The use of restricted latent class models for defining and testing nonparametric and parametric IRT models

Vermunt, J.K.

Published in:
Applied Psychological Measurement

Publication date:
2001

Document Version
Peer reviewed version

Citation for published version (APA):
Vermunt, J. K. (2001). The use of restricted latent class models for defining and testing nonparametric and parametric IRT models. *Applied Psychological Measurement*, *25*(3), 283-294.

The use of restricted latent class models for defining and testing nonparametric and parametric IRT models

Jeroen K. Vermunt

Department of Methodology and Statistics

Faculty of Social and Behavioral Sciences

Tilburg University

Abstract

This paper presents a general class of ordinal logit models, which involve specifying (in)equality constraints on sums of conditional response probabilities. By using these constraints in latent class analysis, one obtains models that are similar to parametric and/or non-parametric item response models. An important implication of this similarity for the field of non-parametric IRT modeling is that latent class methodology can be used to estimate these models by means of maximum likelihood, which make it possible to test their assumptions by means of likelihood-ratio statistics.

Index terms: order-restricted inference, restricted latent class analysis, polytomous item response theory, stochastic ordering, inequality constraints, parametric bootstrapping.

1 Introduction

Two kinds of similarities between latent class models and IRT models have been shown in the psychometric literature. On the one hand, Croon (1990, 1991) and Hoijtink and Molenaar (1997) demonstrated that order-restricted latent class models can be used to estimate non-parametric IRT models with maximum likelihood and Bayesian methods, respectively. Their approaches consist of specifying simple inequality restrictions on the cumulative conditional response probabilities (or item step response functions). On the other hand, Heinen (1996) demonstrated the similarity between log-linear latent class models and parametric IRT models. More precisely, he showed that discretized variants of the most important parametric IRT models can be obtained by imposing certain constraints on the log-linear parameters of latent class models.

This paper integrates and extends the above work using a general class of log-linear (in)equality constraints on sums of conditional response probabilities. The presented approach is based on work in the field of generalized log-linear modeling (Lang and Agresti, 1994; Bergsma, 1997) and order-restricted inference with categorical variables (see, for instance, Robertson et al., 1988; Dardanoni and Forcina, 1998; and Vermunt, 1999). It is shown that restrictions of these forms can be used not only to specify non-parametric IRT models such as the monotone homogeneity model for polytomous items (Molenaar, 1997), but also to define non-parametric variants of most parametric IRT models for polytomous items. In addition, hybrid IRT models combining parametric with non-parametric features can be obtained.

The next section introduces the equality and inequality constraints that can be used to specify models for ordinal variables using a simple example of a two-way table. Then, it is shown how to use these constraints in the context of latent class analysis, yielding discretized variants of parametric and non-parametric IRT models. Technical details on model estimation and testing are given in the Appendix.

2 Models for an ordinal independent and an ordinal dependent variable

This section introduces the equality and inequality constraints that will be used in the next section to specify restrictions on the relationship between a latent trait and a set of items. The logit models for an ordinal independent and an ordinal dependent variable are illustrated by means of an empirical example taken from the ISAC-A questionnaire on crying (Becht, Poortinga & Vingerhoets, in press). Table 1 contains the two-way cross-tabulation of the two ordinal variables of interest: the scores on the "Crying from Distress" scale, which were collapsed into five nearly equal size categories, and the questionnaire item "Feeling Relieved after Crying", which is not from the same scale. These variables will be denoted by X and Y, their number of levels by I (=5) and J (=3), and their category indices by i and j. In the following, we assume that X (Crying from Distress) serves as independent and Y (Feeling Relieved after Crying) as dependent variable, which means that we are interested in the conditional distribution of Y given X. The substantive research question of interest is whether individuals who tend to cry more often experience more benefits from crying; that is, whether there is a positive relationship between "Crying from Distress" and "Feeling Relieved after Crying".

[INSERT TABLE 1 ABOUT HERE]

The standard way of modeling relationships between such ordinal categorical variables is by means of a logit model that imposes equality constraints on certain odds-ratios. Four types of odds can be used for this purpose (Agresti, 1990: section 9.3; Mellenbergh, 1995): cumulative odds ($\Omega_{i,j}^{cum}$), adjacent category (or local) odds ($\Omega_{i,j}^{adj}$), or one of two types of continuation odds ($\Omega_{i,j}^{conI}$ and $\Omega_{i,j}^{conII}$). These are defined as

$$\begin{aligned}
\Omega_{i,j}^{cum} &= P(Y \leq j-1|X=i)/P(Y \geq j|X=i)\,, \\
\Omega_{i,j}^{adj} &= P(Y = j-1|X=i)/P(Y = j|X=i)\,, \\
\Omega_{i,j}^{conI} &= P(Y = j-1|X=i)/P(Y \geq j|X=i)\,,
\end{aligned}$$

$$\Omega^{conII}_{i,j} = P(Y \leq j-1|X=i)/P(Y=j|X=i)\,,$$

respectively, with $2 \leq j \leq J$ and $1 \leq i \leq I$. Below, the symbol $\Omega_{i,j}$ will be used as a generic symbol referring to any of these odds.

