PSYCHOMETRIKA—VOL. 69, NO. 4, 641–654
DECEMBER 2004

THE ORDER-RESTRICTED ASSOCIATION MODEL: TWO ESTIMATION ALGORITHMS AND ISSUES IN TESTING

FRANCISCA GALINDO-GARRE

ACADEMIC MEDICAL CENTER, UNIVERSITY OF AMSTERDAM

JEROEN K. VERMUNT

TILBURG UNIVERSITY

This paper presents a row-column (RC) association model in which the estimated row and column scores are forced to be in agreement with an a priori specified ordering. Two efficient algorithms for finding the order-restricted maximum likelihood (ML) estimates are proposed and their reliability under different degrees of association is investigated by a simulation study. We propose testing order-restricted RC models using a parametric bootstrap procedure, which turns out to yield reliable p values, except for situations in which the association between the two variables is very weak. The use of order-restricted RC models is illustrated by means of an empirical example.

Key words: Row-column association models, order-constraints, ML estimation algorithms, parametric bootstrap.

1. Introduction

Nowadays, several statistical tools are available to analyze ordinal categorical data, such as correspondence analysis, regression models for transformed cumulative probabilities, and log-linear and log-bilinear association models (see, e.g., Agresti, 2002; Clogg & Shihadeh, 1994). Goodman (1979) presented a class of log-linear and log-bilinear models to study the bivariate association between ordinal variables. This family contains four types of association models—uniform, row, column and row-column (RC)—suited for the analysis of ordinal data.

Nevertheless, association models are not really ordinal models because ordinal models assume a monotone relationship, no more and no less. The uniform association model assigns a priori scores to the categories of the row and column variables, which means that the variables are treated as interval level. The row model assumes that the column scores are known and that the row scores are unknown parameters. This model treats the column variable as an interval level variable and (since there is no guarantee that the estimated row scores have the assumed order) the row variables as nominal. The same applies to the column model. In the log-bilinear RC association model, both the row and column scores are estimated without order restrictions. Again, there is no guarantee that the category scores have the right order since the same ML estimates would be obtained if the levels are permuted in any way. Therefore, some restrictions should be imposed on the row and column scores to analyze ordinal relations.

Several methods have been proposed to overcome the problem that row or column scores do not have the assumed ordering. A first class of methods adapts the Goodman (1979) unidimensional Newton algorithm to deal with inequality restrictions. Both the work of Agresti, Chuang, and Kezouh (1987) on order-restricted row models and of Ritov and Gilula (1991) on order-restricted RC models fit within this framework. A different type of method based on

Francisca Galindo performed this research as a part of her PhD. dissertation project at Tilburg University.

Requests for reprints should be addressed to F. Galindo-Garre, Department of Clinical Epidemiology and Biostatistics, Academic Medical Center, P.O. Box 226600, 1100 DE Amsterdam, The Netherlands. Email: F.GalindoGarre@amc.uva.nl

0033-3123/2004-4/1998-0695-A $00.75/0

using prior distributions for row and column scores was proposed by Agresti and Chuang (1986). Recently, Bartolucci and Forcina (2002) showed how to define the RC model within the marginal modeling framework, that is, as a reduced-rank structure on the matrix of log-odds ratios. Within this framework, inequality constraints can be introduced to obtain order-restricted variants of RC models for various types of odds ratios.

This paper presents two simple algorithms, which are adaptations of Goodman's (1979) algorithm to order-restricted RC association models. The first method is a pooling adjacent violators algorithm (Robertson, Wright, & Dykstra, 1988) and the second one is an active-set algorithm (Gill & Murray, 1974). Both methods can also be used for ML estimation of order-restricted row models. The proposed algorithms overcome some of the problems associated with the procedure of Ritov and Gilula (1991). Moreover, the methods are simpler than the more general procedure proposed by Bartolucci and Forcina (2002) and can easily be implemented in existing RC modeling programs or routines that are based on Goodman's algorithm.

This paper is built up as follows. First, unrestricted and order-restricted RC models are described. Second, ML estimation of their model parameters is discussed. Then, the performance of the proposed algorithms, described in the former paragraph, and the Ritov–Gilula method are evaluated by a simulation study. Next, the testing of this model is investigated, and the parametric bootstrap (Efron & Tibshirani, 1993, sec. 21.5) is used to approximate p values for a two-hypotheses test, and its performance is investigated in a simulation study. Finally, the new approach is illustrated by means of an empirical example.

2. Description of the RC Model

Let n_{ij} and m_{ij} denote an observed and an expected cell count, respectively, in an $I \times J$ table. The assumed model for the expected frequencies is a log-bilinear RC association model, that is,

$$\log m_{ij} = \lambda + \lambda_i^R + \lambda_j^C + \phi \mu_i \nu_j. \tag{1}$$

The λ, λ_i^R, and λ_j^C parameters are standard log-linear effects, ϕ is the association parameter, and μ_i and ν_j are unknown row and column scores. For identification, some restrictions have to be imposed on the row and column scores and on the log-linear parameters, for instance,

$$\begin{aligned} \sum_i \mu_i &= \sum_j \nu_j = 0, \\ \sum_i \mu_i^2 &= \sum_j \nu_j^2 = 1, \\ \sum_i \lambda_i^R &= \sum_j \lambda_j^C = 0. \end{aligned} \tag{2}$$

The uniform, row, and column association models can be seen as special cases of the RC model.

