Latent structure models for ranking data

Croon, M.A.; Luijkx, R.

Published in:
Probability models and statistical analyses for ranking data

Publication date:
1993

4

Latent Structure Models for Ranking Data

M. A. Croon and R. Luijkx [1]

ABSTRACT In this paper several latent structure models for analyzing data that consist of complete or incomplete rankings are discussed. First, attention is given to some latent class extensions of the Bradley-Terry-Luce model for ranking data. Next, various latent class models based on log-linear modeling of ranking data are described. Within this latter family of latent class models, a main distinction is made between models based on the assumption of quasi-independence within the latent classes, and models in which some form of association between the ranking positions is allowed to exist within the classes. All models are applied to a real data set from a large scale cross-national survey on political values.

4.1 Introduction

Latent Structure Models

Latent structure models are extensively used in the social and behavioral sciences, and their popularity in these circles is easily explained. One of the main problems with which empirical research in these sciences has to cope pertains to the imperfect and unreliable way in which theoretically important constructs are 'measured' or operationalized. Concepts such as 'intelligence' 'neuroticism', 'group cohesiveness',or 'political trust' simply elude direct measurement, and variation among respondents on such theoretical constructs can only be assessed by means of imperfect indicator variables. These indicator variables hopefully reflect variation on the underlying theoretical concept, but are probably also influenced by a host of other irrelevant disturbing factors. As a consequence, empirical investigators in the social and behavioral sciences have long been interested in methods by means of which the relation between underlying unobservable latent variables and observable manifest variables can be described and analyzed, and that is exactly what latent structure models do.

Depending upon the nature of the manifest and latent variables, many

[1] Faculty of Social Sciences, Tilburg University, Tilburg, The Netherlands

different forms of latent structure models may be formulated. By way of *factor analysis* (or by means of the related technique of *covariance structure analysis*) one may analyze the correlation or covariance structure among a large number of quantitative manifest variables in terms of a relatively small number of quantitative latent variables. *Latent trait models*, on the other hand, aim at the analysis of categorical (mostly dichotomous) responses to aptitude or attitude items in terms of underlying continuous latent traits. Finally, *latent class analysis* was developed to analyze the association between qualitative variables.

Although in social and behavioral research respondents are quite often asked to rank a given set of alternatives on a particular evaluation criterion, no special attention has yet been paid to the problem of developing latent structure models for ranking data. In this paper, we will describe several latent structure models for ranking data and illustrate the usefulness of these methods. The basic idea behind all models that we will discuss is that a heterogeneous population of respondents may be partitioned into a small number of homogeneous subpopulations, within each of which the choice or ranking processes are assumed to satisfy a relatively simple model. Seen in this way, these latent structure models are instances of *finite mixture models.*

Ranking Tasks: Some Notation and Terminology

Assume that a set of n stimuli is presented to the subjects who are instructed to select and rank the m alternatives which, in their view, score highest on the evaluation criterion defined by the investigator. Such a ranking task will be called a 'rank m out of n' task. If $m = n - 1$, we obtain complete rankings of the stimuli; for $m < n - 1$, the rankings are incomplete. In this paper we assume that ties are not allowed in the rankings. If we denote the alternatives by the first n integers, and arbitrary alternatives by either subscripted or unsubscripted symbols as i, j, k and l, the respondents' rankings can be represented by ordered $m-$tuples $(i_1, i_2, \cdots, i_m)$. In this m-tuple, i_1 represents the alternative that occupies the first position in the ranking, i_2 represents the alternative that occupies the second position in the ranking, etc. In general, i_r represents the alternative that occupies position r in the ranking, with $1 \leq r \leq m$.

The Data for Illustration

All the models described in this paper will be illustrated on data obtained from the cross-national survey *Political Action* (See [1]). In this survey respondents from five different western countries (West-Germany, The Netherlands, the United States, Great-Britain and Austria) were asked to select and rank their three most preferred alternatives from the following set of eight political goals:

1. Maintain a high rate of economic growth.

2. Make sure that this country has strong defense forces.

3. Give people more say in how things are decided at work and in their country.

4. Try to make our cities and countryside more beautiful.

5. Maintain a stable economy.

6. Fight against crime.

7. Move toward a friendlier, less impersonal society.

8. Move toward a society where ideas are more important than money.

In this paper only the U.S. data will be used.

The selection of these eight political goals was based on Inglehart's theory of value orientations in which a clear distinction is drawn between a materialistic and a post-materialistic value orientation (See [8]). The materialistic value orientation is characterized by a strong concern for social and economic stability, while the post-materialistic value orientation is mainly concerned with the more humane, ecological and spiritual aspects of social life. In this respect, it is clear that the political goals 1,2,5 and 6 tap the materialistic value orientation, whereas the remaining goals 3,4,7 and 8 tap the post-materialistic value orientation.

The ranking in which alternative i is in the first, alternative j in the second and alternative k in the third position will be denoted by the ordered triple (i, j, k). Its observed frequency will be denoted by f_{ijk}, and its theoretical probability by p_{ijk}.

For all models discussed in this paper specific FORTRAN computer programs were developed since none of the available standard packages for log-linear and latent class analysis seemed capable of dealing in an efficient way with the particular features shown by ranking data. As we shall see, the fact that particular patterns of structural zeros emerge if one summarizes ranking data in the form of a contingency table has to be taken into account in a log-linear and latent class analysis of ranking data. Upon request these program codes are available from the first author.

