ELSEVIER

Journal of Statistical Planning and Inference 107 (2002) 45–73

journal of statistical planning and inference

www.elsevier.com/locate/jspi

The analysis of contingency tables under inequality constraints

Alan Agresti[a,*], Brent A. Coull[b]

[a] *Department of Statistics, University of Florida, 204 Griffin-Floyd Hall, Box 118545, Gainesville, FL, 32611-8545, USA*

[b] *Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115, USA*

Abstract

This article surveys ways of taking order restrictions into account in the analysis of contingency tables. The methods are best suited for tables in which at least one classification is ordinal. Primary topics include (1) inequality constraints for parameters for a set of independent binomial random variables, (2) inequality constraints for odds ratios, (3) inequality constraints for parameters in loglinear or logit models, and (4) structured terms such as linear trends in loglinear and logit models. We discuss methods for contingency tables with two rows or two columns that result from comparing several binomial parameters or two multinomial distributions, two-way contingency tables with arbitrary numbers of rows and columns, multi-way contingency tables in which two-way analyses are relevant within levels of a stratification factor, and square contingency tables having the same categories in each dimension. We discuss the advantages and disadvantages of various order-restricted approaches and highlight possible problems for future research.

1. Introduction

In recent years statisticians have increasingly recognized that many benefits can result from using methods that take into account orderings among categories of classifications in contingency tables. The benefits include the potential for considerable improvements in power over methods that ignore the ordinal information and greater parsimony in model-building from a reduction in the number of parameters needed to describe association structure. In particular, ordinal models can represent predictions about monotone or linear trends in effects and associations.

* Corresponding author. Tel.: +1-352-392-1941; fax: +1-352-392-5175.
E-mail address: aa@stat.ufl.edu (A. Agresti).

PII: S0378-3758(02)00243-4

One way to utilize ordered categories is to assume inequality constraints on parameters for those categories that describe association structure. This article surveys order-restricted descriptive and inferential analyses of contingency tables that are based on such constraints. These methods are based on restrictions such as the monotone increase of a set of binomial parameters, uniformly positive (or uniformly negative) log odds ratios between two categorical variables, association structure using monotonically ordered parameters in loglinear and logit models, order restrictions under stratification for a control variable, and order restrictions for comparing margins of multivariate categorical responses. In all cases, the methods relate to inequality constraints for a set of probabilities, odds ratios, or model parameters.

Section 2 introduces analyses for $r \times 2$ contingency tables that compare binomial parameters at r ordered levels of a factor. Section 3 presents order-restricted methods for $r \times c$ contingency tables, such as in comparing several multinomial distributions with ordered categories. These methods are specified in terms of inequality constraints on various odds ratios. The binomial methods of Section 2 have connections with those for the $2 \times c$ multinomial setting. Section 4 discusses methods for $r \times c$ contingency tables that are specified in terms of inequality constraints on parameters in various parametric models. Section 5 presents extensions to multi-way contingency tables and to comparing marginal distributions of multivariate categorical responses. Throughout, the article highlights problems for future research. Section 6 focuses on this and discusses reasons that inequality-constrained methods are not more prominent in the toolbox used by most data analysts.

Although our focus is on inequality-constrained methods for the contingency table format, we mention in passing that a large literature also exists on such methods for probabilities $(\pi_1, \pi_2, \ldots, \pi_c)$ in a single c-category multinomial distribution or for a set of Poisson means. For instance, Dykstra and Lee (1991) showed that for $\{\pi_j\}$ that are isotonic (order-preserving) with respect to the index set, maximum likelihood (ML) and various other estimators are related and can be expressed in terms of least squares projections, El Barmi and Dykstra (1994, 1996) characterized the ML estimates under the restriction that $\{\pi_j\}$ lie within a convex set, and Bhattacharya (1997) considered other cases. Chacko (1966) provided a test of H_0: $\pi_1 = \cdots = \pi_c$ against H_1: $\pi_1 \leqslant \cdots \leqslant \pi_c$ using a chi-squared-type statistic that approximates the likelihood-ratio (LR) statistic, and Bhattacharya (1995) considered this ordering in the context of testing for and against an increasing failure rate. Using LR tests, Robertson (1978) extended Chacko's results to testing for and against an arbitrary ordered alternative, and Dykstra and Robertson (1982a) considered testing for and against the "star-shaped" alternative

$$H_1\colon \pi_1 \geqslant \frac{\pi_1 + \pi_2}{2} \geqslant \cdots \geqslant \frac{\pi_1 + \cdots + \pi_{c-1}}{c-1} \geqslant \frac{1}{c}.$$

These tests all yield test statistics having an asymptotic chi-bar-squared distribution, the form of which is described in Section 2.1. Lee (1987a) reported that $n \geqslant 5c$ is sufficient for this approximation when testing $H_0: \ \pi_1 = \cdots = \pi_c$ against an order restriction but that slightly larger sample sizes are required when testing an order restriction against the general alternative. Regarding ML fitting algorithms having usefulness that extends to contingency tables, Dykstra and Lemke (1988) considered the setting in which log probabilities are in a closed, convex cone, and more generally, El Barmi and Dykstra (1998) incorporated both general convexity constraints on the multinomial probabilities and translated cone constraints on the log of the probability vector. Other relevant references include Weisberg (1972), Lee (1977, 1980), Baras (1983a, b), Greenberg (1985), Schervish (1985), and Sedransk et al. (1985).

