p_\pi \mathbf{q}_\pi}{\sum_{\pi\in\mathcal{S}} p_\pi} \cdot \frac{\sum_{\pi\in\mathcal{S}} p_\pi \mathbf{q}_\pi^t}{\sum_{\pi\in\mathcal{S}} p_\pi} \right),
\end{aligned} \tag{30}$$

and in particular,

$$\frac{\partial}{\partial\beta} \sum_\pi p_\pi \mathbf{q}_\pi^t = \mathbf{X}^t \left\{ \sum_\pi p_\pi \mathbf{q}_\pi \mathbf{q}_\pi^t - \left(\sum_\pi p_\pi \mathbf{q}_\pi\right) \left(\sum_\pi p_\pi \mathbf{q}_\pi^t\right) \right\}, \tag{31}$$

It now follows from (28) that

$$\begin{aligned}
\frac{\partial \ln \tilde{L}(\beta)}{\partial\beta} &= \sum_{j=1}^J \frac{\partial}{\partial\beta} \ln p_{\varpi_j} = \sum_{j=1}^J \frac{\frac{\partial}{\partial\beta} p_{\varpi_j}}{p_{\varpi_j}} \\
&= \sum_{j=1}^J \mathbf{X}^t \left(\frac{\sum_{\pi\in\mathcal{S}_{\varpi_j}} p_\pi \mathbf{q}_\pi}{\sum_{\pi\in\mathcal{S}_{\varpi_j}} p_\pi} - \sum_\pi p_\pi \mathbf{q}_\pi \right) \\
&= \mathbf{X}^t \left(\sum_{j=1}^J \frac{\sum_{\pi\in\mathcal{S}_{\varpi_j}} p_\pi \mathbf{q}_\pi}{\sum_{\pi\in\mathcal{S}_{\varpi_j}} p_\pi} - J \sum_\pi p_\pi \mathbf{q}_\pi \right).
\end{aligned} \tag{32}$$

Similarly, it follows from (30) and (31) that

$$\begin{aligned}\frac{\partial^2 \ln \tilde{L}(\beta)}{\partial \beta^2} &= \frac{\partial}{\partial \beta}\left(\sum_{j=1}^{J} \frac{\sum_{\pi \in \mathcal{S}_{\varpi_j}} p_\pi \mathbf{q}_\pi^t}{\sum_{\pi \in \mathcal{S}_{\varpi_j}} p_\pi} - J \sum_{\pi} p_\pi \mathbf{q}_\pi^t\right)\mathbf{X} \\ &= \left(\sum_{j=1}^{J} \frac{\partial}{\partial \beta} \frac{\sum_{\pi \in \mathcal{S}_{\varpi_j}} p_\pi \mathbf{q}_\pi^t}{\sum_{\pi \in \mathcal{S}_{\varpi_j}} p_\pi} - J \frac{\partial}{\partial \beta} \sum_{\pi} p_\pi \mathbf{q}_\pi^t\right)\mathbf{X} \\ &= \mathbf{X}^t \left(\sum_{j=1}^{J} \left\{ \frac{\sum_{\pi \in \mathcal{S}_{\varpi_j}} p_\pi \mathbf{q}_\pi \mathbf{q}_\pi^t}{\sum_{\pi \in \mathcal{S}_{\varpi_j}} p_\pi} - \frac{\sum_{\pi \in \mathcal{S}_{\varpi_j}} p_\pi \mathbf{q}_\pi}{\sum_{\pi \in \mathcal{S}_{\varpi_j}} p_\pi} \cdot \frac{\sum_{\pi \in \mathcal{S}_{\varpi_j}} p_\pi \mathbf{q}_\pi^t}{\sum_{\pi \in \mathcal{S}_{\varpi_j}} p_\pi} \right\} \right. \\ &\quad \left. - J \left\{ \sum_{\pi} p_\pi \mathbf{q}_\pi \mathbf{q}_\pi^t - \left(\sum_{\pi} p_\pi \mathbf{q}_\pi\right)\left(\sum_{\pi} p_\pi \mathbf{q}_\pi^t\right) \right\} \right)\mathbf{X}. \qquad (33)\end{aligned}$$

In particular, in case of full rankings we have that the log-likelihood (20) simplifies to 21. Consequently, (32) becomes

$$\frac{\partial \ln \tilde{L}(\beta)}{\partial \beta} = \mathbf{X}^t \left(\sum_{j=1}^{J} \mathbf{q}_{\pi_j}^t - J \sum_{\pi} p_\pi \mathbf{q}_\pi^t\right),$$

and (33) becomes

$$\frac{\partial^2 \ln \tilde{L}(\beta)}{\partial \beta^2} = \mathbf{X}^t \left(-J \left\{ \sum_{\pi} p_\pi \mathbf{q}_\pi \mathbf{q}_\pi^t - \left(\sum_{\pi} p_\pi \mathbf{q}_\pi\right)\left(\sum_{\pi} p_\pi \mathbf{q}_\pi^t\right) \right\}\right)\mathbf{X}.$$