4.2 Latent Class Analyses Based on the Bradley-Terry-Luce Model

The BTL choice model

Although the Bradley-Terry-Luce model (in what follows, the BTL- model, for short) was first introduced by Bradley and Terry [4] in 1952 as a statistical model for analyzing choices between pairs of stimuli, its history seems to date back to at least 1929, when the set-theoretician Zermelo [16] arrived at basically the same model in an attempt to develop a mathematically sound way to rank chess masters on the basis of the results of round-robin tournaments. As a model for individual choice behavior, the BTL-model was thoroughly investigated by Luce [9] in his monograph *'Individual Choice Behavior'*. Luce showed how the BTL-model may be derived from an *Axiom of Choice*.

Let S denote the set of alternatives used in a choice experiment and let R be a subset of S: $R \subseteq S$. Let i be an arbitrary element of R, and hence of S. Let $p_R(i)$ and $p_S(i)$ denote the probabilities of selecting item i from either R or S, and let $P_S(R)$ represents the probability that one of the elements of R is selected when the entire set S of alternatives is presented. Then, Luce's choice axiom states that the choice probabilities satisfy the following condition:

$$p_S(i) \quad = \quad p_S(R).p_R(i)$$

Luce [9] showed that if a subject's choices satisfy this choice axiom, there exists a scale on which each alternative i has a (positive) scale value u_i such that:

$$p_R(i) \quad = \quad \frac{u_i}{\sum_{j \in R} u_j}$$

The scale values u_i are uniquely defined up to multiplication by a positive constant. In the case of a pairwise choice between alternatives i and j, we have $R = \{i, j\}$, and, hence, if p_{ij} denotes the probability of choosing i over j, we have:

$$p_{ij} \quad = \quad \frac{u_i}{u_i + u_j}$$

The BTL-model can be parametrized in another way. By defining

$$a_i = \ln\ u_i,$$

we obtain:

$$p_R(i) \quad = \quad \frac{\exp a_i}{\sum_{j \in R} \exp a_j}$$

with $-\infty < a_i < +\infty$. For pairwise choices, we have:

$$p_{ij} \quad = \quad \frac{\exp a_i}{\exp a_i + \exp a_j}$$

We will now discuss two different extensions of the BTL-model which have been proposed in the past for the analysis of rankings.

The BTL-model as a random utility model

The first adaptation starts from the well known fact that the BTL-model is compatible with a particular random utility model as defined in [2] or [10]. This point has been thoroughly investigated by Yellott [12, 13], but was already signaled by Bradley [3]. The basic assumptions of random utility models may be stated in the following way. Every time a stimulus is presented to a subject, it elicits a subjective impression of worth or value. The magnitude of this subjective impression may be represented by a real number. Instead of assuming that a stimulus always elicits the same subjective impression, one assumes that the magnitude of the subjective impression is a random variable. Let $\tilde{U}_i$ represent the random variable that corresponds to the fluctuating subjective impressions elicited by stimulus i. Then, the probability that alternative i will be chosen from set R is given by

$$p_R(i) \quad = \quad Prob(\tilde{U}_i = \max_{k \in R} \tilde{U}_k)$$

For pairwise choices we obtain

$$p_{ij} \quad = \quad Prob(\tilde{U}_i \geq \tilde{U}_j)$$

The BTL-model is compatible with the random utility model in which the random variables $\tilde{U}_i$ are independently distributed as extreme value distributions with constant scale parameters, but with possibly different location parameters. Without loss of generality, we may set the constant scale parameter equal to one, and obtain the following expression for the density function of the extreme value distribution for the random variable $\tilde{U}_i$:

$$f(u_i) \quad = \quad \exp\left[-(u_i - a_i) - \exp(u_i - a_i)\right]$$

for $-\infty < u_i < +\infty$, and in which a_i is the location parameter of the distribution.

Under this interpretation of the BTL-model, one easily derives expressions for the ranking probabilities in a ranking task. The probability $p_{i_1,i_2,\cdots,i_m}$ that in a 'm out of n' ranking task the incomplete ranking $(i_1, i_2, \cdots, i_m)$ is observed is given by:

$$p_{i_1,i_2,\cdots,i_m} \quad = \quad Prob(\tilde{U}_{i_1} \geq \tilde{U}_{i_2} \geq \cdots \geq \tilde{U}_{i_m} \geq \max_{k \notin \{i_1,i_2,\cdots,i_m\}} \tilde{U}_k)$$

Let $\mathcal{I} = \{1, 2, \cdots, m\}$ and define

$$\mathcal{J}_r \quad = \quad \mathcal{I} \backslash \{i_1, i_2, \cdots, i_{r-1}\}$$

for a given ordering $(i_1, i_2, \cdots, i_m)$. Note that $\mathcal{J}_1 = \mathcal{I}$.

If the random variables $\tilde{U}_i$ follow independent extreme value distributions, one may prove

$$p_{i_1,i_2,\cdots,i_m} = \prod_{r=1}^{m} \left(\frac{\exp a_{i_r}}{\sum_{j\in\mathcal{J}_r}^{n} \exp a_j} \right)$$

For a 'rank three out of n' task, the expressions for the ranking probabilities simplify to:

$$p_{ijk} = \frac{\exp a_i}{\exp a_i + \exp a_j + \exp a_k} \times \frac{\exp a_j}{\exp a_j + \exp a_k}$$

This expression shows that under this random utility BTL ranking model the probability of obtaining a particular ranking such as (i, j, k) is given by the product of the probability of selecting i from $\{i, j, k\}$ and the probability of selecting j from the set that remains after the first selection has been made, i.e. the probability of selecting j from $\{j, k\}$. A similar interpretation of ranking probabilities as products of successive selection probabilities also applies in the general case of a 'rank m out of n' task.