2. Inequality-constrained methods for $r \times 2$ contingency tables

We discuss first the case of a binary response variable Y with a single predictor X having r ordered levels. Throughout this article, the columns refer to levels of the response variable, so in this case the data form a $r \times 2$ contingency table. Often, it is reasonable to predict that $P(Y=1 \mid X=x)$, the probability of a "success" at level x of the predictor, increases as x increases. Sometimes a set of ordered scores $x_1 < x_2 < \cdots < x_r$ naturally apply to the levels of X, such as in dose-response studies with the ordered levels refering to doses or log doses of the drug. To describe the predicted trend, one might construct a generalized linear model in which X has a linear effect, on some scale. For instance, one might use the linear logit model

$$\text{logit}[P(Y=1 \mid X=x_i)] = \alpha + \beta x_i, \ i=1,\ldots,r \tag{1}$$

with the expectation that $\beta > 0$. Alternatively, one might treat the explanatory variable as a factor but impose a monotonicity constraint on effects of factor levels; that is,

$$P(Y=1 \mid X=x_i) = \alpha + \beta_i, \tag{2}$$

with $\beta_1 \leqslant \beta_2 \leqslant \cdots \leqslant \beta_r$. This article discusses both types of ordinal approach, linear trends and inequality constraints, with main emphasis on the latter since it seems less fully developed and discussed in other survey papers and in textbooks.

Let n_{i1} denote the number of successes out of n_i trials at level x_i of X, and let $n_{i2}=n_i-n_{i1}$ and $n=\sum_i n_i$. We assume that $\{n_{i1},\ i=1,\ldots,r\}$ are independent binomial variates with parameters $\{\pi_i=P(Y=1 \mid X=x_i),\ i=1,\ldots,r\}$. Let $p_i=n_{i1}/n_i,\ i=1,\ldots,r$, denote the sample proportions. We first consider model (2), postponing discussion of model (1) to Section 4, which presents more general logit and loglinear models.

2.1. Inequality constraints on several binomial parameters

Bartholomew (1959) presented one of the first tests for contingency tables employing inequality constraints. For the several binomial problem, he proposed a test of H_0: $\pi_1 = \pi_2 = \cdots = \pi_r$ against the simple order H_a: $\pi_1 \leqslant \pi_2 \leqslant \cdots \leqslant \pi_r$; equivalently, in model (2), H_0: $\beta_1 = \cdots = \beta_r$ and H_a: $\beta_1 \leqslant \beta_2 \leqslant \cdots \leqslant \beta_r$. Under H_0, the ML estimator of π_i is the overall sample proportion $p = (\sum_i n_{i1})/n$. If $p_1 \leqslant p_2 \leqslant \cdots \leqslant p_r$, then the order-restricted ML estimators $\{\hat{\pi}_i\}$ of $\{\pi_i\}$ are $\{\hat{\pi}_i = p_i\}$. Otherwise, one uses the pool adjacent violators algorithm (Ayer et al., 1955; Robertson et al., 1988, Section 1.2) to pool "out-of-order" categories for which $p_i > p_{i+1}$ until the resulting sample proportions are monotone increasing. The order-restricted ML estimates $\{\hat{\pi}_i\}$ for the original categories are the sample proportions for this partition of categories.

Bartholomew's test statistic equals the usual Pearson chi-squared statistic for testing independence, applied to the counts in the collapsed table that combines rows having sample proportions falling out of order. The large-sample null distribution of Bartholomew's test statistic, however, is not the usual chi-squared distribution. Like most other tests in inequality-constrained inference for categorical data, the asymptotic distribution is *chi-bar-squared*. This distribution refers to a mixture of independent chi-squared random variables of form $\sum_{d=1}^{r} \rho_d \chi^2_{d-1}$, where χ^2_d is a chi-squared variate with d degrees of freedom (with $\chi^2_0 \equiv 0$), and $\{\rho_d\}$ are a set of probabilities. For a test statistic T with observed value t, the P-value for this distribution equals

