Order-Restricted Score Parameters in Association Models for Contingency Tables

ALAN AGRESTI, CHRISTY CHUANG, and ABBAS KEZOUH*

The row effects and column effects models for two-way contingency tables have parameters for the row and column categories pertaining to the association between the variables. For classifications having ordered categories, it is often reasonable to assume that the association parameters have a corresponding ordering. This article proposes order-restricted estimates of the association parameters in these models. The maximum likelihood solution can be determined by the solution of a simple isotonic regression of some of the model sufficient statistics. The models are primarily log-linear in form and can be expressed in terms of odds ratios for 2 × 2 subtables consisting of adjacent rows and adjacent columns. For the order-restricted solution, these local log-odds ratios have uniform sign. Goodness-of-fit statistics for this solution are related to corresponding statistics for collapsed tables and to statistics for testing equality of sets of the parameters.

The row effects model discussed in this article has been proposed by Haberman (1974), Simon (1974), and Goodman (1979), among others. This model contains parameters for the rows in the contingency table that describe the structure of the association and can be used to describe dependence in corresponding logit models. This article deals with applications of the model in which there is a monotonic relationship between the variables, in the sense that the population values of the local log-odds ratios are uniformly nonnegative or uniformly nonpositive. For instance, one might expect a nonnegative relationship for the data analyzed by Haberman (1974) and by Goodman (1979) on mental health and socioeconomic status, and a nonpositive relationship for the data analyzed in Section 2 of this article on age and severity of disturbances in dreams. By using the methods described in this article, one can obtain monotone estimates of the association parameters, which imply a monotone relationship between the variables. With this approach, one obtains a simpler description of the relationship, and better estimates, when the parameter scores truly are ordered.

KEY WORDS: Collapsed tables; Isotonic regression; Log-linear model; Odds ratio; Ordinal variable; Row effects model.

1. INTRODUCTION

Let $\{n_{ij}\}$ denote the cell frequencies in an $r \times c$ cross-classification of ordinal variables X and Y. Let $\{m_{ij}\}$ denote the corresponding expected frequencies. Haberman (1974), Simon (1974), Goodman (1979), and others modeled the association between X and Y by using the structural form

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \mu_i \nu_j, \qquad (1.1)$$

for which the local log-odds ratio equals

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

This model is referred to as the *row effects* model when the $\{\mu_i\}$ are unknown parameters and the $\{\nu_j\}$ are fixed, strictly monotone scores; the *column effects* model when the $\{\nu_j\}$ are parameters and the $\{\mu_i\}$ are fixed, strictly monotone scores; the multiplicative *row-and-column effects* model when both sets are parameters; and the *linear-by-linear association* model when $\mu_i \nu_j = \beta u_i v_j$ with the $\{u_i\}$ and $\{v_j\}$ being fixed, strictly monotone scores. Goodman (1979; 1985) showed that model (1.1) is a discrete analog for a family of distributions that includes the bivariate normal, and gave examples of its use. Clogg (1982) used the $\{\mu_i\}$ and $\{\nu_j\}$ in this model as category scores in scaling variables assumed to have such an underlying distribution. When researchers use structural form (1.1), their initial choice of model may depend on whether the classifications have natural scorings. Any classification having parameter scores is treated as nominal, in the sense that there is no inherent order to the $\{\mu_i\}$ or $\{\nu_j\}$, and their maximum likelihood estimates can have any order.

In many applications, it is reasonable to assume that the association is monotonic in the sense of uniformly nonnegative, or uniformly nonpositive, local log-odds ratios. For model (1.1) this implies that the scores have the same, or the reverse, monotone ordering as the categories. In this article we show that when an ordering constraint is imposed on the parameter scores in the row effects model or in the column effects model, maximum likelihood estimates can be determined from the ordinary fit for an appropriately collapsed table. We give necessary and sufficient conditions for this collapsing, and we show that it can be determined from an isotonic regression involving the model sufficient statistics. A goodness-of-fit statistic for the order-restricted solution is shown to be related to corresponding statistics for collapsed tables and to a statistic for testing equality of sets of score parameters.