The Pendergrass-Bradley approach

Pendergrass and Bradley [11] proposed a different extension of the BTL-model to the analysis of rankings. In the case the subjects are required to rank three alternatives $\{i, j, k\}$, these authors assume that the probability of obtaining the complete ranking (i, j, k) is proportional to the product of the three paired comparison probabilities which are induced by the ranking:

$$p_{ijk} = C \cdot p_{ij} \cdot p_{ik} \cdot p_{jk}$$

The proportionality constant C is chosen so that the sum of all ranking probabilities p_{ijk} is equal to one.

If the paired comparison probabilities satisfy the BTL-model, one may derive

$$p_{ijk} = \frac{\exp(2a_i + a_j)}{\sum_{r,s\neq r} \exp(2a_r + a_s)}$$

By applying the basic principle of this approach, we obtain for the probability that the incomplete ranking $(i_1, \cdots, i_m)$ is observed in a 'rank m out of n' task the following expression:

$$p_{i_1,\cdots,i_m} = \frac{\exp\left(\sum_{r=1}^{m}(n-r)a_{i_r}\right)}{Q}$$

in which Q is the sum of terms like $\exp\left(\sum_{r=1}^{m}(n-r)a_{i_r}\right)$ over all possible incomplete rankings .

Latent class models for the analysis of ranking data based on the BTL model

For a discussion on how to obtain the maximum likelihood estimates of the scale parameters a_i under both models, and on how to test their statistical fit, we refer to [5]. Unfortunately, application of these methods to data from large surveys seldom results in an acceptable fit. The main reason for this consistent negative result probably lies in the fact that these models are unable to capture 'differences of opinion' in large populations, which are usually quite heterogeneous with respect to social and political attitudes.

In an attempt to extend the applicability of the BTL-model to the analysis of rankings in large samples from heterogeneous populations, Croon [5, 6] eveloped finite mixture models in which the BTL ranking models are coupled with the basic assumptions of latent class models. The point of departure of this approach is the assumption that the original heterogeneous population from which the respondents were sampled can be partitioned into a relatively small number of homogeneous subpopulations, the latent classes. Each respondent is assumed to belong to exactly one of these latent classes, but latent class membership is an unobserved variable. Assume that T latent classes are needed in a particular analysis and let t denote an arbitrary class. The scale values of alternative i in latent class t will be denoted by a_{it}. Let $\rho = (i_1, \cdots, i_m)$ be an arbitrary incomplete ranking. If we denote the probability of obtaining ranking ρ in latent class t by $p_{\rho,t}$, we obtain for the random utility ranking model:

$$p_{\rho,t} = \prod_{r=1}^{m} \left(\frac{\exp a_{i_r t}}{\sum_{j \in \mathcal{J}_r} \exp a_j} \right)$$

If π_t denotes the proportion of subjects belonging to latent class t, we derive

$$p_\rho = \sum_{t=1}^{T} p_{\rho,t} \cdot \pi_t$$

for the probability p_ρ of obtaining ranking ρ in a random sample from the entire population.

Similar expressions hold for the PB ranking models. For more information on these latent class models, and on the way in which the model parameters can be estimated by means of an E-M algorithm, we refer to [5].

An illustration

We give here the results of some analyses on the incomplete rankings of the eight political goals in the US sample (N=2090). These analyses were based on the random utility adaptation of the BTL-model. (We will not discuss the results of the analysis using the Pendergrass-Bradley approach, which gave very similar results.) The number of latent classes was systematically varied from 1 to 6. In the following table we give for each latent class number the log likelihood ratio statistic and the associated number of degrees of freedom. By means of the log likelihood ratio one tests the hypothesis that the model with a particular number of classes provides an acceptable description of the data against the general alternative that the set of ranking frequencies are multinomially distributed. This log likelihood ratio statistics is asymptotically distributed as a chi square distribution with the corresponding number of degrees of freedom. The general formula for computing the degrees of freedom is: $n(n-1)(n-2) - nT$, with n being the number of alternatives and T the number of latent classes.

t	L	df
1	1073.31	328
2	573.64	320
3	488.08	312
4	429.74	304
5	401.78	296
6	377.51	288

From this table we see that the value of the log likelihood statistic drastically decreases when the number of classes is increased, but, unfortunately, even the solution with six classes fails to provide a statistically acceptable fit to the data. Presumably, the latent class model based on this adaptation of the BTL model still remains a much too simple model to capture the diversity of political attitudes in the U.S. sample. Although we have certainly to reject the two-classes solution, it may be of some interest to take a closer look at it. If Inglehart's theory on value orientations is correct, one expects that one of the latent classes would represent the 'materialistic' respondents while the other would represent the 'post-materialists'. The following table gives the scale values of the eight political goals in the two classes.

i	*Class* 1	*Class* 2
1	−.05	−1.49
2	.77	−1.17
3	.00	.70
4	−1.08	−.31
5	1.33	.45
6	1.07	.15
7	−1.64	.56
8	−.40	1.12

In latent class 1, the materialistic alternatives 2, 5 and 6 score relatively high, while the post-materialistic items 4,7, and to a lesser extent also alternative 8, score low. The first latent class seems to represent the materialistic respondents. The interpretation of the second latent class as the subpopulation of post-materialistic respondents is probably also quite adequate since in this class the post-materialistic items 3, 7 and 8 score high, while the materialistic items 1, 2, and to a lesser extent alternative 6, score low. However, note that not all items conform to the expected pattern:

- In class 1 item 1 scores too low, whereas item 3 scores too high.
- In class 2 item 4 scores too low, whereas item 5 scores too high.