$$P(T > t) = \sum_{d=1}^{r} \rho_d P(\chi^2_{d-1} > t).$$

For applications of the chi-bar-squared distribution to inequality constrained problems with a set of parameters, ρ_d is the null probability that the inequality-constrained estimators consist of d distinct sets on which these estimates are level (either because of pooling or by chance). The $\{\rho_d\}$ are then called "level probabilities." Robertson et al. (1988, pp. 74–82) discussed calculation of $\{\rho_d\}$ for the case of independent samples from normal populations with common means but different variances. They then depend on the relative sizes of the r variances of sample means, and they equal a sum of products of multivariate normal orthant probabilities. Unfortunately, even in the normal case, $\{\rho_d\}$ do not have closed form except in certain simple cases such as equal variances of the sample means. Robertson et al. provided a general recursive algorithm for their calculation in the normal case. The level probabilities needed for the order-restricted binomial problem and others discussed in this article for contingency tables are traditionally approximated by the corresponding ones for the normal case, with their adequacy for large-sample distribution theory being based on the approximate normality of the sample proportions. The Robertson et al. text provided tables of critical values for the chi-bar-squared distribution for $r=3$ and 4 (pp. 411–413) and for larger r values for the equal sample size case (pp. 416, 418).

Barlow et al. (1972, p. 193), Shi (1991), and Oluyede (1994b) presented related large-sample tests comparing binomial parameters. The former test is based on the approximate normality of the sample proportions, using tests for inequality constraints on normal means and an inverse sine transformation to stabilize the variance. This would not seem preferable to Bartholomew's test, since it also relies on large-sample approximations (the delta method and an asymptotic chi-bar-squared distribution) and since it requires modification when $n_{i1}=0$ or $n_{i1}=n_i$. In fact, Kulatunga et al. (1996) noted that it is too liberal, tending to have actual size considerably larger than the nominal level. Ramsey (1972) took a Bayesian approach to the ordered binomial problem by placing a Dirichlet prior on the successive probability differences, and Sackrowitz (1982) recommended an adjusted estimate for situations when the ordinary ML estimate is inadmissible under squared error loss. Robertson et al. (1988, p. 167) presented the LR statistic for the inequality-constrained binomial problem as a special case of a LR test comparing parameters for independent samples from a single-parameter exponential family distribution (Robertson and Wegman, 1978). The LR statistic for testing independence (I) in the $r \times 2$ table, assuming the order-restricted (O) model (2), is

$$G^2(I|O) = \sum n_{i1} \log(\hat{\pi}_i/p) + \sum n_{i2} \log[(1-\hat{\pi}_i)/(1-p)].$$

This has the same general form as for the unrestricted case; the main difference is that the expected values are computed under the restricted hypothesis. Again, the limiting distribution is chi-bar-squared.

Robertson and Wright (1985) established the asymptotic chi-bar-squared distribution for the binomial LR test under the partly specified ordering $\pi_i \geqslant \pi_0$ for all i, where π_0 is the parameter under a control condition; see also Eeden (1956) and Dinh and Nguyen (1994) for ML estimation in this situation. In the exponential family setting, Tsai (1993) used Roy's union-intersection (UI) principle to test against a general restricted alternative. The resulting test statistic reduces to Rao's efficient score test when testing against the unrestricted alternative. Robertson et al. (1988, p. 164) also provided a LR statistic of chi-bar-squared form for testing the null hypothesis of an inequality constraint against the general alternative that at least two parameters are unequal.

The large-sample chi-bar-squared approximation for the distribution of LR statistics and related statistics such as the Pearson may be inadequate if some null expected cell counts are small or if a severe imbalance occurs in the binomial sample sizes. For testing equality of binomial parameters against the simple ordered alternative, Bennett (1962) presented a small-sample test for the equal sample size case, and Amundsen and Ljøgodt (1979) presented simpler tests applied to 2×2 tables obtained by ignoring or collapsing rows. Soms (1985) described test statistics expressible as linear combinations of the sample proportions, and Agresti and Coull (1996) presented tests using the LR statistic. Liu (1998) presented a score test for testing equality against an ordering that equates to the simple ordered alternative when the binomial parameters follow a nondecreasing order. All of these tests use the standard exact conditional approach of eliminating the unknown common value of $\{\pi_i = \pi\}$ under H_0 by conditioning on its sufficient statistic, $\sum_i n_{i1}$. The resulting hypergeometric distribution defined on the $r \times 2$ contingency table is free of unknown parameters, and it induces an exact distribution for the test statistic. One can rapidly simulate from this distribution to estimate as precisely as required the exact conditional P-value (Patefield, 1981). If n is small or if the data mainly clump at a small number of levels of X or mainly have outcomes at one of the levels of Y, the exact conditional distribution can be highly discrete. As in other discrete problems, this causes conservativeness. Using a mid-P value (the probability of more extreme test statistic values plus half the probability of the observed value; see Lancaster, 1961) ameliorates this problem, though it loses the guarantee that the actual size of the test is bounded above by the nominal size.