Model (1.1) treats the variables symmetrically. In applications where it is natural to treat one of the variables as a response, this model may be more informative when viewed as a logit model for adjacent response categories, as in Simon (1974) and in Goodman (1983). For instance, if the column variable is a response that is assigned scores $\{\nu_j = j\}$, then the row effects model corresponds to the logit model

$$\log(m_{i,j+1}/m_{ij}) = (\lambda_{j+1}^Y - \lambda_j^Y) + \mu_i = \alpha_j + \mu_i.$$

If the $\{\mu_i\}$ are regarded as treatment effects on the logit scale, then the order restriction implies that these effects have the same ordering as the levels of X.

2. ORDER-RESTRICTED ROW EFFECTS MODEL

For the row effects model, $\nu_1 < \cdots < \nu_c$ are fixed and the $\{\mu_i\}$ are parameters. Here, we fit this model subject to the order restriction $\mu_1 \leqslant \cdots \leqslant \mu_r$, corresponding to a

* Alan Agresti is Professor, Department of Statistics, University of Florida, Gainesville, FL 32611. Christy Chuang is Research Statistician, Biostatistics Unit, Upjohn Company, Kalamazoo, MI 49001. Abbas Kezouh is Statistician, Université des Sciences et de la Technologie Houari Boumedienne, Algiers, Algeria. This work was partially supported by National Institutes of Health Grant GM33210 for Agresti and National Cancer Institute Grant CA11198 for Chuang. The authors are grateful to the referees for several helpful comments.

Journal of the American Statistical Association
June 1987, Vol. 82, No. 398, Theory and Methods

nonnegative association. Analogous results apply to the solution for the order restriction $\mu_1 \geqslant \cdots \geqslant \mu_r$ or to order-restricted solutions for the column effects model. For identifiability, one can also impose a constraint such as $\sum \mu_i = 0$. Let $\{\hat{\mu}_i\}$ denote the ordinary maximum likelihood estimates, and let $\{\hat{\mu}_i^*\}$ denote the order-restricted estimates. The following result is shown in Appendix A.

Theorem 1. A necessary and sufficient set of equations for characterizing the order-restricted solution $\{\hat{m}_{ij}^*\}$ of the row effects model is

$$\hat{m}_{i+}^* = n_{i+}, \qquad i = 1, \ldots, r;$$

$$\hat{m}_{+j}^* = n_{+j}, \qquad j = 1, \ldots, c, \qquad (2.1)$$

where a plus (+) subscript indicates summation over that index,

$$\sum_{i \leqslant b} \left[\sum_j v_j \hat{m}_{ij}^* \right] \leqslant \sum_{i \leqslant b} \left[\sum_j v_j n_{ij} \right], \quad b = 1, \ldots, r, \quad (2.2)$$

with

$$\sum_{i \leqslant r_k} \sum_j v_j \hat{m}_{ij}^* = \sum_{i \leqslant r_k} \sum_j v_j n_{ij}, \qquad k = 1, \ldots, a, \quad (2.3)$$

where $\{r_1, \ldots, r_a\}$ are such that

$$\hat{\mu}_1^* = \cdots = \hat{\mu}_{r_1}^* < \hat{\mu}_{r_1+1}^* = \cdots = \hat{\mu}_{r_2}^* < \cdots < \hat{\mu}_{r_{a-1}+1}^* = \cdots = \hat{\mu}_{r_a}^*. \quad (2.4)$$

The likelihood equations for the ordinary solution of the model are Equations (2.1) and (2.3) with $r_k = k$ $(k = 1, \ldots, r)$. Let $R_k = \{r_{k-1} + 1, \ldots, r_k\}$ $(k = 1, \ldots, a)$, with $r_0 = 0$. Then Equations (2.1) and (2.3) imply the likelihood equations for the row effects model fitted to the collapsed table in which the rows in each of $R_1, \ldots, R_a$ are combined. The order-restricted solution is, therefore, the same as the solution for the model with equality constraints for the score parameters in each R_k, and this solution can be determined from the ordinary solution for an appropriately collapsed table. The new aspect here occurs in determining the partition $\{R_k\}$ for which these equations give the order-restricted solution. The following result, proved in Appendix B, is useful for this purpose.