Modeling a certain type of odds corresponds to modeling a certain type of item step response function (ISRF). Note that an important difference with the IRT models to be discussed in the next section is that we do not have a latent trait, but condition on an observable scale score X. As is explained in more detail by Van der Ark (2001), the ISRFs corresponding to the four types of odds are

$$\begin{aligned}
ISRF^{cum}_{i,j} &= P(Y \geq j|X=i), \\
ISRF^{adj}_{i,j} &= \frac{P(Y=j|X=i)}{P(Y=j|X=i)+P(Y=j-1|X=i)}, \\
ISRF^{conI}_{i,j} &= \frac{P(Y \geq j|X=i)}{P(Y \geq j-1|X=i)}, \\
ISRF^{conII}_{i,j} &= \frac{P(Y=j|X=i)}{P(Y \leq j|X=i)}.
\end{aligned}$$

The cumulative odds or $ISRF^{cum}_{i,j}$ is used in graded response models, adjacent category odds or $ISRF^{adj}_{i,j}$ in partial credit models, and the first type of continuation odds or $ISRF^{conI}_{i,j}$ in sequential models (Mellenbergh, 1995; Van der Ark, 2001). The second type of continuation odds described above differs from the first one with respect to order in which respondents are assumed to evaluate the categories of the response variable. By using $\Omega^{conI}_{i,j}$ one assumes that respondents evaluate the response alternatives from low to high, while by using $\Omega^{conII}_{i,j}$ one assumes the reversed process. So, actually there are two different types of sequential models.

In practical research situations, one will make a choice between these four types of odds or ISRFs; that is, one will specify a model for the type of odds or ISRF that fits best to the assumed process underlying the individual responses. Here, we will use all four types in order to illustrate generality of the presented approach.

A logit model that takes into account that both the dependent and independent variable are ordinal is (Agresti, 1990)

$$\log \Omega_{i,j} = \alpha_j - \beta\, x_i\,.$$

Here, x_i denotes the fixed score assigned to category i of X, and α_j and β are the intercept and the slope of the logit model. In most cases, x_i will be equal-interval scores (for example, 1, 2, 3, 4, etc.), but it is also possible to use other scoring schemes for the X variable. The above logit model is equivalent to the following model for the ISRFs:

$$ISRF_{i,j} = \frac{\exp(\alpha_j + \beta\, x_i)}{1 + \exp(\alpha_j + \beta\, x_i)}.$$

As can been seen, the ISRFs are assumed to have equal slopes. This implies, for instance, that with adjacent category ISRFs, one obtains a model that is similar to a partial credit model. A difference is, of course, that X is an observed variable rather than a latent trait. When we use equal-interval x_i, the model described in equation (**??**) implies that the log odds-ratios between adjacent levels of X are assumed to be constant, i.e.,

$$\log \Omega_{i,j}/\Omega_{i+1,j} = \log \Omega_{i,j} - \log \Omega_{i+1,j} = \beta$$

for all i and j. The fact that these differences between log odds do not depend on the values of X and Y can also be expressed by the following two sets of equality constraints:

$$(\log \Omega_{i,j} - \log \Omega_{i+1,j}) - (\log \Omega_{i,j+1} - \log \Omega_{i+1,j+1}) = 0\,, \tag{1}$$

$$(\log \Omega_{i,j} - \log \Omega_{i+1,j}) - (\log \Omega_{i+1,j} - \log \Omega_{i+2,j}) = 0\,. \tag{2}$$

What is important is to note is that the restrictions implied by standard ordinal logit models can also be defined in terms of equalities on the log odds-ratios; that is, by eliminating the model parameters α_j and β. This feature is used in the constrained-optimization procedure described in the Appendix.

[INSERT TABLE 2 ABOUT HERE]

Table 2 reports the observed log cumulative odds-ratios for the data reported in Table 1. Note that a log odds-ratio larger than zero is in agreement with the postulated positive relationship between "Crying from Distress" and "Feeling Relieved after Crying". As can be seen, the data contain four violations of an ordinal relationship. The question to be answered is whether this could be the result of sampling fluctuation.

[INSERT TABLE 3 ABOUT HERE]

Table 3 reports the test results for the estimated models. The test statistic for the independence model shows that there is a significant association between the two variables. As can be seen, none of the four types of logit models that impose the constraints described in equations (1) and (2) fits the data. This shows that the parametric assumptions of the standard ordinal logit models are too restrictive for this data set.