The inequality-constrained model (2) extends to multiple predictors. Gebhardt (1970), Dykstra and Robertson (1982a, b), Lee (1983), and Block et al. (1994) considered estimation for such models, and Eddy et al. (1995) and Fygenson (1997) modeled binary responses in this context. In practice, it might be sensible to treat some predictors with structured forms such as linear trends, some as qualitative factors, and some with inequality constraints on parameters. So far, there seems to be little literature on such models in the context of generalized linear modeling for discrete data. For two

Table 1
Responses from a clinical trial comparing four treatments on extent of trauma due to subarachnoid hemorrhage

Treatment group	Outcome				
	Death	Vegetative state	Major disability	Minor disability	Good recovery
Placebo	59	25	46	48	32
Low dose	48	21	44	47	30
Medium dose	44	14	54	64	31
High dose	43	4	49	58	41

Source of data: Dr. Christy Chuang-Stein, Pharmacia & Upjohn Company.

predictors and a binary response, Eddy et al. (1995) studied the isotonic model

$$\text{logit}[P(Y=1\,|\,x_1,x_2)] = f(x_1,x_2),$$

where f is monotone increasing in each argument. They noted, however, that no test yet exists for comparing this model to the linear logit special case $f(x_1,x_2)=\alpha+\beta_1x_1+\beta_2x_2$ with $\beta_1>0$ and $\beta_2>0$. For such an isotonic model, it would be useful also to have a mechanism for constructing a confidence interval for $P(Y=1\,|\,x_1,x_2)$ at fixed values of x_1 and x_2; presumably a bootstrap method could be used when the computations are feasible. It would also be useful to develop descriptive measures for summarizing the effects of the predictors in such models.

2.2. Example

Throughout this article, we illustrate order-restricted methods for contingency tables using Table 1, taken from Chuang-Stein and Agresti (1997). This 4×5 contingency table refers to a clinical trial regarding the outcome for patients who experienced trauma due to subarachnoid hemorrhage. The five response categories are ordered, ranging from ‘death’ to ‘good recovery.’ The table has four ordered treatment groups, corresponding to three dose levels of a medication and a vehicle infusion serving as a control group. (Due to company confidentiality requirements it is not possible to say more about the drug.) A study objective was to determine whether a more favorable outcome tends to occur as the dose increases.

To illustrate an inequality-constrained test for several binomials, we compare the four treatment groups in terms of whether they fall into response category 1 (death) or one of the other categories; that is, we collapse Table 1 into a 4×2 table. The sample proportions of death are (0.281, 0.253, 0.213, 0.221). The ML fitted probabilities under the restriction $\pi_1\geqslant\pi_2\geqslant\pi_3\geqslant\pi_4$ are (0.281, 0.253, 0.216, 0.216), where $0.216=(44+43)/(207+195)$. The LR statistic for testing equality against this ordered alternative equals 3.27. Using the recursive formulas for the level probabilities in Robertson et al. (1988, pp. 74–82), we obtain estimates (0.251, 0.459, 0.249, 0.041) of $\{\rho_d\}$. The large-sample chi-bar squared distribution (Robertson et al., 1988, p. 164) yields a

P-value of

$$P = 0.251(0) + 0.459P(\chi_1^2 \geqslant 3.27) + 0.249P(\chi_2^2 \geqslant 3.27)$$
$$+ 0.041P(\chi_3^2 \geqslant 3.27) = 0.0954.$$

Simulating from the exact conditional null distribution, we obtain $P = 0.0955$, with a 95% confidence interval for the true exact P of (0.0951, 0.0959). Here asymptotic and exact results are essentially identical, but this need not be true in general. For comparison, the chi-squared test of independence against the unrestricted alternative yields LR statistic 3.30 ($\text{df} = 3$) and $P = 0.35$.

By contrast, suppose we combine the first two response categories and separately combine the final three categories, and test for a monotone decrease in the probability of death or vegetative state. The sample proportions (0.400, 0.363, 0.280, 0.241) are then identical to the order-restricted fit, so the LR statistic is identical for the ordered and the unrestricted alternatives. The P-values are 0.0003 for both the order-restricted chi-bar-squared and exact tests (again, the P-values are different in general) and 0.002 for the unrestricted alternative.