Theorem 2. For the ordinary row effects model, the $\{\hat{\mu}_i\}$ have the same ordering as the sample row means $\{M_i = \sum_j v_j n_{ij}/n_{i+}\}$.

Hence the partition $\{R_k\}$ must be such that

$$\left[\sum_{i \in R_k} \left(\sum_j v_j n_{ij} \right) \Big/ \sum_{i \in R_k} n_{i+} \right]$$

is strictly increasing in k. This partition can easily be determined using the following result, proved in Appendix C.

Theorem 3. The partition $\{R_k\}$ for the order-restricted row effects model is identical to the partition of level sets obtained in minimizing $\sum (M_i - M_i^*)^2 n_{i+}$ in the class of functions isotonic with respect to the simple order of the rows, that is, such that $M_1^* \leqslant \cdots \leqslant M_r^*$.

In other words, the level sets are the same as those obtained in the isotonic regression of the sample row means with respect to the simple order on the rows, using the row marginal totals as the weights. For the order restriction $\mu_1 \geqslant \cdots \geqslant \mu_r$, Theorem 3 applies to the class of functions isotonic with respect to the reverse of the simple order.

The correct partition $\{R_k\}$ can be determined with an algorithm used in isotonic regression, such as the pooling adjacent violators algorithm (see Barlow, Bartholomew, Bremner, and Brunk 1972). At each step in this algorithm, one compares each pair of rows i and $i + 1$, combining them and placing them in the same level set if the row means satisfy $M_i > M_{i+1}$. One continues to enlarge the level sets from violations in the ordering of row means until, at a particular stage, the row means are monotone increasing. The isotonic regression is unique (Barlow et al. 1972, p. 12), and the final solution does not depend on the order in which the algorithm is applied to the rows.

A useful consequence of Theorem 3 is that one can determine the partition $\{R_k\}$ before actually fitting the row effects model. Once the partition is obtained, standard software that can perform the iterative fitting of the row effects model (see Agresti 1984, app. D) can be used to obtain the order-restricted solution. Alternatively, because of the correspondence between the ordering of the row means and the estimated parameter scores (see Th. 2), one can discover the partition through successive fitting of the model, whereby, at each stage, scores are equated that violate the desired order. The order-restricted solution can also be obtained using a general-purpose algorithm for nonlinear optimization subject to linear inequality constraints. McDonald and Diamond (1983) described some of these algorithms. We have developed a program that incorporates the E04UAF subroutine from the NAG library (1984), which uses a sequential augmented Lagrangian method, the maximization being solved by a quasi-Newton method. We have also used the BMDP–3R (Dixon 1979) nonlinear regression program, by using one of its options to form G^2 as the recognized loss function. Copies of these programs are available from Dr. Chuang.

Equations (2.1)–(2.3) are implied by the likelihood equations for fitting the row effects model to a transformed table $\{n_{ij}^*\}$ in which the frequencies in R_k are replaced by expected frequency estimates for the independence model fitted to those rows $(k = 1, \ldots, a)$; that is,

$$n_{ij}^* = \left[\sum_{x \in R_k} n_{xj} \right] \left[n_{i+} \Big/ \sum_{x \in R_k} n_{x+} \right], \quad \text{if } i \text{ in } R_k.$$

Hence the order-restricted solution can be regarded as the composite fit in which the row effects model is applied after the independence model has been fitted to each set of rows R_k. We use this result in Appendix D to show that the goodness of fit of an order-restricted row effects model can be decomposed into the goodness of fit of independence models to the $\{R_k\}$ plus the goodness of fit of the row effects model to the collapsed table in which each set of rows R_k is combined into a single row. Specifically, let $G^2(R) = 2 \sum \sum n_{ij} \log(n_{ij}/\hat{m}_{ij})$ denote the likelihood ratio statistic for fitting the ordinary row effects model to the original table, let $G^2(R^*)$ denote the fit of the order-restricted model for that table, and let $G^2(R')$ denote the fit of the row effects model for the collapsed table. In addition, let $G^2(I)$ and $G^2(I')$ denote the fit of the independence model to the original and collapsed tables, respectively, and let $G^2(I_k)$ denote the fit of the independence model to the set of rows R_k $(k = 1, \ldots, a)$. Then, we have the following result.