As was shown by Goodman (1979), the parameters of the RC model are directly related to the log-local odds ratios, that is,

$$\log \frac{m_{ij} m_{i+1j+1}}{m_{i+1j} m_{ij+1}} = \phi(\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j).$$

Although the RC model was originally proposed by Goodman (1979) for the analysis of two-way tables having ordered categories, there is no guarantee that the ML solution will be ordinal unless, assuming $\phi > 0$, the row and column scores are constrained to be monotonically increasing or decreasing.

3. ML Estimation of the Unrestricted RC Model

For the ML estimation discussed below, it is somewhat easier to use a slightly different formulation of the above RC model, such as

$$\log m_{ij} = \lambda + \lambda_i^R + \lambda_j^C + \rho_i \sigma_j, \tag{3}$$

where

$$\sum_i \rho_i = \sum_j \sigma_j = 0. \tag{4}$$

The equivalence between the formulation in Equations (1) and (3) becomes clear if one notes that $\phi \mu_i \nu_j = \rho_i \sigma_j$, where $\rho_i = \phi^\gamma \mu_i$ and $\sigma_j = \phi^\delta \nu_j$ for any γ and δ whose sum is one. The simplest choice is obtained when $\gamma = \delta = \frac{1}{2}$ because it gives equal weights to the row and column scores, and ensures that the sum of squares of row scores and the sum of squares of column scores are equal.

The likelihood equations for the log-linear parameters have the well-known form

$$\sum_j n_{ij} - \sum_j m_{ij} = 0,$$

$$\sum_i n_{ij} - \sum_i m_{ij} = 0.$$

These likelihood equations can be solved using simple iterative proportional fitting (IPF) adjustments. The likelihood equations for the ρ_i and σ_j parameters are

$$\sum_j n_{ij}\sigma_j - \sum_j m_{ij}\sigma_j = 0 \quad \text{and} \quad \sum_i n_{ij}\rho_i - \sum_i m_{ij}\rho_i = 0,$$

respectively. It should be noted that the conditions described in the above likelihood equations are necessary but not sufficient for a solution to be the ML solution. Because the log-likelihood function is not necessarily concave, there may be local maxima. A manner to decrease the probability of ending up with a local maximum is to re-estimate the model various times using different sets of (random) starting values for the row and column scores.

As was already shown by Goodman (1979), the likelihood equations for the RC model can be solved by means of a simple uni-dimensional Newton algorithm (see also Clogg, 1982; Becker, 1990). This method solves these equations with the following updates of the ρ_i and σ_j parameters:

$$\rho_i^{(t)} = \rho_i^{(t-1)} - \frac{g(\rho_i)^{(t)}}{H(\rho_i)^{(t)}} = \rho_i^{(t-1)} - \frac{\sum_j n_{ij}\sigma_j^{(t-1)} - \sum_j m_{ij}^{(t-1)}\sigma_j^{(t-1)}}{-\sum_j m_{ij}^{(t-1)}(\sigma_j^{(t-1)})^2}, \tag{5}$$

and

$$\sigma_j^{(t)} = \sigma_j^{(t-1)} - \frac{g(\sigma_j)^{(t)}}{H(\sigma_j)^{(t)}} = \sigma_j^{(t-1)} - \frac{\sum_i n_{ij}\rho_i^{(t)} - \sum_i m_{ij}^{(t)'}\rho_i^{(t)}}{-\sum_i m_{ij}^{(t)'}(\rho_i^{(t)})^2}, \tag{6}$$

where $\rho_i^{(t)}$ and $\sigma_j^{(t)}$ denote the tth approximation for the ρ_i and σ_j parameters and $m_{ij}^{(t)}$ denote the tth approximation for the expected frequencies. The numerator g in Equations (5) and (6) is the first partial derivative of the log-likelihood function with respect to the parameter concerned (an element of the gradient vector) and the denominator H is the second partial derivative (a diagonal element of the Hessian matrix).

As can be seen, each iteration cycle consists, besides the standard IPF adjustments for the log-linear parameters, of two steps: one in which the ρ_i are updated, treating the σ_j as fixed, and one in which the σ_j are updated, treating the ρ_i as fixed. It is important to note that the ρ_i parameters are updated independently of one another. The same applies to the σ_j parameters. After updating the ρ_i parameters, they are centered (see the condition given in Formula (4)) and the estimated expected frequencies are updated, which yields $m_{ij}^{(t)'}$. The same procedure is followed after updating the σ_j parameters.

The above uni-dimensional Newton method can also be used for estimating row and column models. This just involves treating either the column or the row scores as fixed rather than as random quantities.

4. ML Estimation of the Order-Restricted RC Model

In terms of the model formulation in Equation (3), ordinality is defined as

$$\rho_i \leq \rho_{i+1},$$

and either

$$\sigma_j \leq \sigma_{j+1}$$

or

$$\sigma_j \geq \sigma_{j+1},$$

depending on whether there is a positive or negative relationship. Thus, the row scores have to be monotonically nondecreasing while the column scores can either be postulated to be monotonically nondecreasing or nonincreasing. Next, we describe four algorithms to estimate the RC model under these constraints.