An alternative is to replace the equalities implied by the above logit model by inequalities; that is, to switch from a parametric to a non-parametric approach. This yields the following less restrictive definition of a positive relationship in terms log odds-ratios:

$$\log \Omega_{i,j} - \log \Omega_{i+1,j} \geq 0\,. \tag{3}$$

As can be seen, we are assuming that all log odds-ratios are at least 0. Such a set of constraints is often referred to as simple stochastic ordering, likelihood ratio ordering, or uniform stochastic ordering for cumulative, adjacent category, and continuation odds, respectively (Dardanoni and Forcina, 1998). The non-negativity constraint on the log odds-ratios can also be formulated in terms of constraints on the ISRFs (see also Van der Ark 2001); that is,

$$ISRF_{i,j} \leq ISRF_{i+1,j}\,.$$

The presented constraints are, actually, special cases of a more general class of equality and inequality constraints described in the Appendix. Hybrid ordinal logit models could be obtained by combining inequality with equality restrictions. For example, combining the restrictions in

(3) with the equality restrictions in (2) yields a model in with the odds are monotonically decreasing, where the decrease is constant between adjacent values of X. The Appendix gives details on constrained maximum likelihood estimation.

The Appendix also describes the parametric bootstrapping procedure that was used to estimate the p values corresponding to the likelihood-ratio statistic. We have to use an alternative procedure for assessing goodness-of-fit because in models with inequality constraints we can not rely on standard asymptotic results.

Models imposing the inequality constraints defined in (3) were estimated for the crying data set introduced above. The lower part of Table 3 gives the tests results for the four order-restricted models. As can be seen, the order-restricted (non-parametric) models fit much better than the standard (parametric) logit models. This means that, contrary to what would be concluded on the basis of the standard ordinal logit models, there is no evidence against a monotonic relationship between "Crying from Distress" and "Feeling Relieved after Crying". Note again that the choice among the four types of non-parametric models should not only depend on the fit of the models, but also on the plausibility of the assumed process generating the responses.

As can be seen from the estimated cumulative log odd-ratios reported in Table 2, the consequence of imposing inequality constraints is that certain log odd-ratios are equated to zero. Similar tables could be presented for the other three types of log odds-ratios.

3 Latent class models for ordinal items

Suppose we have a latent class model (LCM) with a single latent trait X and K items denoted by Y_k, with $1 \leq k \leq K$. The number of latent classes, or the number of levels of the (discretized) latent trait, is denoted by I, the number of levels of item Y_k by J_k, and i and j_k denote a particular level of X and Y_k.

The basic assumption of the latent class model is that the items are independent of one another within latent classes, mostly referred to as the local independence assumption (see, for instance, Goodman, 1974; or Bartholomew and Knott, 1999). A LCM with a single latent variable can be defined as

$$P(Y_1 = j_1, Y_2 = j_2, ..., Y_K = j_K) = \sum_{i=1}^{I} P(X = i) \prod_{k=1}^{K} P(Y_k = j_k | X = i).$$

Here, the $P(X = i)$ refer to the unspecified distribution of the latent trait, and the $P(Y_k = j_k | X = i)$ are the item response probabilities. In the standard latent class model, no restrictions are imposed on these probabilities.

In order to obtain a monotone relationship between X and an ordinal item Y_k, we can use the logit constraints described in the previous section; for instance, equality restrictions like the ones described in the equations (1) and (2) can be imposed on the conditional response probabilities $P(Y_k = j_k | X = i)$. Denoting an odds for item k by $\Omega^k_{i,j}$, these equality constraints are now

$$\left(\log \Omega^k_{i,j} - \log \Omega^k_{i+1,j}\right) - \left(\log \Omega^k_{i,j+1} - \log \Omega^k_{i+1,j+1}\right) \;=\; 0\,, \tag{4}$$

$$\left(\log \Omega^k_{i,j} - \log \Omega^k_{i+1,j}\right) - \left(\log \Omega^k_{i+1,j} - \log \Omega^k_{i+2,j}\right) \;=\; 0\,. \tag{5}$$

The first set of constraints renders the odds-ratios for item k category independent, and the second renders them class independent.