Theorem 4.

$$G^2(R^*) = G^2(R') + \sum G^2(I_k)$$
$$= G^2(R') + G^2(I) - G^2(I').$$

We next consider briefly the asymptotic distributions of $G^2(R^*)$ and $\hat{\mu}^* = (\hat{\mu}_1^*, \ldots, \hat{\mu}_r^*)$, when the model holds with ordered score parameters. If $\mu_1 < \mu_2 < \cdots < \mu_r$, then under multinomial or Poisson sampling assumptions, $\Pr(\hat{\mu}_1 < \cdots < \hat{\mu}_r) \to 1$ as $n \to \infty$. Hence $\Pr(\hat{\mu}^* = \hat{\mu}) \to 1$ and $\Pr\{G^2(R^*) = G^2(R)\} \to 1$ as $n \to \infty$, so asymptotically the order-restricted fit and the ordinary fit are identical. In particular $\hat{\mu}^*$ has an asymptotic normal distribution with asymptotic covariance matrix obtained from the inverse of the information matrix for the ordinary model. Suppose next that the $\{\mu_i\}$ are monotonic, but that at least two scores are equal. For illustrative purposes, suppose that the row effects model holds with $\mu_1 < \cdots < \mu_k = \mu_{k+1} < \cdots < \mu_r$. Then, with limiting probability .5 as $n \to \infty$, $\hat{\mu} = \hat{\mu}^*$ and $G^2(R^*) = G^2(R)$, which has an asymptotic $\chi^2_{(r-1)(c-2)}$ distribution. In addition, with limiting probability .5, $\hat{\mu}_1 < \cdots < \hat{\mu}_k$, $\hat{\mu}_k > \hat{\mu}_{k+1}$, $\hat{\mu}_{k+1} < \cdots < \hat{\mu}_r$, and $\hat{\mu}_1^* < \cdots < \hat{\mu}_k^* = \hat{\mu}_{k+1}^* < \cdots < \hat{\mu}_r^*$, so $G^2(R^*) = G^2(R') + G^2(I_k)$. For a fixed selection of rows k and $k + 1$ for the collapsing, $G^2(R')$ and $G^2(I_k)$ are asymptotically independent $\chi^2_{(r-2)(c-2)}$ and χ^2_{c-1} random variables, respectively, so $G^2(R') + G^2(I_k)$ is an asymptotic $\chi^2_{(r-1)(c-2)+1}$ random variable. Hence for large n, $G^2(R^*)$ has an asymptotic distribution that is an equal mixture of $\chi^2_{(r-1)(c-2)}$ and $\chi^2_{(r-1)(c-2)+1}$ distributions. If more than two of the scores are equal, the asymptotic distribution is a more complex mixture of χ^2 distributions that will not be considered here. Suppose finally that some $\mu_k > \mu_{k+1}$. Then $\Pr(\hat{\mu}_k > \hat{\mu}_{k+1}) \to 1$, and $G^2(I_k)$ for the table involving rows k and $k + 1$, and similarly $G^2(R^*)$, is unbounded with probability 1 as $n \to \infty$.

To illustrate order restrictions in the row effects or column effects model, we consider the data in Table 1. The table, analyzed most recently by Anderson (1984), categorizes boys by age and by severity of disturbed dreams. We assigned midpoint scores (6, 8.5, 10.5, 12.5, 14.5) to the age variable. When we assume equal-interval scores

Table 1. Severity of Disturbances of Dreams in Boys, by Age: Observed Data, With Estimated Expected Frequencies (in parentheses) for Order-Restricted Column Effects Model

Age	Not severe 1	2	3	Very severe 4
5–7	7 (4.06)	4 (5.15)	3 (5.02)	7 (6.78)
8–9	10 (14.25)	15 (11.27)	11 (11.01)	13 (12.47)
10–11	23 (20.55)	9 (10.14)	11 (9.90)	7 (9.42)
12–13	28 (31.92)	9 (9.83)	12 (9.59)	10 (7.66)
14–15	32 (29.23)	5 (5.62)	4 (5.48)	3 (3.68)

for severity of disturbed dreams, the linear-by-linear association model gives $G^2 = 14.61$ with 11 residual df, a large improvement over $G^2 = 32.46$ with 12 df given by the independence model. In the column effects model we treat the severity scores as parameters and obtain a somewhat better fit, with $G^2(C) = 9.75$ and 9 residual df. The estimated parameters $\hat{\nu} = (.189, -.061, -.008, -.120)$ suggest a negative trend in the association, though $\hat{\nu}_2 < \hat{\nu}_3$.