3. Inequality-constrained methods for odds ratios in contingency tables

We next consider $r \times c$ contingency tables. We first suppose that both X and Y are ordinal. Denote the cell counts by $\{n_{ij}\}$, with total n, and let $n_{i+} = \sum_j n_{ij}$ and $n_{+j} = \sum_i n_{ij}$. Let

$$\pi_{j|i} = P(Y = j \mid X = i) \quad \text{and} \quad \gamma_{j|i} = P(Y \leqslant j \mid X = i)$$

for $i = 1, \ldots, r,\ j = 1, \ldots, c$. One can describe the association in this table using various types of odds ratios, with a set of $(r-1)(c-1)$ of them fully specifying the association structure. Possible sets of such odds ratios include:

1. *Cumulative* (or "local–global") odds ratios,
$$\theta_{ij} = \frac{\gamma_{j|i}/(1-\gamma_{j|i})}{\gamma_{j|i+1}/(1-\gamma_{j|i+1})},$$
2. *Local* odds ratios,
$$\theta_{ij} = \frac{\pi_{j+1|i+1}/\pi_{j|i+1}}{\pi_{j+1|i}/\pi_{j|i}},$$
3. *Global* odds ratios,
$$\theta_{ij} = \frac{P(Y \leqslant j|X \leqslant i)/P(Y > j|X \leqslant i)}{P(Y \leqslant j|X > i)/P(Y > j|X > i)},$$
4. *Continuation* odds ratios,
$$\theta_{ij} = \frac{P(Y = j|X = i)/P(Y > j|X = i)}{P(Y = j|X > i)/P(Y > j|X > i)}.$$

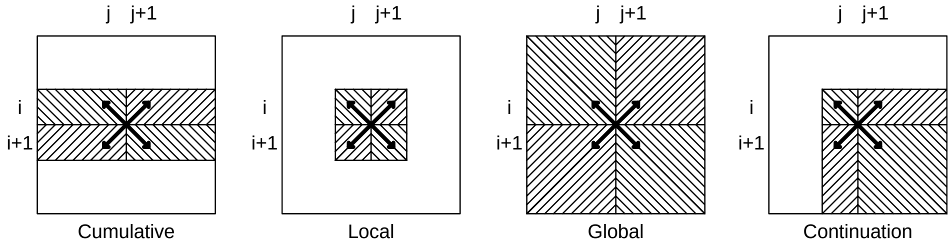

Fig. 1. Four sets of $(r-1)(c-1)$ odds ratios for ordinal variables: cumulative, local, global, and continuation.

Fig. 1 illustrates these four odds ratios, which are special cases of *generalized odds ratios* formed with four adjacent cells by collapsing row and column classifications into dichotomies. For other examples of such odds ratios and for a variety of results about relationships among them, see Lehmann (1966), Dale (1984), Grove (1984, 1986), Douglas et al. (1990), Barnhart and Sampson (1994), and Oluyede (1994a).

One can construct a separate and nonequivalent set of continuation odds ratios by applying that formula after reversing the category order for both variables. Either of these forms refer to a set of tables for which the usual LR statistic for testing independence partitions into the sum of $(r-1)(c-1)$ components, where each component is the LR statistic computed for a 2×2 table (Lancaster, 1949); this is not the case with the other odds ratios discussed. The local, global, and continuation odds ratios are invariant to switching the roles of X and Y. The cumulative odds ratios are not invariant. Because of their asymmetric form, they make most sense when Y alone is a response variable, whereas the others are also relevant when both variables are response variables. The conditional distribution of Y at level $i+1$ of X is *stochastically larger* than the one at level i of X if $\gamma_{j|i} \geqslant \gamma_{j|i+1}$ for all j. The condition of uniformly nonnegative log cumulative odds ratios is equivalent to Y being stochastically increasing in X.

For any particular set of odds ratios, one can define a positive association between X and Y by nonnegative values of all $(r-1)(c-1)$ log odds ratios. Denote this condition by L for the local odds ratios, G for the global odds ratios, S for the cumulative odds ratios (i.e., S for stochastic ordering), and C for the continuation odds ratios. Tables satisfying the L restriction are sometimes referred to as totally positive of order two (Douglas and Fienberg, 1990; Douglas et al., 1990). Of these conditions,

$$L \text{ implies } S \text{ implies } G \quad \text{and} \quad L \text{ implies } C \text{ implies } G.$$

One could conduct inequality-constrained inference for $r\times c$ contingency tables by expressing the inference in terms of one of these conditions. For instance, one could obtain ML estimates of cell probabilities subject to such a condition, and one could construct tests of independence against it. In applications that distinguish between response and explanatory variables, it is usually natural to treat the r rows as independent multinomial samples, treating $\{n_{i+}\}$ as fixed. In this setting, most work deals with comparing two groups $(r=2)$.

3.1. Comparisons of two multinomial distributions for various odds ratios

This section considers the special case, $r = 2$; that is, X is binary and the response variable Y has c ordered categories. Such tables often result from the comparison of two groups in terms of which tends to make the higher response, in some sense. In this case, Oh (1995) discussed estimation of cell probabilities under S, L, G, and C alternatives under single multinomial, product multinomial, and Poisson sampling schemes.