4.1. The Ritov–Gilula Algorithm

Ritov and Gilula (1991) proposed to obtain ML estimates of the order-restricted RC model by a pooling adjacent violators algorithm, which is a well-known class of procedures in the field of ordered statistical inference (see Robertson, Wright, & Dykstra, 1988). The amalgamation of categories that are out of order is not determined directly on the ρ_i and σ_j parameters but on the quantities $E_i(\sigma)$ and $F_j(\rho)$, which are defined as

$$E_i(\sigma) = \sum_j \frac{n_{ij}\sigma_j}{n_{i.}}, \tag{7}$$

$$F_j(\rho) = \sum_i \frac{n_{ij}\rho_i}{n_{.j}}. \tag{8}$$

Note that these are the sufficient statistics for the unrestricted row and column parameters divided by the corresponding marginal frequencies.

Ritov and Gilula (1991) proved that pooling adjacent violators of $E_i(\sigma)$ and $F_j(\rho)$ using the marginal observed frequencies $n_{i.}$ and $n_{.j}$ as weights yield information on which categories scores have to be equated. The necessary conditions for the ML solution of the order-restricted RC model are that the pooled $E_i(\sigma)$ and $F_j(\rho)$ are monotone and that the likelihood equations are fulfilled. The likelihood equations concern the table in which the equated categories are collapsed. This implies, for instance, that if rows 3, 4, and 5 are equated, the unrestricted likelihood equations for these three rows have to be summed.

Agresti, Chuang, and Kezouh (1987) used the same principle of pooling adjacent violating $E_i(\sigma)$ and $F_j(\rho)$ in the estimation of order-restricted R and C models. There, however, either the σ_j or the ρ_i are known quantities, which makes it possible to determine the categories that have to be collapsed from the data. Since in RC models both the σ_j and the ρ_i are unknown, the order violations in $E_i(\sigma)$ and $F_j(\rho)$ are not independent of one another.

According to Ritov and Gilula (1991), asymptotically, amalgamation can be done independently for rows and columns using the unrestricted ML estimates for σ_j and ρ_i in the above formulas for $E_i(\sigma)$ and $F_j(\rho)$. This should yield information on which ρ_i and σ_j parameters must be equated to obtain the order-restricted solution.

Their procedure, thus, consists of four steps:

1. estimate the unrestricted RC model,
2. compute $E_i(\sigma)$ and $F_j(\rho)$ using the unrestricted estimates for ρ_i and σ_j,
3. determine which row and column scores should be equated by pooling adjacent violators in the $E_i(\sigma)$ and $F_j(\rho)$, and
4. estimate the RC model with the necessary equality restrictions.

The equality restrictions can, for instance, be imposed by estimating an unrestricted RC model for the table in which the equated categories are collapsed.

Despite that Ritov and Gilula show that their method works asymptotically, in practice it often fails to find the equality restrictions yielding the global order-restricted ML solution. This is caused by the fact that restrictions on rows and columns are not independent of one another. Although asymptotically—which means that the model holds in the population and that the sample size goes to infinity—it does not make a difference whether we determine $E_i(\sigma)$ using the order-restricted or the unrestricted estimates for the column scores, in practice, it makes a difference. The same applies for $F_j(\rho)$.

4.2. A Naive Algorithm

As pointed out by Ritov and Gilula (1991), there is no guarantee, in the nonasymptotic case, that the global maximum can be found with their algorithm. An alternative naive algorithm can be formulated that will always find the global maximum for the order-restricted RC model. It consists of estimating independently all possible models that arise from imposing equality constraints between adjacent row and column scores. Given that there are $I-1$ possible equalities on adjacent row scores and $J-1$ possible equalities on adjacent column scores, there are $2^{(I+J-2)}$ different RC models with the relevant equality constraints. Each of these $2^{(I+J-2)}$ models has to be estimated. As the ordered ML solution it is selected the model that gives the highest log-likelihood value among the models with correctly ordered row and column scores.

Because $2^{(I+J-2)}$ different models have to be estimated, this naive algorithm is a very time-consuming method. However, since it always finds the order-restricted ML solution, it is very well suited as a benchmark for alternative procedures.

4.3. A Pooling Adjacent Violators Algorithm

Instead of estimating all possible models as is done in the naive method, it is also possible to transform the Ritov–Gilula procedure into a pooling adjacent violators (PAV) algorithm that converges to the order-restricted ML solution. The proposed modification is to determine $E_i(\sigma)$ and $F_j(\rho)$ at each iteration cycle rather than from the unrestricted ML solution. In other words, the necessary order restrictions on the row scores are determined given the current order-restricted estimates for column scores and vice versa. This algorithm fits very well within the framework of

the uni-dimensional Newton algorithm, in which row scores are updated given the current values of the column scores, and column scores are updated given the current values of the row scores.

In the proposed PAV algorithm, at each iteration cycle the updating of the row scores consists of three steps: (1) determine which rows should be equated given the current estimates of the column scores using the method proposed by Ritov and Gilula, (2) perform an unrestricted update of the row scores, and (3) pool the row scores that should be equated using the marginal observed frequencies as weights. Equivalent steps are applied to obtain improved estimates for the column scores. More details are provided in the Appendix.