Anderson fitted the column effects model to these data. Noting the closeness of $\hat{\nu}_2$, $\hat{\nu}_3$, and $\hat{\nu}_4$, he suggested using a simpler model in which these scores are constrained to be equal. We could, instead, assume that $\nu_1 \geqslant \cdots \geqslant \nu_4$ and conjecture that $\hat{\nu}_2 < \hat{\nu}_3$ is simply due to sampling error. In other words, given that the column effects model holds, we could assume that boys of higher ages are stochastically lower on severity of disturbed dreams. This order-restricted column effects model gives $\hat{\nu}^* = (.189, -.034, -.034, -.120)$ for Table 1. The $\{\hat{m}_{ij}^*\}$ are also shown in the table. This order-restricted fit has $G^2(C^*) = 10.13$, very nearly as good as the unrestricted fit.

3. COMMENTS

Goodman (1985) presented tests about equality of score parameters for model (1.1). His test statistics can be related to goodness-of-fit statistics for order-restricted solutions. For instance, to test $\nu_2 = \nu_3$ given that the column effects model holds, Goodman (1985) showed that one can use the statistic $T = G^2(I) - G^2(I') + G^2(C') - G^2(C)$, where I' and C' refer to the collapsed table in which columns two and three are combined. When the model holds with $\nu_2 = \nu_3$, T has an asymptotic chi-squared distribution with df = 1. Now if an order-restricted fit gives equality only to $\hat{\nu}_2^*$ and $\hat{\nu}_3^*$, then $G^2(C^*) = G^2(I) - G^2(I') + G^2(C') = G^2(C) + T$. For instance, to test $H_0: \nu_2 = \nu_3$ for Table 1, Goodman's statistic is $T = .38$, based on df = 1. This is identical to $G^2(C^*) - G^2(C)$ for that table. The statistic T should here be regarded only as an informal index, since H_0 was suggested by the data. Since $\hat{\nu}_2 < \hat{\nu}_3$ was the only out-of-order pair, this statistic supports our conjecture that this violation may simply reflect sampling error. When there is substantial evidence that the true scores are out of order, so that some true local log-odds ratios are positive and some are negative, then the ordinary fit is useful for describing the nonmonotonicity. In many applications, though, such as when we posit a bivariate normal distribution for underlying continuous variables, it is quite reasonable to assume a uniform sign for the local log-odds ratios. Then, if there is not strong evidence of nonmonotonicity, an order-restricted solution is appealing because simpler interpretations follow from it.

If the row effects (or column effects) model holds and the true scores are strictly monotonic, then the order-restricted estimates and the ordinary estimates can have quite different small-sample distributions. The ordinary estimates are less likely to be monotonic as n or the strength of association decreases. A simulation study by Kezouh (1984) indicates that order-restricted estimates can have much smaller mean squared error than the ordinary estimates when the true scores are monotonic. Such behavior is consistent with the improved performance obtained with isotonic regression methods in other contexts (see, e.g., Lee 1981). In addition, although the order-restricted solution produces groups of row scores that are equal, the purpose of this procedure is not necessarily that of collapsing the table to obtain simpler scales. For instance, if $\hat{\mu}_k > \hat{\mu}_{k+1}$ and $\hat{\mu}_k^* = \hat{\mu}_{k+1}^*$, one would often expect larger samples to produce $\hat{\mu}_k < \hat{\mu}_{k+1}$ and a properly ordered scale.