In the $2 \times c$ case, the L condition is also called *likelihood-ratio ordering* (Lehmann, 1966). Dykstra and Lemke (1988) showed how to obtain the ML fit for this condition, and Dykstra et al. (1995) gave the LR test of whether two multinomial distributions are identical against it. The L condition is equivalent to c monotone decreasing binomial probabilities when one reverses orientation and considers the distribution of X given Y (i.e., fixing the column totals rather than the row totals). The LR test of identical multinomials against L (and the ML fit) is apparently equivalent to the one for comparing c independent binomial proportions against that simple ordered alternative in the distribution of X given Y; that is, the test considered by Robertson et al. (1988, p. 167). For comparing two multinomial distributions, Gautam et al. (1999) proposed using as test statistic the maximum possible squared correlation between the classifications, with maximum taken over all possible increasing scores for the columns. They showed that their test is asymptotically equivalent to the likelihood-ratio test for the two-sided alternative that one population is larger than the other with respect to the L condition.

For $2 \times c$ tables, the global and cumulative odds ratios are identical, so the conditions S and G are equivalent. The ML estimates of the two sets of multinomial probabilities, assuming a stochastic ordering, were derived by Brunk et al. (1966). Let $r_j = (n_{11} + \cdots + n_{1j})/(n_{21} + \cdots + n_{2j})$, $j = 1, \ldots, c$. The c columns are divided into subsets as follows: The first subset ends at the column v_1 for which r_{v_1} is the minimum of $\{r_1, \ldots, r_c\}$. If this does not include all columns, then the next subset consists of columns $\{v_1 + 1, \ldots, v_2\}$ such that r_{v_2} is the minimum of $\{r_{v_1+1}, \ldots, r_c\}$. One continues in this manner, forming the collection of subsets. Grove (1980) provided a geometric representation of this construction. In this construction of subsets, suppose a particular subset consists of columns $a, a+1, \ldots, b$. Then the ML fitted value under the stochastic ordering restriction for cell (i, j) in those columns equals

$$n_{ij}\,\frac{(n_{+a} + \cdots + n_{+b})(n_{i+})}{(n_{ia} + \cdots + n_{ib})(n)}.$$

Note that when $a = b$ (i.e., a subset contains a single column), the fitted value in that cell equals the fitted value for the independence model, namely $(n_{+j}n_{i+})/n$. The construction yields a single subset when the minimum ratio of cumulative proportions for row 1 to row 2 equals 1, the value for the final column. In that case, the sample counts themselves satisfy condition S, and the ML fitted values are simply those sample counts. Dykstra et al. (1996) showed that ML and various other estimation procedures for S are related and can be expressed in terms of a weighted least squares projection.

Grove (1980), Robertson and Wright (1981), and Bhattacharya and Dykstra (1994) discussed the LR test of whether two multinomial distributions are identical against the S alternative. See Lee (1987b), Lucas and Wright (1991), and Silvapulle (1994) for related tests. Lee et al. (1993) considered the S alternative with the additional constraint that the probabilities for one of the populations are monotone. Takeuchi and Hirotsu (1982) and Nair (1987) proposed cumulative chi-squared-type (CCS-type) tests designed for the S alternative and the ordered binomial alternative of Section 2. These tests are based on weighted sums of Pearson chi-square statistics from the 2×2 tables that result from the $c-1$ collapsings of the response into a binary outcome. These authors gave the asymptotic distribution and power for this class of statistics, and generalized the method to an arbitrary number of rows. Cohen, Sacrowitz, and colleagues (Berger and Sackrowitz, 1997; Cohen and Sackrowitz, 1998; Cohen et al., 2000) discussed a complete class of tests for the S alternative and provided conditions for tests to be unbiased, conditionally unbiased, and in the complete class. Their tests have a rather complex construction for finding the P-value, but the latter paper is highly critical of the LR test for the S alternative, as discussed later.

The C alternative, which corresponds to the condition of *uniform stochastic ordering* (Dykstra et al., 1991), has received considerably less attention than L or S. Grove (1984) proposed the LR statistic for comparing two multinomial populations against C, and Oluyede (1993a, b) presented an asymptotically equivalent statistic with Pearson components.

All asymptotic distributions for the various tests for $2 \times c$ tables are chi-bar-squared. Unfortunately, the weights $\{\rho_d\}$ for such distributions are unknown parameters because they depend on the common unknown multinomial probabilities; presumably, this may be problematic for small or unbalanced samples. For the S alternative, Wang (1996) showed that the asymptotic approximation performs well for small to moderate samples when $\{n_{i+}\}$ are approximately equal (e.g., $n_{1+} = n_{2+} = 30$) but rapidly deteriorates as the sample sizes become unbalanced. Grove (1980) for S and Dykstra et al. (1995) for L suggested bypassing such calculations by instead using the simpler approximation of the P-value that applies when the column totals are equal; in that case, simple recursive formulas exist for the weights in the chi-bar-squared distribution and tables of critical values exist for the distribution (Robertson et al., 1988, pp. 82, 416). To bypass possibly poor asymptotic approximations, we recommend simulating exact conditional tests for the LR test statistic (Agresti and Coull, 1998b).