4.4. An Active-Set Algorithm

An alternative to the above PAV algorithm is to modify the uni-dimensional Newton method into an active-set algorithm (Gill & Murray, 1974). Active-set or activated-constraints algorithms are commonly used to solve optimization problems with inequality constraints. In our case, row and column scores can be updated using Equation (5) or (6) as long as they are not out of order, that is, as long as they belong to the inactive set. If there are order violations, say at iteration t, the scores that are out of order have to be equated and are thus moved to the active set. Once scores belong to the active set, in subsequent iterations, it must be checked whether an unrestricted update would again yield an order violation. If so, they have to remain equal, otherwise they are allowed to become unequal again (to be moved to the inactive set). More formally:

- For row (column) scores belonging to the inactive set, perform an unrestricted update using Equation (5) or (6) and check whether there are order violations ($\rho_i^t > \rho_{i+1}^t$). If so, equate the scores that are out of order (e.g., $\rho_i^t = \rho_{i+1}^t = c$). It is not important which provisional value, c, is taken when equating the scores. For instance, c may be the unweighted mean or, as in the PAV algorithm, the marginally weighted means of the corresponding unrestricted scores
- For row (column) scores belonging to the active set, check whether an unrestricted update using Equation (5) or (6) again yield an order violation. If so, they have to remain equal; otherwise they are allowed to become unequal. Note that treating parameters as equal implies summing the numerators (g) and the denominators (H) of the unrestricted updates.

In contrast to the Ritov–Gilula algorithm, which determines the order violations from the unrestricted ML solution, the PAV and the active-set algorithms settle the order violations at each iteration. Even though both algorithms make use of the uni-dimensional Newton updating scheme, they use different approaches to find the necessary equality constraints. While the PAV procedure uses $E_i(\sigma)$ and $F_j(\rho)$ to determine which scores should be equated after an unrestricted update, the active-set method equates scores that are out of order and keeps them equal in the next iterations as long as the gradients show that an unrestricted update would yield an order violation. As is shown in the Appendix, for scores that are equal at iteration $t-1$ and that should remain equal at iteration t, the two updating schemes are equivalent. This means that once the set of necessary constraints is found, both algorithms converge to the same order-restricted solution.

4.5. A Simulation Study

In order to show that the PAV and the active-set algorithms perform better than the Ritov–Gilula algorithm, we carried out a Monte Carlo study. In the evaluation of these procedures, we assumed that the naive algorithm always reaches the equalities on the rows and columns that produce the order-restricted ML solution, and we used it as reference. Since the PAV and the active-set algorithms produce about the same order-restricted estimates, we only used the active-set algorithm in the simulation. More specifically, we investigated, under several conditions,

whether the active-set algorithm yields the solution obtained with the naive method more often than the Ritov–Gilula algorithm.

In the simulation study, a 5×3 contingency table was taken as starting point, and samples were generated from a model of the form presented in Equation (1). The λ_i^R and λ_j^C parameters were assumed to be equal to zero and the centering and scaling constraints on the μ_i and ν_j parameters were as described in Equation (2). The influence of three factors was investigated:

1. strength of the association between the variables (3 conditions): $\phi = 0$, $\phi = 0.3$, or $\phi = 3.0$;
2. relative distances between rows and between columns (3 conditions): equal-distant row and column scores, two equal row scores ($\mu_1 = \mu_2$), or two equal row and two equal column scores ($\mu_1 = \mu_2$ and $\nu_2 = \mu_3$);
3. sample size (2 conditions): $N = 1000$ or $N = 100$.

A thousand data sets were drawn under the 14 conditions obtained by crossing the above three factors. Note that distances between rows and columns are not varied when $\phi = 0$. For each data set, we estimated the ordered RC model using the naive method, the method proposed by Ritov and Gilula (1991), and the active-set method using $\phi = 1$ and equal-distant row and column scores as starting values. In the active-set method, we equated scores that are out of order in the PAV manner (see the Appendix), which makes this method almost equivalent to the PAV algorithm. It may be expected that it becomes harder to find the ML solution with a weaker association, with less distant scores, and with a smaller sample size, that is, when there is a higher probability of having several order violations.

Table 1 reports the proportion of samples in which the value of the likelihood-ratio statistic (L^2) obtained with the active-set and Ritov–Gilula methods is larger than the value obtained with the naive method (the ML solution). We also report the average difference in L^2 between the last two methods across the 1000 replications. L^2, which is taken as a measure of the fit of the model, represents the distance between estimated frequencies and the data. This statistic minimizes with the parameter values maximizing the log-likelihood (more details about L^2 can be found in the next section). As can be seen, the Ritov–Gilula method is reliable only with the largest sample size ($N = 1000$) and the strongest association ($\phi = 3$). With the small sample size, this method performs badly even for the strongest association, which is not surprising given that it is based

TABLE 1.

Proportion of Simulated Data Sets in which the L^2 Value of the Active-Set or Ritov–Gilula Method is Larger than the One of the Naive Method

		$N = 1000$			$N = 100$		
Population		Active-set	Ritov–Gilula		Active-set	Ritov–Gilula	
$\phi = 0.0$		.077[1]	.694	(.553)[2]	.084[1]	.658	(.625)
$\phi = 0.3$,	Equal-distance μ_i and ν_j	.005[1]	.283	(.123)	.042[1]	.727	(.571)
	$\mu_1 = \mu_2$	.005[1]	.322	(.184)	.063[1]	.756	(.594)
	$\mu_1 = \mu_2, \nu_2 = \nu_3$	.001[1]	.387	(.235)	.093[1]	.782	(.720)
$\phi = 3$,	Equal-distance μ_i and ν_j	.000	.000	(.000)	.000	.007	(.046)
	$\mu_1 = \mu_2$	.000	.000	(.000)	.000	.036	(.049)
	$\mu_1 = \mu_2, \nu_2 = \nu_3$	.000	.000	(.000)	.000	.067	(.071)

[1]These are results obtained with our default starting values. With three sets of random starting values, we get a perfect match, or a value of .000.