It would be useful in some applications to fit the model having both row and column effects, subject to order restrictions such as $\mu_1 \leqslant \mu_2 \leqslant \cdots \leqslant \mu_r$ and $\nu_1 \leqslant \nu_2 \leqslant \cdots \leqslant \nu_c$. By using the same argument as in Appendix A, one can show that an order-restricted fit necessarily satisfies equations that correspond to likelihood equations for fitting the ordinary row-and-column effects model to a collapsed table. Based on the same arguments given in Appendix D, the goodness of fit of the order-restricted model can be expressed in terms of the fit to the collapsed table by $G^2(RC^*) = G^2(RC') + G^2(I) - G^2(I')$. To obtain an order-restricted solution, one can use a nonlinear optimization program. However, we have been unable to obtain a sufficient condition for the solution. The difficulty here is that the log-likelihood is not necessarily concave, since the row-and-column effects model is not log-linear. If the model holds and there is dependence, then the likelihood function is locally concave around the global maximum when the sample size is large, since the information matrix at the true parameter values is positive definite. Hence a judicious choice of initial trial values improves the chance of obtaining proper convergence. The discovery of sufficient conditions for the solution of the ordinary or order-restricted row-and-column effects model, plus the development of an algorithm for fitting these models, are important problems for future research.

APPENDIX A: PROOF OF THEOREM 1

If the $\{n_{ij}\}$ are independent Poisson random variables with means $\{m_{ij}\}$, then the log-likelihood is

$$L = \sum\sum n_{ij}\log m_{ij} - \sum\sum m_{ij} = \mu n + \sum n_{i+}\lambda_i^X + \sum n_{+j}\lambda_j^Y + \sum \mu_i \sum \nu_j n_{ij} - \sum\sum \exp(\mu + \lambda_i^X + \lambda_j^Y + \mu_i\nu_j).$$

To maximize a function subject to constraint functions, one can refer to the Kuhn–Tucker conditions. These are discussed, with particular attention to convex or concave functions, in Mangasarian (1969, chap. 7). In the context of maximizing L subject to the linear inequality constraints $\mu_i \leqslant \mu_{i+1}$ $(i = 1, \ldots, r-1)$, these conditions are as follows:

1. $\hat{\mu}_1^* \leqslant \cdots \leqslant \hat{\mu}_r^*$.
2. $\nabla L + (\nabla g)'\eta = 0$, where ∇L is the column vector of partial derivatives of L taken with respect to the model parameters, ∇g is a matrix whose jth row contains the partial derivatives of the jth constraint function g_j with respect to the model parameters, where $g_j = \mu_j^* - \mu_{j+1}^* \leqslant 0$ $(j = 1, \ldots, r-1)$ and η is a column vector of Lagrange multipliers satisfying conditions 3 and 4.
3. $\eta_i \geqslant 0$ $(i = 1, \ldots, r-1)$.
4. $g'\eta = \sum_{i=1}^{r-1}\eta_i(\hat{\mu}_i^* - \hat{\mu}_{i+1}^*) = 0$.

The second of these conditions corresponds to the equations

$$\hat{m}_{i+}^* = n_{i+}, \quad i = 1, \ldots, r;$$

$$\hat{m}_{+j}^* = n_{+j}, \quad j = 1, \ldots, c;$$

$$\sum_j \nu_j\hat{m}_{ij}^* - \sum_j \nu_j n_{ij} + \eta_i - \eta_{i-1} = 0, \quad i = 1, \ldots, r,$$

where $\eta_0 = \eta_r = 0$. The nonnegativity of $\{\eta_i\}$, therefore, is equivalent to Equation (2.2). Because of the constraints on $\{\hat{\mu}_i^*\}$ and $\{\eta_i\}$, the fourth condition is equivalent to the condition that $\eta_i = 0$ whenever $\hat{\mu}_i^* < \hat{\mu}_{i+1}^*$. Thus if the partition for the solution is as shown in (2.4), it follows that the fourth condition is equivalent to (2.3).

The Kuhn–Tucker conditions are necessarily satisfied by a solution to the order-restricted problem (Mangasarian 1969, pp. 103 and 105–106). The row effects model is log-linear, so L is concave, from which it follows that these are also sufficient conditions (Mangasarian 1969, p. 94).