Heinen (1996:120-133) showed that a LCM for polytomous items with restrictions of the form (5) on the adjacent category log odds-ratios yields a model that is similar to the nominal response model (Bock, 1972). He also demonstrated that when constraints (4) and (5) are imposed at the same time, one obtains a discretized variant of the partial credit model (PCM; Masters, 1982) with item-specific slopes, usually referred to as the generalized partial credit model (Muraki, 1992). These constraints imply that we assume the following parametric function for the ISRF of level j of item k:

$$ISRF^k_{i,j} = \frac{\exp(\alpha^k_j + \beta^k x_i)}{1 + \exp(\alpha^k_j + \beta^k x_i)}. \tag{6}$$

Constraints of these types can, however, not only be imposed on the adjacent category odds as Heinen suggested, but also on the cumulative and continuation odds, which yields discretized variants of the graded response model (GRM; Samejima, 1969) and the sequential response model (SRM; Tutz, 1990; also see, Mellenbergh, 1995). The only difference between a parametric IRT model and the corresponding restricted LCM model is that the distribution of the trait is described by a small number of points (classes) with unknown weights (sizes) rather than by a known, in most cases normal, distributional from.

If one realizes how parametric IRT models are estimated in practice, the similarity to latent class models becomes even greater. In one of the standard estimation methods for parametric IRT models, marginal maximum likelihood, one uses normal quadrature to solve the integrals appearing in the likelihood function, which means that one implicitly works with a discretized latent variable. Rasch models are usually estimated by conditional maximum likelihood, which involves conditioning on the (discrete) total score. It has been shown that the equivalent estimates for the Rasch model can be obtained by a restricted LCM (Lindsay, Clogg, and Grego, 1991).

Rather than obtaining monotonicity with equality constraints, this can also be accomplished by imposing inequality constraints of the form (3) on the various types of log odds-ratios. If we translate these constraints to the LCM context, we obtain

$$\log \Omega^k_{i,j} - \log \Omega^k_{i+1,j} \geq 0, \tag{7}$$

which indicates that the $\Omega^k_{i,j}$ decreases or remains equal as i increases. Croon (1990, 1991) and Hoijtink and Molenaar (1997) proposed using such constraints with cumulative odds, which yields a discretized variant of the non-parametric GRM (Hemker et al., 1997), also known as the polytomous monotone homogeneity model (Molenaar, 1997). The more general approach presented here, makes it also possible to impose non-negative log odds-ratios restrictions on the adjacent category and continuation odds, yielding non-parametric variants of the PCM and SRM.

Besides the within-item restrictions on the conditional response probabilities described so far, it may also be relevant to impose equality or inequality constraints between items. The most relevant between-item equality restriction is

$$\left(\log \Omega^{k}_{i,j} - \log \Omega^{k}_{i+1,j}\right) - \left(\log \Omega^{\ell}_{i,j} - \log \Omega^{\ell}_{i+1,j}\right) = 0\,,$$

which amounts to equating the discrimination parameters of items k and ℓ; that is, $\beta^{\,k} = \beta^{\,\ell}$. The most interesting inequality constraint across items is

$$\log \Omega^{k}_{i,j} - \log \Omega^{\ell}_{i,j} \geq 0. \tag{8}$$

This inequality specifies that item k is easier than item ℓ for latent class i. Imposing these constraints on all item pairs (k, ℓ) in combination with the inequality constraints described in equation (7) on the cumulative odds yields a more restricted variant of the polytomous monotone homogeneity model; that is, the model of strong double monotonicity (Sijtsma and Hemker, 1998), in which the restrictions concern the same item category (j) for all K items.

So far, we assumed that we had either a parametric or a non-parametric model; that is, we imposed either equality or inequality restrictions on odds-ratios. However, it is also possible to combine parametric with non-parametric features. For example, combining inequality constraints (7) with equality constraints (4) yields a model with category-independent odds-ratios in combination with ordered latent classes. When used in combination with adjacent category odds, such a hybrid model has the form of an ordered-restricted row-association structure (Agresti et. al., 1986; Vermunt, 1999).

Technical details on maximum likelihood estimation of LCMs with equality and/or inequality restrictions are given in the Appendix. With respect to model estimation, it is important to note that in terms of requested computer time it is no problem to estimate models with a large number (say 50 or 100) of items: estimation will never take more than a few minutes, which is similar to parametric IRT models. Another practical estimation issue is the multimodality of the log-likelihood function: to decrease the chance of reporting a suboptimal solution, the same model has to be run with several sets of random starting values.

The Appendix discusses computation of p values by parametric bootstrapping. Application of this method in the context order-restricted latent class models was first proposed by Ritov and Gilula (1993). More precisely, they applied the method in an order-restricted correspondence analysis model, which is a two-class LCM with non-negative adjacent category log odds-ratios. A simulation study by the same authors showed that the bootstrapping procedure yields reliable estimates of the p values associated with the G^2 statistic. However, because of its computational intensivity, application of the bootstrap procedure may be problematic in larger problems, say with more than 10 items. One has to realize that, depending on the requested precision of the estimated p value, the same model has to be estimated 100 to 1000 times.