[2]Between braces we report the average difference in L^2 compared to naive method.

on asymptotic properties. The active-set method is more reliable under all conditions. However, Table 1 shows that the active-set method may also fail to find the global maximum, especially with a small sample and a weak association. In all these cases, the global maximum can be found by repeating the estimation using random starting values (three random starts sufficed in all cases). Therefore, using the ML criterium, we can conclude that the active-set algorithm will always find the best order-restricted solution if various random starting values are used.

5. Model Selection

One way of testing the order-restricted RC model is by comparing its log-likelihood value to the one of the unrestricted RC model. This test, which will be denoted as L^2_{01}, was studied by Ritov and Gilula (1991). A second test, denoted as L^2_{02}, is the goodness-of-fit test often referred to as the G^2 statistic. In our case, it involves comparing the log-likelihood values of the order-restricted RC model and the saturated model.

Let $\hat{m}^0_{ij}$ denote the estimated frequency in cell (i, j) under the order-restricted RC model, $\hat{m}^1_{ij}$ the corresponding estimated frequency under the unrestricted RC model, and $\hat{m}^2_{ij}$ the one under the saturated model. The latter is, of course, equal to the observed frequency n_{ij}. The two likelihood-ratio statistics are defined as follows:

$$L^2_{0k} = 2\sum_{ij} n_{ij} \log\left(\frac{\hat{m}^k_{ij}}{\hat{m}^0_{ij}}\right), \tag{9}$$

where $1 \leq k \leq 2$. A complication that arises from the use of these test statistics is that their asymptotic distribution depends on the number of constraints that needs to be activated, something that is not known a priori. Ritov and Gilula (1991) derived the asymptotic distribution of L^2_{01} as a mixture of chi-squared distributions whose weights equal the probabilities of having a certain number of constraints activated. This distribution is called a chi-bar-squared distribution.

The chi-bar-squared distribution may also be derived, for example, from the work of Shapiro (1985) or Bartolucci and Forcina (2002). Under the null hypothesis, the p value corresponding to a certain value of L^2_{01}, say c, is expressed as

$$P(L^2_{01} \geq c) = \sum_{l=0}^{l_{\max}} P(l) P(\chi^2_l > c),$$

where χ^2_l denotes a chi-squared random variable with l degrees of freedom. The $P(l)$ are non-negative weights summing to one and representing the probabilities that l out of the $l_{\max}$ possible constraints are activated. Though there is no expression for exact computation of the weights if $l_{\max} > 3$, Dardanoni and Forcina (1998, p. 1117) proposed a tractable method for estimating their values. The expression for $P(L^2_{02} \geq c)$ is obtained by replacing χ^2_l with $\chi^2_{l+df_1}$, where df_1 is the number of degrees of freedom of the unrestricted RC model.

Rather than using an asymptotic approach to obtain the p value associated with L^2_{01} and L^2_{02}, it may also be estimated using parametric bootstrap. This is a conceptually simple method based on an empirical reconstruction of the sampling distribution of the test statistic. Parametric bootstrap has been extensively used in the literature. For example, Ritov and Gilula (1993) proposed such a procedure in ML correspondence analysis with ordered category scores, Schoenberg (1997) advocated using bootstrap testing methods in a general class of constrained maximum likelihood problems, and Langeheine, Pannekoek, and Van de Pol (1996) proposed the use of bootstrap in categorical data analysis for dealing with sparse tables, which is another situa-

tion in which we cannot rely on asymptotic distribution functions for the test statistics. Recently, Vermunt (1999, 2001) proposed using this procedure to test the goodness-of-fit of models with inequality constraints on the parameters.

Suppose we want to perform both the L^2_{01} and L^2_{12} test by means of a parametric bootstrap. After estimating the unrestricted and order-restricted RC models with the data set at hand, B frequency tables with the same number of observations as the original data are simulated from the estimated probabilities under the order-restricted RC model. For each of these tables, we estimate both the unrestricted and order-restricted RC model and compute the values of the L^2_{01} and L^2_{02} statistics. The corresponding estimated p value is the proportion of simulated tables in which the L^2_{01} (L^2_{02}) value is at least as large as the one obtained with the original table. The standard errors of the estimated p values equal $(p(1-p)/B)^{1/2}$.

To evaluate the proposed bootstrap procedure, we performed a simulation study using the same 14 conditions as in the simulation reported in the previous section. Under each of these conditions we generated 1000 samples and estimated the p value with the bootstrap procedure for L^2_{01} and L^2_{02} using $B = 400$. Given the fact that the estimated model is true, parametric bootstrap can be decided to perform well if the proportion of samples rejected using a particular significant level is approximately equal to the correspondent nominal value. Table 2 reports the proportion of samples in which the order-restricted RC model is rejected at significance levels $\alpha = .50$ and $\alpha = .05$. As can be seen, the rejection proportion obtained with the parametric bootstrap is not always in agreement with these nominal levels.

For data simulated under the independence model, $\phi = 0$, both tests are somewhat too liberal. This means that, at the chosen α level, the order-restricted RC model is rejected more often than should be expected.