4.3 Latent Class Analyses Based on a Quasi-independence Model

Log-linear models for ranking probabilities.

In search for more flexible latent class models, a study of the log-linear analysis of ranking data was made. For more information on the log-linear analysis of 'rank 3 out of n' data, we refer to [7], but see also [12, 13] for similar ideas.

In the case of 'rank 3 out of n' data, the saturated log-linear model for the theoretical ranking probabilities $p_{ijk} > 0$ (with $i \neq j, i \neq k, j \neq k$) may be stated in the following way:

$$\ln p_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)}$$

In this model, u is a normalizing constant; the terms u_1, u_2 and u_3 represent the main effects of the various alternatives corresponding to the first, second and third position in the ranking; the terms u_{12}, u_{13} and u_{23} represent the first-order interaction effects between the ranking positions; finally, the term u_{123} represents the second- order interaction between all ranking positions. The first- and second- order interaction terms are only defined for pairs and triples of distinct subscripts. Moreover, in order to obtain an identified log-linear model, some ANOVA-like restrictions have to be imposed on the main and interaction effects. The basic idea behind this log-linear model for 'rank 3 out of n' data is that the ranking frequencies can be inscribed in a $n \times n \times n$ contingency table whose three successive dimensions correspond to the three positions in the incomplete rankings. Since an alternative cannot occupy two or more different positions in the same ranking, only the $n(n-1)(n-2)$ cells that correspond to the possible rankings may contain a non-zero frequency. The remaining $n^3 - n(n-1)(n-2) = n(3n-2)$ cells necessarily contain structural zeros.

The quasi-independence log-linear model

The quasi-independence log-linear model is obtained by assuming that all first- and second-order interaction effects are zero. We then have

$$\ln\ p_{ijk} \quad = \quad u + u_{1(i)} + u_{2(j)} + u_{3(k)}$$

for any triple (i, j, k) of different subscripts. As identifying constraints we impose

$$\sum_{i=1}^{n} u_{1(i)} = \sum_{i=1}^{n} u_{2(i)} = \sum_{i=1}^{n} u_{3(i)} = 0$$

This model may also be written multiplicatively:

$$p_{ijk} \quad = \quad v \cdot v_{1(i)} \cdot v_{2(j)} \cdot v_{3(k)}$$

with, as identifying constraints,

$$\sum_{i=1}^{n} v_{1(i)} = \sum_{i=1}^{n} v_{2(i)} = \sum_{i=1}^{n} v_{3(i)} = 1$$

The concept of quasi-independence is an adaptation of the usual concept of independence to the case of contingency tables with structurally empty cells.

In the general case of a 'rank m out of n' task, we may write in terms of the multiplicative model

$$p_{i_1,\cdots,i_m} \quad = \quad v \cdot \prod_{r=1}^{m} v_{r(i_r)}$$

with

$$\sum_{i=1}^{n} v_{r(i)} \quad = \quad 1$$

for all $r = 1, \ldots, m$. The parameter v is a normalizing factor, which is needed to ensure that the sum of the ranking probabilities over all feasible rankings is equal to one.

It may be of some interest to note here that the Pendergrass-Bradley model for ranking probabilities is a submodel of the quasi-independence model introduced above. Under the Pendergrass-Bradley model, there exist scale values v_i such that we have $v_{r(i)} = v_i^{n-r}$ for all $r = 1, \cdots, m$. The random utility variant of the BTL-model for ranking data, on the other hand, is a submodel of the log-linear model in which the m-th position is independent of the configuration of the first $m - 1$ positions, i.e. of the model in which all interaction terms in which the m-th position is involved are equal to zero.

The latent class model based on the quasi-independence model

The quasi-independence model can easily be incorporated in a latent class model. Assume T latent classes are needed, and let an arbitrary class be denoted by t. The parameters v_r are assumed to be specific for each class; they will be denoted by $v_{r(i)t}$. As identifying constraints we impose for all r and all t:

$$\sum_{i=1}^{n} v_{r(i)t} = 1$$

Then, we may write for the probability of obtaining ranking $(i_1, \cdots, i_m)$ in latent class t:

$$p_{i_1,\cdots,i_m,t} = v_t \prod_{r=1}^{m} v_{r(i_r)t}$$

in which v_t is the normalizing factor for latent class t. If π_t represents the latent proportion of class t, we finally have:

$$p_{i_1,\cdots,i_m} = \sum_{t=1}^{T} p_{i_1,\cdots,i_m,t} \cdot \pi_t$$

The E.M. algorithm for estimating the quasi-independence latent class model

The maximum likelihood estimates of the model parameters can be obtained by means of an E.M-algorithm. We will restrict ourselves to the case of 'rank 3 out of n' data in our discussion of this algorithm.

The iterations of the E.M. algorithm consist of two steps: an E(expectation)-step and a M(aximization)-step.