3.2. Analyses for odds ratios in $r \times c$ tables

We now turn our attention to $r \times c$ tables, for which the research literature is more sparse. Denote the LR statistics for testing independence against the various ordered alternatives by $G^2(I|G), G^2(I|S), G^2(I|C), G^2(I|L)$. These all have the general form $\sum_i \sum_j n_{ij} \log(\hat{\pi}^a_{j|i}/\hat{\pi}^I_{j|i})$, where $\{\hat{\pi}^I_{j|i} = n_{+j}/n\}$ are the estimated probabilities under independence and $\{\hat{\pi}^a_{j|i}\}$ are the ML estimates under the alternative. Denote the LR statistic for the general, unrestricted, alternative by $G^2(I)$.

Patefield (1982) suggested the L condition as a natural alternative for tests of independence with ordinal data. He noted, however, the computational complexity of

LR tests because of the lack of closed form for the fitted values under that alternative. Using I-projection geometry, Dykstra and Lemke (1988) provided an algorithm for finding the ML fitted values under L and linked this order-restriction to the positive dependence concept of concordance; however they did not derive an asymptotic distribution for $G^2(I|L)$. The algorithm in El Barmi and Dykstra (1998) can also be used for this purpose. Cohen and Sackrowitz (1992) studied other ordinal statistics that are sometimes used for this alternative. They showed that certain tests, such as one based on numbers of concordant and discordant pairs, are sometimes inadmissible, and suggested a refinement of the usual P-value to reduce the degree of discreteness.

For the S alternative, Feltz and Dykstra (1985) noted that no closed form exists for the ML fit when $r > 2$. Wang (1996) suggested using an approximate test, such as one based on the bootstrap. Bhattacharya and Nandram (1996) and Evans et al. (1997) presented Bayesian analyses for this alternative.

Again, there is less work for the C condition. Grove (1984) and Oluyede (1994a) proposed large-sample chi-bar-squared tests (not LR) of independence against that alternative.

Dardanoni and Forcina (1998) provided a unified inference for the S and L alternatives. They demonstrated that each has form

$$\mathbf{g}(\boldsymbol{\pi}) = \mathbf{X}\boldsymbol{\beta}, \quad \mathbf{K}\boldsymbol{\beta} \geqslant \mathbf{0}, \tag{3}$$

where $\mathbf{g}$, $\mathbf{X}$, and $\mathbf{K}$ are specific to the type of ordering, and they applied a constrained Fisher scoring algorithm to maximize the log-likelihood subject to (3). They proved that the asymptotic distributions of the LR tests of independence against a particular alternative are of chi-bar squared form. They also showed that the LR tests of an inequality-constrained hypothesis against the general alternative have asymptotic distributions stochastically bounded above by a sum of chi-squared and chi-bar squared random variables.

Even for statistics that have limiting chi-bar-squared distributions, as in the $2 \times c$ case the weights must be estimated. We are unaware of any work evaluating conditions under which the chi-bar-squared approximations hold well for the $r \times c$ case, but the results of Wang (1996) suggest that the approximation may fail for small samples or unbalanced data sets. Agresti and Coull (1998b), who used an optimization program to obtain the ML fits for G, S, L, and C alternatives, suggested Monte Carlo simulation of exact conditional tests based on the LR test statistic.

Of the conditions $\{G, S, C, L\}$, L is the most restrictive. If L holds, in other words, then so do the other conditions. The nesting of L within S and C, which themselves are nested within G, implies that $G^2(I|L) \leqslant G^2(I|S) \leqslant G^2(I|G)$ and $G^2(I|L) \leqslant G^2(I|C) \leqslant G^2(I|G)$ and implies corresponding stochastic orderings of the null distributions. When the sample satisfies L, all four order-restricted fits and LR test statistics are identical (in fact, identical to $G^2(I)$). Hence, the P-values have the ordering induced by the null distributions, that for $G^2(I|L)$ being smallest. In practice, if the data *nearly* satisfy all the order restrictions, it is our experience that the L criterion usually provides the smallest P-value. Agresti and Coull (1998b) noted that if the true local log odds ratios are strictly positive, then the LR test based on it is asymptotically more powerful than the others.