4 Example

To illustrate the various types of LCMs for ordinal items presented above, we use again data from the ISAC-A questionnaire on crying (Becht, Poortinga & Vingerhoets, in press). We selected four items related to "Crying from Distress". Their exact wording is: 1) I cry when I feel frightened (item a29), 2) I cry when I am in despair (item a39), 3) I cry when I feel rejected by others (item a40), and 4) I cry when I feel that I am in a blind-alley situation (item a41). The original seven point scales (1=never,7=always) were collapsed into 3 levels: 1-2, 3-5, and 6-7.

[INSERT TABLE 4 ABOUT HERE]

Table 4 reports the test results for the estimated unrestricted and restricted LCMs for the example data set. As can be seen from the goodness-of-fit tests for the unrestricted LCMs, the four-class models fit the data best. Also the difference between the G^2 values of the three- and four-class models ($\Delta G^2 = 33.4$) is clearly significant with an estimated bootstrap p value of

0.00. The four-class model contains only a few order violations, which is an indication that it is capturing a single dimension in the items. Therefore, we will retain the four-class model as a basis for testing the validity of the restrictions corresponding to parametric and non-parametric IRT models.

The second part of Table 4 gives the test results for the four discretized parametric IRT models obtained by specifying four-class models with equality constraints of the forms (4) and (5) on the four types of odds, which amounts to restricting the ISRFs to have the same slopes (see equation (6)). The reported p-values show that none of these models fits the data very well, which indicates that the constraints implied by parametric IRT models are too restrictive for this data set.

The third part of Table 4 reports the goodness-of-fit measures for the LCMs that impose inequality constraints of the form (7). Each of these models fits the data at a 5% significant level. As can be seen, the G^2 value of cumulative model (GRM) is the lowest one, the one of the adjacent category model (PCM) the highest one, and the G^2 values of the two continuation odds models (SRMs) are in between these two values. Actually, this is what can be expected since there is a hierarchy between the various types of inequality constraints: the inequality constraints on the adjacent category odds imply the inequality constraints on the two types of continuation odds and the cumulative odds, and the inequality constraints on one of the two types of continuation odds imply the inequality constraints on the cumulative odds (see, for instance, Hemker et al., 1997; and Van der Ark, 2001).

As can be seen from the last part of Table 4, also the double monotonicity models based on constraints (7) and (8) fit the data well (all p values are larger than 0.05). Note that in the cumulative model no additional constraints are activated by the data compared to the model of monotone homogeneity. This can be seen from the fact that the reported df is equal for the two non-parametric graded response models. On the other hand, when applying the strong double monotonicity restrictions on the adjacent category and continuation odds, a few additional constraints are activated by the data.

5 Final remarks

This paper presented a general class of models for restricting the conditional response probabilities in LCMs. The log-linear equality and inequality restrictions can be used to specify a broad class of ordered LCMs. In addition, it was shown that these LCMs can be seen as discretized versions of parametric and non-parametric IRT models. The most important implication of this similarity for the field of non-parametric IRT modeling is that latent class methodology can be used to estimate these models by means of maximum likelihood. This makes it possible to test their assumptions using standard likelihood-ratio tests.

Several interesting extensions are straightforward within the presented latent class framework. The most important ones are models with several latent traits, models with covariates, and models with local dependencies. Each of these extensions can be implemented using the framework introduced in this paper.

Appendix

This appendix presents the more general form of the equality and inequality constraints described in this paper and explains how to obtain the restricted maximum likelihood estimates of the model probabilities. In addition, attention is paid to computation of p values by means of parametric bootstrapping. The procedures described below are implemented in an experimental version of the ℓ_{EM} program (Vermunt, 1993, 1997).

Equality constraints

ML estimation of standard ordinal logit models is straightforward. Nevertheless, I would like to propose a less straightforward manner for parameter estimation that will prove useful when working with inequality constraints. The method is based on the work of Lang and

Agresti (1994) and Bergsma (1997) in the field of generalized log-linear modeling, also known as marginal modeling.

The basic idea is to define the estimation of the probabilities $P(Y = j|X = i)$ or $\pi_{j|i}$ as a restricted optimization problem. The rth restriction on the $\pi_{j|i}$'s of the form

$$\sum_t c_{rt} \log \sum_{ij} a_{ijt} \pi_{j|i} = 0 \,.$$

The a_{ijt}'s, which take on values one or zero, can be used to define the appropriate sums of probabilities on which the odds are based. The index t is used to denote the tth sum. The c_{rt}'s are used to define the relevant linear restrictions on the log odds-ratios; that is, restrictions like the ones defined in equations (1) and (2).

Assuming a multinomial sampling scheme, ML estimation now involves finding the saddle point of the following Lagrange equation:

$$\begin{aligned} \mathcal{L} &= \sum_{ij} n_{ij} \log \pi_{j|i} + \sum_i \gamma_i \left(\sum_j \pi_{j|i} - 1 \right) \\ &\quad + \sum_r \lambda_r \left(\sum_t c_{rt} \log \sum_{ij} a_{ijt} \pi_{j|i} \right) \,, \end{aligned}$$

where γ_i and λ_r denote Lagrange multipliers, and n_{ij} is an observed cell count.