1. During the **E-step** the observed ranking frequencies f_{ijk} are distributed over the T classes in the following way:

$$f_{ijkt} = f_{ijk} \times p_{t|ijk}$$

   in which the conditional probability $p_{t|ijk}$ is given by

$$p_{t|ijk} = \frac{p_{ijkt} \cdot \pi_t}{\sum_t p_{ijkt} \cdot \pi_t}$$

   This conditional probability is computed on the basis of the provisory values of the model parameters.

2. During the **M-step**, the quasi-independence model is fitted, separately in each class, to the 'completed' set of ranking frequencies f_{ijkt}. This is done by using the Iterative Proportional Fitting Algorithm. Let e_{ijkt} denote the expected frequency corresponding to the

observed frequency f_{ijkt} under the quasi-independence model. These expected frequencies are obtained by means of the following iterative computing algorithm:

Step 1

$$e_{ijkt}^{(s)} = e_{ijkt}^{(s-1)} \times \frac{f_{i++t}}{e_{i++t}^{(s-1)}}$$

Step 2

$$e_{ijkt}^{(s+1)} = e_{ijkt}^{(s)} \times \frac{f_{+j+t}}{e_{+j+t}^{(s)}}$$

Step 3

$$e_{ijkt}^{(s+2)} = e_{ijkt}^{(s+1)} \times \frac{f_{++kt}}{e_{++kt}^{(s+1)}}$$

We use here, and also in what follows, the + subscript to denote summation over the corresponding subscript. So, for instance,

$$f_{i++t} = \sum_{j \neq i} \sum_{k \neq i,j} f_{ijkt}$$

That we have to use the Iterative Proportional Fitting Algorithm in fitting the quasi-independence model is due to the fact that this model does not allow for an analytic solution of the maximum likelihood optimization problem.

During each M-step, the latent proportions are also estimated again:

$$\pi_t = \frac{e_{+++t}}{N}$$

An example

The following table contains the global results of some latent class analyses based on the quasi- independence model. We have used once again the U.S. data.

t	L	df	p
1	920.36	314	0
2	385.70	292	0.0002
3	318.43	270	0.0228
4	269.86	248	0.1625

From a statistical point of view, only the solution with four classes is acceptable; the solutions with a smaller number of classes all result in a statistically unacceptable fit. In order to see in which respects these four classes differ among themselves, we report the first-choice parameters $v_{1(i)t}$ in the following table:

i	*Class* 1	*Class* 2	*Class* 3	*Class* 4
1	.113	.012	.007	.025
2	.131	.175	.023	.038
3	.063	.023	.005	.230
4	.009	.000	.005	.024
5	.404	.620	.068	.158
6	.241	.141	.878	.000
7	.008	.000	.006	.127
8	.030	.029	.009	.399
π_t	.318	.229	.163	.289

The second- and third choice parameters $v_{2(i)t}$ and $v_{3(i)t}$ showed a pattern similar to that of the first-choice parameters. These results indicate that under the quasi-independence model three slightly different 'materialistic' classes seem to exists in the U.S.A. The first three classes are all characterized by a strong preference of some 'materialistic' items, and by a resolute rejection of the 'post-materialistic' political goals. The differences between the three 'materialistic' classes are more difficult to interpret, and seem to be rather item-specific. Seventy-one percent of the American sample is estimated to belong to one of the materialistic classes. The fourth class probably represents the 'post-materialistic' subpopulation, although some of the alternatives do not conform to the pattern that could be expected here: In this class the alternative item 4, which is a very unpopular item in the U.S., scores too low, while the materialistic item 5, which is the most popular item in this sample, scores too high.

4.4 Models that Allow for Association Between Choices within the Classes

A General Model Allowing for Association between Choices

Latent class models based on a quasi-independence model do not always lead to satisfactory results. Often models of this kind only provide a statistically acceptable fit to the data if the number of latent classes is made large enough.

In a search for alternative latent class models, which possibly could explain the data in terms of a smaller number of latent classes, we first considered the log-linear model which includes all first-order, but not the second- and higher-order interaction effects. This first-order interaction model is in some sense the most simple extension of the quasi-independent model. In this section we restrict ourselves to a discussion of ranking data from a 'rank 3 out of n' task.

For the case of 'rank 3 out of n' data, the latent class model with first-order interactions can be written as

$$\ln p_{ijkt} = u_t + u_{1(i)t} + u_{2(j)t} + u_{3(k)t} + u_{12(ij)t} + u_{13(ik)t} + u_{23(jk)t}$$

In this model, which we refer to as the A_0-model, latent classes differ with respect to the main effects as well as with respect to the first-order interaction terms. It is interesting to note that, for $T = 1$, we simply obtain the hierarchical submodel of the saturated log-linear from which all second-order interaction terms are removed. Our limited experiences with this very general A_0-model, however, have been quite negative for $T \geq 2$.

We observed quite often that the final solutions under this model had many of their parameter estimates on the boundary of the parameter space. This was especially the case for the estimates of the first-order interaction terms. Some rather difficult identification problems are probably involved here.

Three submodels with Invariant First-Order Interaction Effects

Since the general A_0-model did not provide an acceptable alternative to the quasi-independence model considered earlier, we have investigated some submodels of it. In particular, we have considered models in which the first-order interaction terms are assumed to be the same in the various latent classes, which may still differ with respect to main effects. In these models the latent classes may differ with respect to the 'popularity' of the items, but the pattern of association between the choices (as described by first-order interaction terms) is assumed to be invariant over the different classes. We first consider the most general model of this kind, the A_1-model before discussing two interesting submodels of it.