When the association is strong, parametric bootstrap yields rejection proportions close to the nominal α level for L^2_{02}. However, for L^2_{01}, it produces too conservative proportions, especially if the distances between row and column scores are large. A rejection proportion is said to be conservative if it is lower than the nominal value. It can, for example, be seen that with $\phi = 3$, equal-distance scores, and $N = 1000$, the proportion of samples in which the order-restricted RC model is rejected is only .001 instead of .05. What happens is that the bootstrap probabilities are almost always higher than .05 because in most replication samples L^2_{01} will be equal to zero. Such a L^2_{01} value of zero indicates that no constraints are activated in the order-restricted RC model and that, therefore, the order-restricted and unrestricted RC yield the same estimated frequencies. The same occurs with the smaller sample size condition.

If the association is weak and the sample size large, parametric bootstrap tends to be somewhat too conservative in both tests. The reason is that bootstrap replications tend to be more in

TABLE 2.
Proportion of Simulated Data Sets in which the Order-restricted RC Model Is Rejected at Three Different α Levels

		T1: Order-RC vs RC				T2: Goodness of fit			
		$N = 1000$		$N = 100$		$N = 1000$		$N = 100$	
Population	$\alpha =$	.50	.05	.50	.05	.50	.05	.50	.05
$\phi = 0$		.651	.071	.656	.072	.576	.060	.583	.064
$\phi = .3$,	Equal-distance μ_i, ν_j	.399	.025	.659	.074	.422	.036	.510	.040
	$\mu_1 = \mu_2$	.440	.033	.679	.078	.431	.036	.605	.078
	$\mu_1 = \mu_2, \nu_2 = \nu_3$	.541	.047	.707	.102	.500	.060	.542	.044
$\phi = 3$,	Equal-distance μ_i, ν_j	.142	.001	.427	.006	.478	.038	.478	.050
	$\mu_1 = \mu_2$	.484	.049	.332	.035	.478	.062	.543	.062
	$\mu_1 = \mu_2, \nu_2 = \nu_3$	.605	.061	.441	.043	.538	.058	.502	.063

agreement with the established order than the original sample (see Geyer, 1995). For example, if $\mu_1 \geq \mu_2$ in the population, the probability of drawing empirical samples in which the inequality induces an activated equality equals 0.5. However, this probability is smaller for the bootstrap samples if the equality is not activated in the empirical sample, something that happens in 50% of the cases. Under $N = 100$, the results are more accurate. It seems as if the extra variability caused by the smaller sample size compensates for this effect.

In conclusion, parametric bootstrap yields more accurate results when testing the goodness-of-fit (L^2_{02}) than when comparing the two nested RC models (L^2_{01}). The rejection proportions for L^2_{02} are close to their nominal values as long as the association between the row and column variables is strong enough. The encountered rejection proportions for the L^2_{01} statistic, however, are lower than their nominal levels, indicating that this test is too conservative.

6. An Empirical Example

Table 3 displays the relationship between number of siblings (S) and happiness (H). This example uses data reported by Clogg (1982, Table 2) in a paper on ordinal log-linear and log-bilinear models. The original three-way cross-classification was collapsed over the variable years of schooling. The question of interest is whether the association is of an ordinal nature, or, more precisely, whether there is a positive association between number of siblings and happiness.

The test results for the estimated models are reported in Table 4. As can be seen, the independence model does not fit the data ($L^2(1) = 26.27$, $df = 8$, $p < .01$), which indicates that there is an association between H and S. A model that is often used for the analysis of ordinal data is the uniform association model (Model 2). Note that in this model, both variables are treated as interval level variables. The uniform association model does not fit the data ($L^2(2) = 20.21$, $df = 7$, $p = .01$), which indicates that the assumption that the local odds ratios are constant is too strong. Nevertheless, the uniform association parameter is significant and, as can be seen from the parameters reported in Table 5, has the "expected" positive sign.

Less restrictive are the row and the column association models, which assume column and row-independent local-odds ratios, respectively. The row model does not fit the data ($L^2(3) = 17.52$, $df = 4$, $p < .01$), which indicates that H may not be treated as an interval level variable. In addition, the estimated scores for S are not ordered: The score for row 4 is much higher than for row 5. The column model fits ($L^2(4) = 8.36$, $df = 6$, $p = .21$), but again some category scores—$H = 1$ and $H = 2$—are out of order.

The RC model is less restrictive than the row and column models since it does not assume that one of the variables is an interval level variable. The unrestricted RC model fits the data quite well: $L^2(5) = 7.33$, $df = 3$, $p = .06$. A problem is, however, that neither the row or the column

TABLE 3.
Cross-Classification of Number of Siblings and Happiness: Observed Frequencies

Number of siblings	Happiness		
	Not too happy	Pretty happy	Very happy
0–1	99	155	19
2–3	153	238	43
4–5	115	163	40
6–7	63	133	32
8 +	99	118	47

TABLE 4.
Test Results for the Estimated Models

Model	L^2 value	df[1]	p value[2]
1. independence	26.27	8	.00
2. uniform association	20.21	7	.01
3. row association	17.52	4	.00
4. column association	8.36	6	.21
5. row-column association	7.33	3	.06
6. ordered row association	18.60	$4+1$	.00
7. ordered column association	8.84	$6+1$	.22
8. ordered row-column association	8.36	$3+1$	.08

[1]The reported number of degrees of freedom for the order-restricted models is the df of the model without constraints plus the number of activated constraints.
[2]The p values of the models with inequality constraints are estimated on the basis of 1000 bootstrap samples. The standard errors of these estimates are less than .01 for $p \leq .11$ and $p \geq .89$, and at most .02 for other p values.

scores have the correct order (see Model 5 in Table 5). More precisely, the order of rows 3 and 4 and of columns 1 and 2 is incorrect. This makes the results difficult to interpret.