Lang and Agresti (1994) and Bergsma (1997) provided two slightly different versions of the Fisher-scoring algorithm to solve this problem. Vermunt (1999) proposed a simple unidimensional Newton method that can be used for a more limited class of restrictions. For more general information on algorithms for constrained optimization see, for instance, Gill and Murray (1974).

Inequality constraints

Estimation with inequality constraints is very similar to the estimation with equality constraints. The sth inequality constraint is of the form

$$\sum_t d_{st} \log \sum_{ij} a_{ijt} \pi_{j|i} \geq 0 \,.$$

ML estimation involves finding the saddle point of

$$\begin{aligned}\mathcal{L} &= \sum_{ij} n_{ij} \log \pi_{j|i} + \sum_i \gamma_i \left(\sum_j \pi_{j|i} - 1 \right) \\ &\quad + \sum_r \lambda_r \left(\sum_t c_{rt} \log \sum_{ij} a_{ijt} \pi_{j|i} \right) + \sum_s \delta_s \left(\sum_t d_{st} \log \sum_{ij} a_{ijt} \pi_{j|i} \right) ,\end{aligned}$$

with

$$\delta_s \geq 0 \,.$$

Here, γ_i, λ_r and δ_t denote Lagrange multipliers.

The only difference compared to the situation in which there are only equality constraints is that the Lagrange multipliers belonging to the inequality constraints should be at least 0. This means that an inequality constraint is activated only if it is violated.

In practice, estimation can be accomplished by transforming the Fisher-scoring algorithm proposed by Lang and Agresti (1994) and Bergma (1997) into an active-set method. Vermunt (1999) showed how to transform a simple uni-dimensional Newton algorithm for ML estimation with equality constraints into an active-set method. In active-set methods, at each iteration cycle the inequality restrictions which are no longer necessary (i.e., if $d_s < 0$) are de-activated, and the ones which are violated are activated. More general information on algorithms for optimization under equality and inequality constraints can, for instance, be found in Gill and Murray (1974).

Restricted latent class models

Let $\pi^k_{j_k|i}$ be $P(Y_k = j_k | X = i)$. In the context of latent class analysis, the equality and inequality restrictions that we used are special cases of the general form

$$\begin{aligned}\sum_k \sum_r c^k_{rt} \log \sum_{ij_k} a_{ij_k t} \pi^k_{j_k|i} &= 0 \,, \\ \sum_k \sum_t d^k_{st} \log \sum_{ij_k} a_{ij_k t} \pi^k_{j_k|i} &\geq 0 \,.\end{aligned}$$

Here, a_{ij_kt} specifies the relevant sums of probabilities. This makes it straightforward to switch from one type of log odds to the other. The c^k_{rt} and d^k_{st} define the linear equality and inequality constraints on logs of sums of probabilities. The first sum over items makes it possible specify between-item constraints.

Estimation can be performed by implementing the Fisher-scoring active-set method discussed above in the maximization (M) step of an EM algorithm (Dempster, Laird, and Rubin, 1977). This is similar to what Croon (1990, 1991) did with a pooling adjacent violaters algorithm, which is a method for dealing with certain types of inequality restrictions. Vermunt (1999) proposed an EM algorithm which implements an active-set algorithm based on unidimensional Newton in the M step.

An advantage of model estimation by means of the EM algorithm is that in the M step the same types of estimation methods can be used as if the latent variable were observed. The expectation (E) step of the EM algorithm is very simple in LCMs.

It is well-known that the log-likelihood function of LCMs may be multimodal and this problem typically becomes worse when imposing inequality constraints. A practical way out is to run the same model with multiple (say 10) sets of random starting values. Within a bootstrap (see below), the best procedure seems to be to start the estimation from the ML estimates.

Model testing

Let H_1 be the hypothesized order-restricted model and H_0 the more restrictive model obtained by transforming all inequality restrictions into equality restrictions. This could, for instance, be non-negative log odds-ratios (H_1) and independence (H_0). Whether H_1 fits the data can be tested using a standard likelihood-ratio statistic defined as

$$G^2 = 2 \sum_{ij} n_{ij} \ln \left(\frac{\hat{\pi}_{j|i}}{p_{j|i}} \right),$$

where $\hat{\pi}_{j|i}$ and $p_{j|i}$ denote estimated and observed probabilities, respectively. A complication in using this test statistic is, however, that it is not asymptotically χ^2 distributed. It has been shown that the above test statistic follows a chi-bar-squared distribution, which are weighted sums of chi-squared distributions, when H_0 holds (see, for example, Robertson, Wright, and Dykstra, 1988:321). Let S denote the number of inequality constraints, which is also the maximum number activated constraints. The p value can be estimated as follows:

$$P(G^2 \geq c) = \sum_{s=0}^{S} P(s) P(\chi^2_{(s)} \geq c)\,;$$

that is, as a weighted sum of asymptotic p values, where the probability of having s activated constraints, $P(s)$, serves as weight. This shows that we have to take into account that the number of activated constraints is a random variable. A problem associated with this formula is, however, that the computation of the $P(s)$'s is - except for some trivial cases - extremely complicated.