Model A_1

For the most general model within this class, we may write for 'rank 3 out of n' data:

$$\ln p_{ijkt} = u_t + u_{1(i)t} + u_{2(j)t} + u_{3(k)t} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)}$$

In the following this model will be referred to as Model A_1. Note that for $T = 1$ this model too is equivalent to the 'no second-order interactions' submodel of the saturated log-linear model, and, hence, to Model A_0 with $T = 1$ as described above.

Next, we consider two submodels of A_1.

Model A_2

As a first interesting submodel of A_1 we will consider the model for which

$$u_{12(ij)} = u_{13(ij)} = u_{23(ij)} = u_{ij}$$

holds for all i, j. Under this model, which will be referred to as model A_2, one may write

$$\ln p_{ijkt} = u_t + u_{1(i)t} + u_{2(j)t} + u_{3(k)t} + u_{ij} + u_{ik} + u_{jk}$$

In this model only one set of invariant first-order interaction terms remains to be estimated.

Model A_3

A second interesting submodel of A_1 is the model in which the First by Second Choice, and the Second by Third Choice first-order interaction terms are included, but not the First by Third Choice interaction terms. Hence, for this model A_3, we may write:

$$\ln p_{ijkt} = u_t + u_{1(i)t} + u_{2(j)t} + u_{3(k)t} + u_{12(ij)} + u_{23(jk)}$$

In this model only interaction terms for pairs of consecutive positions in the ranking are included. Note that models A_2 and A_3 are both submodels of model A_1, but are themselves not hierarchically related to each other.

Estimating the Parameters by Means of an E.M. Algorithm. Let f_{ijk} be the observed frequency of ranking (i, j, k) and assume that T latent classes are needed for an analysis based either on model A_1, model A_2, or model A_3. Let N denote the sample size. The maximum likelihood estimates of the parameters of the three models can be obtained by means of an E.M. algorithm.

Each iteration of this algorthim consists of two steps:

- An **Expectation step** during which the frequencies f_{ijkt} with which ranking (ijk) occurs in latent class t is estimated again:

$$f_{ijkt} = f_{ijk} \times \frac{p_{ijkt} \cdot \pi_t}{p_{ijk}}$$

with

$$p_{ijk} = \sum_t p_{ijkt} \cdot \pi_t$$

The probability p_{ijkt} of observing ranking (ijk) in latent class t is computed on the basis of the provisory values of the parameter estimates. The way in which these probabilities are computed depends on the model under consideration.

- A **Maximization Step** during which the maximum likelihood estimates of the model are determined again on the basis of the completed set of frequencies f_{ijkt}. The expression for the latent proportions π_t is extremely simple:

$$\pi_t = \frac{f_{+++t}}{N}$$

The estimation of the parameters of the log-linear model is more involved, since one has to rely on a subordinate iterative process, such as the Iterative Proportional Fitting Algorithm. More information on these estimation procedures are given in the next paragraph.

The Iterative Proportional Fitting Algorithm for Models A_1, A_2 and A_3 with Complete Data. We assume that the frequency f_{ijkt} with which ranking (ijk) occurs in class t is observed. The corresponding expected frequency will be denoted by e_{ijkt}.

For model A_1 the iterations of Iterative Proportional Fitting Algorithm consist of the following 6 steps:

1.

$$e^{(s+1)}_{ijkt} = e^{(s)}_{ijkt} \times \frac{f_{i++t}}{e^{(s)}_{i++t}}$$

2.

$$e^{(s+2)}_{ijkt} = e^{(s+1)}_{ijkt} \times \frac{f_{+j+t}}{e^{(s+1)}_{+j+t}}$$

3.

$$e^{(s+3)}_{ijkt} = e^{(s+2)}_{ijkt} \times \frac{f_{++kt}}{e^{(s+2)}_{++kt}}$$

4.

$$e^{(s+4)}_{ijkt} = e^{(s+3)}_{ijkt} \times \frac{f_{ij++}}{e^{(s+3)}_{ij++}}$$

5.

$$e^{(s+5)}_{ijkt} = e^{(s+4)}_{ijkt} \times \frac{f_{+jk+}}{e^{(s+4)}_{+jk+}}$$

6.

$$e_{ijkt}^{(s+6)} = e_{ijkt}^{(s+5)} \times \frac{f_{i+k+}}{e_{i+k+}^{(s+5)}}$$

For Model A_2, the iterations of the Iterative Proportional Fitting Algorithm consist of 4 steps, the first three being identical with the corresponding steps of the algorithm for the A_1 model. The fourth step itself consists of $n(n-1)$ substeps, each one corresponding to a pair (i, j) of distinct subscripts. During the substep that corresponds to the pair (i, j), the following computations take place for all $k = 1, \cdots, n$ (with $k \neq i$ and $k \neq j$) and for all t:

$$\begin{aligned} e_{ijkt}^{(new)} &= e_{ijkt}^{(old)} \times \frac{S_{ij}}{U_{ij}^{(old)}} \\ e_{ikjt}^{(new)} &= e_{ikjt}^{(old)} \times \frac{S_{ij}}{U_{ij}^{(old)}} \\ e_{kijt}^{(new)} &= e_{kijt}^{(old)} \times \frac{S_{ij}}{U_{ij}^{(old)}} \end{aligned}$$

with

$$S_{ij} = f_{ij++} + f_{i+j+} + f_{+ij+}$$

and

$$U_{ij}^{old} = e_{ij++}^{(old)} + e_{i+j+}^{(old)} + e_{+ij+}^{(old)}$$

For Model A_3, the iterations of the Iterative Proportional Fitting Algorithm consist of five steps, which are identical to the first five steps of the algorithm for fitting Model A_1.