APPENDIX B: PROOF OF THEOREM 2

For any two rows a and b, the ordinary maximum likelihood estimates satisfy

$$\log(\hat{m}_{aj}\hat{m}_{b,j+1}/\hat{m}_{a,j+1}\hat{m}_{bj}) = (\nu_{j+1} - \nu_j)(\hat{\mu}_b - \hat{\mu}_a), \quad j = 1, \ldots, c-1.$$

Since $\nu_j < \nu_{j+1}$ for all j, these $c - 1$ local log-odds ratios all have the same sign as $\hat{\mu}_b - \hat{\mu}_a$. It follows (see Agresti 1984, p. 22) that

$$\log\left[\left(\sum_{k=1}^{j}\hat{m}_{ak}\right)\left(\sum_{k=j+1}^{c}\hat{m}_{bk}\right)\Big/\left(\sum_{k=j+1}^{c}\hat{m}_{ak}\right)\left(\sum_{k=1}^{j}\hat{m}_{bk}\right)\right], \quad j = 1, \ldots, c-1$$

have the same sign as $\hat{\mu}_b - \hat{\mu}_a$. The sign of this log-odds ratio determines a stochastic ordering of the conditional distributions $\{\hat{m}_{aj}/\hat{m}_{a+}, j = 1, \ldots, c\}$ and $\{\hat{m}_{bj}/\hat{m}_{b+}, j = 1, \ldots, c\}$. Thus the conditional distributions in the rows are stochastically ordered according to the values of the $\{\hat{\mu}_i\}$. It follows that the $\{\hat{\mu}_i\}$ have the same ordering as the fitted row means $\{\sum_j \nu_j\hat{m}_{ij}/\hat{m}_{i+}, i = 1, \ldots, r\}$. The likelihood equations $\{\hat{m}_{i+} = n_{i+}, i = 1,$

$\ldots, r\}$ and $\{\sum_j v_j \hat{m}_{ij} = \sum_j v_j n_{ij}, i = 1, \ldots, r\}$ imply that the $\{\hat{\mu}_i\}$ also have the same ordering as the sample row means $\{M_i\}$.

APPENDIX C: PROOF OF THEOREM 3

Consider the problem of minimizing $\sum (M_i - M_i^*)^2 n_{i+}$ subject to $M_1^* \leq \cdots \leq M_r^*$. The Kuhn–Tucker conditions for the solution $(\hat{M}_1^*, \ldots, \hat{M}_r^*)$ are as follows:

1. $\hat{M}_1^* \leq \cdots \leq \hat{M}_r^*$.
2. $(\hat{M}_i^* - M_i)n_{i+} + \eta_i - \eta_{i-1} = 0$ $(i = 1, \ldots, r)$, where $\eta_0 = \eta_r = 0$ and where the $\{\eta_i\}$ satisfy conditions 3 and 4.
3. $\eta_i \geq 0$ $(i = 1, \ldots, r - 1)$.
4. $\sum_{i=1}^{r-1} \eta_i(\hat{M}_i^* - \hat{M}_{i+1}^*) = 0$.

If $\{\hat{M}_i^*\}$ is constant on level sets $S_k = \{s_{k-1} + 1, \ldots, s_k\}$ $(k = 1, \ldots, a)$, then conditions 2–4 correspond to equations

$$\sum_{i \leq b} M_i n_{i+} \geq \sum_{i \leq b} \hat{M}_i^* n_{i+}, \qquad b = 1, \ldots, r \tag{C.1}$$

$$\sum_{i \leq s_k} M_i n_{i+} = \sum_{i \leq s_k} \hat{M}_i^* n_{i+}, \qquad k = 1, \ldots, a. \tag{C.2}$$

Having obtained the $\{\hat{M}_i^*\}$ using an algorithm for isotonic regression, suppose that one obtains the fit of the row effects model corresponding to the likelihood equations

$$\hat{m}_{i+}^* = n_{i+}, \quad i = 1, \ldots, r;$$

$$\hat{m}_{+j}^* = n_{+j}, \quad j = 1, \ldots, c;$$

$$\sum_j v_j \hat{m}_{ij}^* = \hat{M}_i^* n_{i+}, \quad i = 1, \ldots, r.$$

Since $M_i n_{i+} = \sum_j v_j n_{ij}$, it follows from (C.1) and (C.2) and from Theorem 2 that this solution satisfies (2.1)–(2.4), with $R_k \equiv S_k$ $(k = 1, \ldots, a)$. Hence, from Theorem 1, this solution is the order-restricted one for the row effects model.