Rather than trying to compute or approximate of the $P(s)$'s, it also possible to determine the p values for the test statistic using parametric bootstrapping methods, which are also known as Monte Carlo studies. This relatively simple method, which involves empirically reconstructing the sampling distribution of the test statistic of interest, is the one followed here. Ritov and Gilula (1993) proposed such a procedure in ML correspondence analysis with ordered category scores. A simulation study by the same authors showed that parametric bootstrapping yields reliable results when applied in these models, which are special cases of the order-restricted LCMs presented in this paper. Langeheine, Pannekoek, and Van de Pol (1996) proposed using bootstrapping in categorical data analysis for dealing with sparse tables, which is another situation in which we cannot rely on asymptotic theory for the test statistics. Agresti and Coull (1996) used Monte Carlo studies in combination with exact tests to determine the goodness-of-fit of order-restricted binary logit models estimated with small samples.

In the parametric bootstrap procedure, T frequency tables with the same number of observations as the original observed table are simulated from the estimated probabilities under H_1.

For each of these tables, we estimate H_1 and compute the value of G^2. This yields an empirical approximation of the distribution of G^2. The estimated p value is the proportion of simulated tables with a G^2 that is at least as large as for the original table. The standard error of the estimated p value equals $\sqrt{p\,(1-p)/T}$.

References

Agresti, A. (1990). *Categorical data analysis.* New York: Wiley.

Agresti, A., Chuang, C., & Kezouh, A. (1986). Order-restricted score parameters in association models for contingency tables. *Journal of the American Statistical Association, 82,* 619-623.

Agresti, A., Coull, B.A. (1996). Order-restricted tests for stratified comparisons of binomial proportions. *Biometrics, 52*, 1103-1111.

Bartholomew, D.J., & Knott, M. (1999). *Latent variable models and factor analysis.* London: Arnold.

Becht M.C., Poortinga Y.H., & Vingerhoets A.J.J.M. (in press). Crying across countries. In: A.J.J.M.Vingerhoets & R.R. Cornelius (Eds.) *Adult crying: A biopsychosocial approach.* Reading UK: Harwood

Bergsma, W.P. (1997). *Marginal models for categorical data.* Tilburg, The Netherlands: Tilburg University Press.

Bock, D.R. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*, 29-51.

Croon, M.A. (1990). Latent class analysis with ordered latent classes. *British Journal of Mathematical and Statistical Psychology, 43,* 171-192.

Croon, M.A. (1991). Investigating Mokken scalability of dichotomous items by means of ordinal latent class analysis. *British Journal of Mathematical and Statistical Psychology, 44,* 315-331.

Dardanoni, V. & Forcina, A. (1998). A unified approach to likelihood inference or stochastic orderings in a nonparametric context. *Journal of the American Statistical Association, 93,* 1112-1123.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Ser. B.*, 39, 1-38.

Gill, P.E., & Murray, W. (1974). *Numerical methods for constrained optimization.* London: Academic Press Inc..

Goodman, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika, 61*, 215-231.

Heinen, T. (1996). *Latent class and discrete latent trait models: similarities and differences.* Thousand Oakes: Sage Publications.

Hemker, B.T., Sijtsma, K., Molenaar, I.W., & Junker, B.W. (1997). Stochastic ordering using the latent trait and the sum score in polytomous IRT models. *Psychometrika, 62*, 331-347.

Hoijtink, H., & Molenaar, I.W. (1997). A multidimensional item response model: constrained latent class analysis using the Gibbs sampler and posterior predictive checks. *Psychometrika, 62,* 171-190.

Lang, J.B., & Agresti, A. (1994). Simultaneously modeling joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association*, 89, 625-632.

Langeheine, R., Pannekoek, J., & Van de Pol, F. (1996). Bootstrapping goodness-of-fit measures in categorical data analysis. *Sociological Methods and Research*, 24, 492-516.

Lindsay, B., Clogg, C.C., & Grego, J. (1991). Semiparametric estimation in the Rasch model and related models, including a simple latent class model for item analysis. *Journal of the American Statistical Association, 86,* 96-107.

Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

Mellenbergh, G.J. (1995). Conceptual notes on models for discrete polytomous item responses *Applied Psychological Measurement, 19*, 91-100.

Molenaar, I.W. (1997). Nonparametric models for polytomous responses. In W.J. van der Linden, & R.K. Hambleton (Eds.), *Handbook of modern item response theory.* (pp. 369-380). New York: Springer.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*, 159-177.

Ritov, Y., & Gilula, Z. (1993). Analysis of contingency tables by correspondence models subject to order constraints. *Journal of the American Statistical Association*, 88, 1380-1387.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph, 17.*

Sijtsma, K., & Hemker, B.T. (1998). Nonparametric polytomous IRT models for invariant item ordering, with results for parametric models. *Psychometrika, 63*, 183-200.

Robertson, T., Wright, F.T, & Dykstra, R. (1988). *Order restricted statistical inference.* New York: Wiley.

Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology, 43*, 39-55.

Van der Ark, L.A. (2001). An overview of relationships in polytomous IRT and some applications. *Applied Psychological Measurement*, this issue.

Vermunt, J.K. (1993). ℓ_{EM}*: log-linear and event history analysis with missing data using the EM algorithm.* WORC PAPER 93.09.015/7. Tilburg University, The Netherlands.

Vermunt, J.K. (1997). ℓ_{EM}*: A general program for the analysis of categorical data.* User's Manual. Tilburg University, The Netherlands.

Vermunt, J.K. (1999). A general class of non-parametric models for ordinal categorical data. M.Sobel and M. Becker (eds.). *Sociological Methodology, 29*, 187-224.

Table 1. Observed cross-classification of "Crying from Distress" and "Feeling Relieved after Crying"

Crying from Distress	Feeling Relieved after Crying: less	same	more
1=low	61	195	438
2	78	158	581
3	38	102	518
4	46	119	572
5=high	53	106	597

Table 2. Log cumulative odds-ratios of "Feeling Relieved after Crying" for adjacent categories "Crying from Distress": observed and estimated under the constraints of equation (3)

	Feeling Relieved after Crying	
Crying from Distress[1]	less versus same or more	less or same versus more
1 versus 2	-.09 / .00	.36 / .39
2 versus 3	.54 / .40	.41 / .34
3 versus 4	-.08 / .00	-.07 / .00
4 versus 5	-.12 / .00	.08 / .10

1. The scale "Crying from Distress" was collapsed into 5 levels denoted by the numbers 1 to 5.

Table 3. Test results for the ordinal logit models estimated with the crying data

Model	G^2 value[1]	df[2]	p value[3]
Independence	76.6	8	.00
Parametric with restrictions (1) and (2)			
Cumulative	23.2	7	.00
Adjacent	34.0	7	.00
Continuation I	21.7	7	.00
Continuation II	39.6	7	.00
Non-parametric with restrictions (3)			
Cumulative	1.3	4	.56
Adjacent	7.3	5	.14
Continuation I	1.3	4	.58
Continuation II	7.3	5	.11

1. G^2 is the likelihood-ratio chi-squared statistic.

2. The reported number of degrees of freedom for the order-restricted models is the number of activated constraints.

3. The p values of the models with inequality constraints are estimated on the basis of 1000 bootstrap samples.

Table 4. Test results for the LCM estimated with the "Crying from Distress" items data

Model	G^2 value[1]	df[2]	p value[3,4]
Unrestricted			
1 class	4597.5	72	.00
2 classes	871.2	63	.00
3 classes	86.8	54	.01
4 classes	53.4	45	.28
Parametric 4-class with restrictions (4) and (5)			
Cumulative (graded response)	187.2	65	.00
Adjacent (partial credit)	223.9	65	.00
Continuation I (sequential)	202.3	65	.00
Continuation II (sequential)	212.5	65	.00
Non-parametric 4-class with restrictions (7)			
Cumulative (graded response)	56.4	48	.28
Adjacent (partial credit)	69.2	52	.07
Continuation I (sequential)	61.7	51	.23
Continuation II (sequential)	60.2	47	.17
Non-parametric 4-class with restrictions (7) and (8)			
Cumulative (graded response)	56.4	48	.30
Adjacent (partial credit)	69.8	53	.08
Continuation I (sequential)	62.0	53	.24
Continuation II (sequential)	60.6	50	.16

1. G^2 is the likelihood-ratio chi-squared statistic.

2. The reported number of degrees of freedom for the order-restricted models equals the df of the unrestricted 4-class model (=45) plus the number of activated constraints.

3. The p values of all models are estimated on the basis of 1000 bootstrap samples.

4. With a sample size of 3821 and only 3 empty cells in the 3^4 table, there are no serious sparseness making model testing problematic. This means that for the models without inequality constraints, we also could have used the asymptotic p values.