In order to guarantee that the Iterative Proportional Fitting Algorithms converge to the maximum of the likelihood function, the starting values of the expected frequencies should satisfy the model under consideration. The easiest way out of this problem is to set all expected frequencies e_{ijkt} initially equal to 1.

After convergence of the Iterative Proportional Fitting Algorithm, the model parameters, such as $u_{1(i)t}, u_{2(i)t}, u_{3(i)t}, u_{12(ij)}, \cdots$, can be determined by solving appropriate systems of linear equations in these unknowns. This system of linear equations expresses the model parameters as functions of the natural logarithms of the expected frequencies e_{ijkt}.

Testing model fit. When the E.M. algorithm has converged, the hypothesis that the model under consideration applies to the data may be tested against the general multinomial hypothesis by means of a log likelihood ratio test. Let $\hat{p}_{ijk}$ be the estimate of the theoretical ranking probability under the particular model under consideration, and let $\hat{f}_{ijk} = N \cdot \hat{p}_{ijk}$ denote the corresponding expected frequency. Then, the log likelihood ratio statistic L is defined as:

$$L = 2 \times \sum_{i,j,k} f_{ijk} \ln \left(\frac{f_{ijk}}{\hat{f}_{ijk}} \right)$$

where the summation runs over all triples of distinct subscripts.

If the model under consideration is true, then the log likelihood statistic is asymptotically distributed as a chi square variate with degrees of freedom equal to the difference between the number of independent parameters under both models.

In the context of latent class analysis, model tests of this kind can be used to the test the hypothesis that the latent class model with a specified number T of classes is true against the general multinomial hypothesis. Let L_T denote the value of log likelihood statistic obtained by a latent class analysis with T classes. For Model A_1 the observed value of the statistic L_T should, for $n \geq 5$, be located under a chi square distribution with $(n^3 - 6n^2 + 11n - 3) - (3n - 2)T$ degrees of freedom; for Model A_2, the number of the degrees of freedom is given by $(n^3 - 4n^2 + 5n - 1) - (3n - 2)T$ if $n \geq 5$; for Model A_3, the number of degrees is $n^3 - 5n^2 + 8n - (3n - 2)T$.

Some Results

The Results of the A_2 Analyses on the U.S. Data. The U.S. ranking data were analyzed on the basis of model A_2 with $T = 1$ and $T = 2$. The global results are shown in the next table.

T	L	$d.f.$	p
1	325.523	273	.016
2	248.833	251	.527

Hence, we see that the solution with two latent classes provides an acceptable fit to the U.S. data. The next table contains the estimates of the main effects parameters in both classes.

i	Class 1 $u_{1(i)1}$	$u_{2(i)1}$	$u_{3(i)1}$	Class 2 $u_{1(i)2}$	$u_{2(i)2}$	$u_{3(i)2}$
1	.11	.29	.05	−1.25	−2.01	−1.33
2	.90	.69	−.09	−.60	−.42	−.10
3	−1.29	−.91	−.12	1.48	1.51	1.14
4	−1.95	−1.67	−1.43	−1.29	−.18	.30
5	2.23	1.49	.68	.59	.58	.33
6	2.37	2.57	2.20	−.09	−.73	−.78
7	−1.59	−1.15	−.48	−.21	.01	−.68
8	−.79	−1.31	−.82	1.38	1.24	1.13
$\bar{u}_{mat}$	1.40	1.26	.71	−.34	−.65	−.47
$\bar{u}_{pmat}$	−1.40	−1.26	−.71	.34	.65	.47

The estimates of the latent proportions were $\hat{\pi}_1 = .613$ and $\hat{\pi}_2 = .387$. The interpretation of these results is rather straightforward:

- The first latent class is a relatively pure 'materialistic' class in which the four materialistic alternatives are rated higher than the four post-materialistic ones. The clear opposition between the two groups of alternatives occurs at all three ranking positions, but it diminishes slightly when going from the first to the third position.

- When looking at the average scale values of the materialistic and post-materialistic alternatives in the second class, it should be clear that this class cannot be considered as a pure 'post-materialistic' class. A few rather striking exceptions make such an interpretation implausible: In this class, the post-materialistic items 4 and 7 score much too low, while the materialistic alternative 5 scores too high. It is probably safer to characterize this class as the class of persons who value the humane and spiritual aspects of life.