Models 6 and 7 are the order-restricted row and column models. Like the unrestricted row and column model, the ordinal row model performs badly ($L^2(6) = 18.60$, $df = 4+1$, $p = .00$) whereas the ordinal column model performs well ($L^2(7) = 8.84$, $df = 6+1$, $p = .22$). This indicates that the row variable, number of siblings (S), may be treated as interval level and the column variable, happiness (H), as ordinal. From the parameters of Models 6 and 7 reported in Table 5, it can be seen that rows 4 and 5 are equated in the row model, and columns 1 and 2 in the column model.

In addition, the order-restricted RC model was specified for the data reported in Table 3. This model performs quite well: $L^2(8) = 8.36$, $df = 3+1$, $p = .08$. Although in the unrestricted RC model (Model 4) there were two order violations in the estimated row and column scores, the ML solution for the ordinal RC model contains only one activated inequality constraint: The score for column 1 is equated to the score for column 2 (see Model 8 in Table 5). This demonstrates that it is dangerous to specify ordinal models by post hoc equality constraints because of the dependence between the row and column scores. Also, the Ritov–Gilula procedure yields a suboptimal solution with an L^2 value of 8.60.

To demonstrate the strength of the active-set algorithm proposed in this paper, an order-restricted RC model is specified that assumes a negative rather than a positive relationship between S and H. For this badly fitting model, it is much harder to determine which row and

TABLE 5.
Estimates for the Association Parameters of Models 2–8

	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
ϕ	0.32	0.41	0.62	0.63	0.37	0.63	0.66
μ_1		−0.59		−0.67	−0.64		−0.71
μ_2		−0.25		−0.24	−0.27		−0.24
μ_3		−0.10		0.13	−0.11		0.10
μ_4		0.73		0.10	0.51		0.25
μ_5		0.22		0.68	0.51		0.60
ν_1			−0.33	−0.27		−0.41	−0.41
ν_2			−0.48	−0.53		−0.41	−0.41
ν_3			0.81	0.80		0.82	0.82

column scores should be equated in the ordered ML solution because there are many order violations. Whereas the Ritov–Gilula procedure equates all column scores yielding an independence model, both the active-set method and the naive method yield a lowest L^2 value of 25.31. This solution contained four activated equality constraints: The first four row scores and the last two column scores were equated.

7. Discussion

This paper presents two algorithms for obtaining order-restricted ML estimates of RC association models: an active-set method and a pooling adjacent violators algorithm. Both algorithms are adaptations of the uni-dimensional Newton method proposed by Goodman (1979) to deal with inequality restrictions. The reported simulation study shows that the new methods perform very well.

As far as model testing is concerned, we studied the performance of a simple parametric bootstrap procedure. For the goodness-of-fit test of the order-restricted model, this procedure yields reliable p values as long as the association is strong enough. When the association is too weak, the bootstrap p values are too conservative. We, therefore, advise the researcher to first test the independence model against the order-restricted RC, that is, to check whether any association between variables is present. Wang (1996) showed that parametric bootstrap yields reasonable rejection proportions for these types of tests. The conditional test between the order-restricted model and the unrestricted RC model is too conservative. Parametric bootstrap will yield downward biased rejection proportions when both models fit the data equally well, that is, when the association is strong and category scores are far apart and the two models cannot be distinguished. The testing procedure proposed by Bartolucci and Forcina (2002) might be preferred in such situations.

Some research has been done into situations in which the parametric bootstrap yields biased p values and some adjustment methods have been proposed, such as the adjusted active set bootstrap (see Geyer, 1995). These kinds of methods are, however, difficult to implement and present certain arbitrariness. Future research may aim at studying whether these methods can be used to resolve the encountered deficiencies of the parametric bootstrap in the context of the order-restricted RC model.

As Goodman's uni-dimensional Newton method can be used for all kinds of extensions of the simple RC model for bivariate associations, the proposed active-set and PAV methods can be used for estimating a much more general class of RC models than discussed in this paper. Examples are models for multi-way cross-classifications assuming order-restricted partial and conditional associations, as well as models for squared two-way tables containing additional log-linear terms like diagonal parameters to correct for the over-representation in the diagonal elements.

Another possible application of the order-restricted RC model is in latent structure analysis. It could be used to specify the nature of the relationship between a discrete latent variable and a set of ordinal indicators. This yields either a variant of the ordinal latent class model proposed by Croon (1990) or a latent trait model in which the underlying latent distribution is approximated by means of a limited number of nodes (Vermunt, 2001). Estimation could be performed by implementing one of the proposed estimation methods in the M step of an EM algorithm (Dempster, Laird, & Rubin, 1977).

Another interesting direction for future research is the use of Bayesian Markov chain Monte Carlo methods for estimating parameters and assessing fit of the order-restricted RC model. It is well known that the specification of inequality constraints is straightforward within this framework (see, for instance, Hoijtink & Molenaar, 1997).

Appendix

In this appendix, we describe the PAV and active-set restricted updates of the scores of row i and $i+1$, as well as demonstrate the similarity between the two methods. The results also apply to the column scores.