APPENDIX D: PROOF OF THEOREM 4

Let $G^2(R_T) = 2 \sum \sum n_{ij}^* \log(n_{ij}^*/\hat{m}_{ij}^*)$, the goodness of fit of the row effects model to the transformed table described in Section 2. The collapsing of the original table in which all of the rows in R_k are combined into a single row $(k = 1, \ldots, a)$ is identical to the corresponding collapsing of the transformed table. For any table, when two rows are combined that have identical sample conditional distributions across the columns, the row effects model gives the same fit (i.e., the same $\hat{m}_{ij}$) for all cells in remaining rows as it does when the rows are separate. Hence it follows that $G^2(R_T) = G^2(R')$. In addition, since the $\{n_{ij}^*\}$ in R_k factor in terms of the product of row and column totals of R_k, and since those totals are the same for the $\{n_{ij}\}$ as the $\{n_{ij}^*\}$, it follows that $\sum \sum n_{ij} \log(n_{ij}^*) = \sum \sum n_{ij}^* \log(n_{ij}^*)$. The $\{\hat{\mu}_i^*\}$ are identical within each R_k. Hence the $\{\hat{m}_{ij}^*\}$ also can be expressed as independence estimates in each R_k, and thus

$$\sum \sum n_{ij} \log(\hat{m}_{ij}^*) = \sum \sum n_{ij}^* \log(\hat{m}_{ij}^*).$$

It follows that

$$\begin{aligned} G^2(R^*) &= 2 \sum \sum n_{ij} \log(n_{ij}/n_{ij}^*) \\ &\quad + 2 \sum \sum n_{ij} \log(n_{ij}^*/\hat{m}_{ij}^*) \\ &= \sum G^2(I_k) + 2 \sum \sum n_{ij}^* \log(n_{ij}^*/\hat{m}_{ij}^*) \\ &= \sum G^2(I_k) + G^2(R'). \end{aligned}$$

From standard chi-squared partitioning arguments, $\sum G^2(I_k) + G^2(I') = G^2(I)$, from which $G^2(R^*) = G^2(R') + G^2(I) - G^2(I')$.

[*Received March 1984. Revised October 1986.*]

REFERENCES

Agresti, A. (1984), *Analysis of Ordinal Categorical Data,* New York: John Wiley.

Anderson, J. A. (1984), "Regression and Ordered Categorical Variables," *Journal of the Royal Statistical Society*, Ser. B, 46, 1–22.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972), *Statistical Inference Under Order Restrictions,* New York: John Wiley.

Clogg, C. C. (1982), "Using Association Models in Sociological Literature: Some Examples," *American Journal of Sociology,* 88, 114–134.

Dixon, W. J. (ed.) (1979), *BMDP Statistical Software,* Berkeley, CA: University of California 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.

——— (1983), "The Analysis of Dependence in Cross-Classifications Having Ordered Categories, Using Log-Linear Models for Frequencies and Log-Linear Models for Odds," *Biometrics,* 39, 149–160.

——— (1985), "The Analysis of Cross-Classified Data Having Ordered and/or Unordered Categories: Association Models, Correlation Models, and Asymmetry Models for Contingency Tables With or Without Missing Entries," *The Annals of Statistics,* 13, 10–69.

Haberman, S. J. (1974), "Loglinear Models for Frequency Tables With Ordered Classifications," *Biometrics,* 30, 589–600.

Kezouh, A. (1984), "Association Models for Cross Classifications Having Ordered Categories," unpublished Ph.D. thesis, University of Florida, Dept. of Statistics.

Lee, C. C. (1981), "The Quadratic Loss of Isotonic Regression Under Normality," *The Annals of Statistics,* 9, 986–988.

Mangasarian, O. L. (1969), *Nonlinear Programming,* New York: McGraw-Hill.

McDonald, J. W., and Diamond, I. (1983), "Fitting Generalised Linear Models With Linear Inequality Constraints," *GLIM Newsletter,* June, 29–36.

Numerical Algorithms Group Limited (NAG) (1984), *NAG Fortran Mini-Manual, Mark 11,* Oxford, U.K.: Author.

Simon, G. (1974), "Alternative Analyses for the Singly-Ordered Contingency Table," *Journal of the American Statistical Association,* 69, 971–976.