A Further Analysis of the First-Order Interaction Terms. Next, we turn to the interpretation of the interaction terms. Instead of giving the complete 8×8 matrix with estimated first-order interaction terms, we will report on the results of a bilinear decomposition analysis of these terms. Assume the first-order interaction terms u_{ij} are inscribed in a $n \times n$ matrix U. Since the terms u_{ij} are undefined for the case $i = j$, the main diagonal of this matrix is structurally empty. We say that the matrix U allows for a 'Bilinear Decomposition of Rank s' if there exist two $n \times s$ matrices X, the left factor matrix, and Y, the right factor matrix, such that

$$u_{ij} = \sum_{q=1}^{s} x_{iq} y_{jq}$$

holds for all $i, j = 1, \cdots, n$ with $j \neq i$. In practice, we are interested in the bilinear decomposition of the lowest rank which still provides an acceptable fit to the incomplete matrix. To this end, we determine, for successive values of s, the decomposition of U which minimizes the following least squares loss function:

$$\phi = \sum_{i,j\neq i} \left(u_{ij} - \sum_{q=1}^{s} x_{iq} y_{jq} \right)^2$$

For more information on the bilinear decomposition model and on the technical details of the estimation procedure, we refer to [7].

In the present example, the rank 1 decomposition left 50.3 % percent of the variance of the interaction terms unexplained. For the rank 2 decomposition, this figure decreased to 23.2 %. The next table gives the result of the latter decomposition.

i	x_{i1}	x_{i2}	y_{i1}	y_{i2}
1	−.60	−.32	.23	−.56
2	−.34	−.43	−.50	−.53
3	.67	−.66	−.47	.07
4	.39	.18	−.56	.67
5	−.23	−.19	.07	−.42
6	−.65	.46	.65	−.03
7	.27	.88	.87	.57
8	.73	.07	−.37	.48

From the information in these coordinate matrices, one may conclude that, to a large extent, the pattern of the first-order interaction terms is dominated or determined by the contrast between the two types of alternatives. An interesting feature of this bilinear decomposition is that the contrast between materialistic and post-materialistic alternatives shows itself most distinctively in the first component of the left factor matrix X, and in the second component of the right factor matrix Y. It is not clear why different components from the left and right factor matrix should be involved in this way.

A Comparison with the A_3 analyses. The U.S. data were also analyzed by means of the A_3-model. The next table gives some global results.

T	L	df	df
1	375.673	230	0
2	242.987	208	.0485
3	191.608	186	.3736

Since models A_2 and A_3 are not related to each other in a hierarchical way, it is difficult to compare the relative fits of both models to the same data. However, it is probably safe to conclude that the two-class solution of the A_2 analysis represents the data better than the two-class solution of the A_3 analysis. This is remarkable since fewer parameters are estimated under the A_2-model than under the A_3. This result seems to indicate that all three kinds of first-order interaction terms (First by Second Choice, Second by Third Choice, and First by Third Choice) are needed in a comprehensive latent class model of this type. Removing one set of these interaction terms has more detrimental effects than setting corresponding terms in the three sets equal to each other.

Acknowledgements:

The second author's contribution was made possible by grant #500-278-009 from the Social Sciences Section of the Dutch National Science Foundation (NWO).

4.5 References

[1] Barnes, S.H. and M. Kaase. *Political Action. Mass Participations in Five Western Countries.* London: Sage. 1979.

[2] Block, H.D. and J. Marschak. Random Orderings and Stochastic Theories of Response. In I. Olkin, S.G. Ghurye, W. Hoeffding, W. Madow and H. Mann (Eds.), *Contributions to Probability and Statistics* pp. 97-132, 1960. Stanford: Stanford University Press.

[3] Bradley, R.A. Another Interpretation of a Model for Paired Comparisons. *Psychometrika*, **30**:315-318, 1963.

[4] Bradley, R.A. and M.E. Terry. Rank Analysis of Incomplete Block Designs I. *Biometrika*, **39**:324-345, 1952.

[5] Croon, M.A. Latent Class Models for the Analysis of Rankings.In: G. de Soete, H. Feger and K.C. Klauer (Eds.), *New Developments in Psychological Choice Modelling* pp. 99-121, 1989a. Amsterdam: Elsevier Science Publishers.

[6] Croon, M.A. The Analysis of Partial Rankings by Log-Linear and Latent Class Models. In: R. Coppi and S. Bolasco (Eds.), *Multiway Data Analysis* pp. 497-506, 1989b. Amsterdam: Elsevier Science Publishers.

[7] Croon, M.A. Log-Linear Analysis of Partial Preference Rankings. *Methodika* 1991. (still to appear).

[8] Inglehart, R. *The Silent Revolution.* Princeton: Princeton University Press. 1977.

[9] Luce, R.D. *Individual Choice Behavior.* New York: Wiley. 1959.

[10] Luce, R.D. and P. Suppes. Preference, Utility and Subjective Probability. In R.D. Luce, R.R. Bush and E. Galanter (Eds.), *Handbook of Mathematical Psychology*, **3**:97-132, 1965. New York: Wiley.

[11] Pendergrass, R.N. and R.A. Bradley. Ranking in Triple Comparisons. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. Madow and H. Mann (Eds.), *Contributions to Probability and Statistics* pp. 133-151, 1960. Stanford: Stanford University Press.

[12] Plackett, R.L. Random Permutations. *Journal of the Royal Statistical Society, Series B*, **30**:517-534, 1968.

[13] Plackett, R.L. The Analysis of Permutations. *Applied Statistics*, **24**:193-202, 1975.

[14] Yellott, J.I. The relationship between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment and the Double Exponential Distribution. *Journal of Mathematical Psychology*, **15**:109-144, 1977.

[15] Yellott, J.I. Generalized Thurstone Models for Ranking: Equivalence and Reversibility. *Journal of Mathematical Psychology*, **22**:48-69, 1980.

[16] Zermelo, E. Die Berechnung der Turnierergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, **29**:436-460, 1929.