Recall that an unrestricted update of a row score at iteration t is of the form

$$\rho_i^{(t)} = \rho_i^{(t-1)} - \frac{g(\rho_i)^{(t)}}{H(\rho_i)^{(t)}}.$$

To simplify notation, we denote $g(\rho_i)^{(t)}$ and $H(\rho_i)^{(t)}$ by g_i and H_i.

In the active-set method, equated scores that should remain equal are updated by

$$\rho_{i,i+1}^{(t)} = \rho_{i,i+1}^{(t-1)} - \frac{g_i + g_{i+1}}{H_i + H_{i+1}}.$$

With $\rho_{i,i+1}^{(t)}$, we denote that $\rho_i^{(t)} = \rho_{i+1}^{(t)} = \rho_{i,i+1}^{(t)}$.

In the PAV algorithm, amalgamation of the scores for rows i and $i+1$ is as follows:

$$\begin{aligned}\rho_{i,i+1}^{(t)} &= \frac{n_{i.}\rho_i^{(t)} + n_{i+1.}\rho_{i+1}^{(t)}}{n_{i.} + n_{i+1.}} \\ &= \frac{n_i\left[\rho_i^{(t-1)} - \frac{g_i}{H_i}\right] + n_{i+1}\left[\rho_{i+1}^{(t-1)} - \frac{g_i}{H_i}\right]}{n_i + n_{i+1}}.\end{aligned}$$

Let us now consider the situation in which the scores for rows i and $i+1$ are equal in iteration $t-1$. In this case, the previous equation simplifies to

$$\rho_{i,i+1}^{(t)} = \rho_{i,i+1}^{(t-1)} - \frac{n_i\frac{g_i}{H_i} + n_{i+1}\frac{g_{i+1}}{H_{i+1}}}{n_i + n_{i+1}}.$$

Using the fact that in this situation $H_{i+1}(t) = H_i(t)\frac{n_{i+1}}{n_i}$, it can be shown that the PAV and active-set method yield equivalent updates. More precisely,

$$\begin{aligned}\frac{n_i\frac{g_i}{H_i} + n_{i+1}\frac{g_{i+1}}{H_{i+1}}}{n_i + n_{i+1}} &= \frac{n_i\frac{g_i}{H_i} + n_{i+1}\frac{g_{i+1}}{H_i\frac{n_{i+1}}{n_i}}}{n_i + n_{i+1}} \\ &= n_i\frac{g_i}{H_i(n_i + n_{i+1})} + n_i\frac{g_{i+1}}{H_i(n_i + n_{i+1})} \\ &= \frac{g_i + g_{i+1}}{H_i\left(1 + \frac{n_{i+1}}{n_i}\right)} = \frac{g_i + g_{i+1}}{H_i + H_{i+1}}.\end{aligned}$$

This is an important result because it shows that once the necessary equality constraints are found, both algorithms converge to the same order-restricted solution.

References

Agresti, A. (2002). *Categorical data analysis*. New York: Wiley.

Agresti, A., & Chuang, C. (1986). Bayesian and maximum likelihood approaches to order restricted inference for models from ordinal categorical data. In R. Dykstra & T. Robertson (Eds.), *Advances in ordered statistical inference* (pp. 6–27). Berlin: Springer-Verlag.

Agresti, A., Chuang, C., & Kezouh, A. (1987). Order-restricted score parameters in association models for contingency tables. *Journal of the American Statistical Association, 82*, 619–623.

Bartolucci, F., & Forcina, A. (2002). Extended RC association models allowing for order restrictions and marginal modeling. *Journal of the American Statistical Association, 97*, 1192–1199.

Becker, M.P. (1990). Algorithm AS253: Maximum likelihood estimation of the RC(M) association model. *Applied Statistics, 39*, 152–167.

Clogg, C.C. (1982). Some models for the analysis of association in multi-way cross-classifications having ordered categories. *Journal of the American Statistical Association, 77*, 803–815.

Clogg, C.C., & Shihadeh, E.S. (1994). *Statistical models for ordinal data*. Thousand Oaks, CA: Sage Publications.

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

Dardanoni, V. & Forcina A. (1998). A unified approach to likelihood inference on 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, Series B., 39*, 1–38.

Efron, B., & Tibshirani, R.J. (1993). *An introduction to the Bootstrap*. London: Chapman and Hall.

Geyer, C.J. (1995). Likelihood ratio tests and inequality constraints. Technical Report No. 610, University of Minnesota.

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

Goodman, L.A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association, 74*, 537–552.

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.

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.

Ritov, Y., & Gilula, Z. (1991). The order-restricted RC model for ordered contingency tables: Estimation and testing for fit. *Annals of Statistics, 19*, 2090–2101.

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.

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

Schoenberg, R. (1997). CML: Constrained maximum likelihood estimation. *The Sociological Methodologist, Spring 1997*, 1–8.

Shapiro, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika, 72*, 133–144.

Vermunt, J.K. (1999). A general nonparametric approach to the analysis of ordinal categorical data. *Sociological Methodology, 29*, 197–221.

Vermunt, J.K. (2001). The use restricted latent class models for defining and testing nonparametric and parametric item response theory models. *Applied Psychological Measurement, 25*, 283–294.

Wang, Y. (1996). A likelihood ratio test against stochastic ordering in several populations. *Journal of the American Statistical Association, 91*, 1676–1683.

Manuscript received 30 NOV 1998
Final version received 21 JAN 2004