The methods discussed in this section treat both X and Y as ordinal. When X and Y are nominal, it would not normally make sense to impose order restrictions. When one (say X) is nominal and one (say Y) is ordinal, restrictions involving Y may be of interest. One possible restriction states that the rows are stochastically ordered on Y, but the ordering of the rows is unknown. For instance, in comparing several treatments on Y, one might expect the response distributions for the treatments to be an unknown family of distributions that differ in location but not dispersion, yet one may have no prior prediction about the relative merits of the treatments. We have not seen any literature on ML estimation or LR tests for this alternative. Though not elegant, one could obtain the ML fit by comparing the log likelihoods for the fits for all $r!$ possible orderings. One could then use this minimum LR statistic as the criterion in a simulated exact test of independence. For this problem, Evans et al. (1997) used Bayesian methods to fit all $r!$ stochastic orderings. The authors evaluated which orderings are most plausible using Bayes factors and by comparing the prior and posterior distributions of a measure of the distance between the cell probabilities and the given stochastic ordering.

We are also unaware of any literature, other than Sampson and Singh (2002), on inequality constraints for partially ordered scales. For instance, X and/or Y could have ordered scales except for a category such as "don't know" or "undecided." Then, one might conduct inference subject to a condition such as L applying to the local odds ratios involving the ordinal categories. More generally, one could formulate "quasi-order-restricted models" by imposing constraints over a certain subset of a contingency table.

The discussion in this section shows that the emphasis in most of the order-restricted literature is on hypothesis testing rather than estimation, especially interval estimation. For any of the conditions discussed, it would seem to be more informative to provide confidence intervals for cell probabilities, odds ratios, or other effect measures, than to test hypotheses that would rarely be plausible. Standard methods would seem to be difficult to apply, but presumably recent computational advances make methods such as the bootstrap increasingly feasible. Model comparison issues also need much exploration. For instance, how does one decide whether a condition such as S provides a significantly better fit than a special case such as L or such as a certain ordinal model that satisfies S (such as model (4) discussed below)? The LR (deviance) statistic comparing models S and L does not have a chi-squared distribution. For instance, the number of parameters is the same for the two models, and one cannot express L as an interior point of the parameter space of S with a certain set of parameters equal to 0 (In fact, when L holds with strictly positive true local log odds ratios, then the likelihood-ratio statistic equals 0 with probability converging to 1 as n increases).

Finally, one could use other association measures besides the odds ratio to detect trends in ordinal contingency tables. These include, for instance, generalizations of Kendall's tau based on concordance and discordance (see, e.g., Cohen and Sackrowitz, 1992). Such statistics can serve as the basis of tests for trend such as the well known linear trend tests proposed by Mantel, Cochran, Armitage and many others. Rao et al. (1987) and Nguyen and Sampson (1987) proposed tests expressed in terms of global odds ratios based on concordance and discordance. For $2 \times c$ tables, Kimeldorf et al. (1992) used the isotonic regression techniques of Section 2.1 for binomial responses to find the maximum and minimum values of the sample correlation and of standard test statistics, among the possible values for all possible sets of increasing column scores. Sampson and Singh (2002) extended this to cases having only a partial ordering of categories. Podgor et al. (1996) showed how to use efficiency robustness principles to combine tests from two or more sets of scores into one robust test for analysis, in such a way as to minimize the worst possible efficiency loss over all the sets of scores. In addition, nonparametric methods that take category ordering into account such as the Jonckeere–Terpstra test and the Kolmogorov–Smirnoff test can be adapted to contingency tables (Cytel Software, 1995, pp. 305–311). Alternatively, one might use the approximate approach of treating an ordinal response as normal with constant variance; one can then use methods developed for testing equality of normal means against order-restricted alternatives. See Chuang-Stein and Agresti (1997) for a discussion of such approaches. We do not consider them here, since they do not utilize the categorical nature of the data. For instance, they do not provide a structure that yields fitted values that one can compare to observed cell counts. In this sense, the model-building approaches described in Section 4 are more appropriate.

3.3. Dose–response example revisited

In analyzing Table 1, we first test whether a more favorable outcome occurs when a patient receives treatment (regardless of dose) by applying the methods of Section 3.1 to the $2 \times c$ table that results from collapsing the final three rows. All four cumulative log odds ratios in the collapsed table exceed 0, so $G^2(I|S) = G^2(I) = 9.6$. Two of the local log odds ratios violate the ordering, and the LR statistic for the L alternative equals 7.9. The exact P-values for the unrestricted, S, and L alternatives are 0.048, 0.023, and 0.012, respectively. The order-restricted tests provide somewhat stronger evidence against independence than the chi-squared test of independence.

For the full table, similar substantive results occur for tests of independence with the various ordered alternatives for odds ratios; for the L, S, and G alternatives, the P-values based on simulated exact conditional tests are all at most 0.002. All the sample global log odds ratios exceed 0, so $G^2(I|G) = 27.8$ is identical to $G^2(I)$. The latter statistic, which is approximately chi-squared with df $= 12$, has $P = 0.006$, so in this case using ordinal inequality constraints does not much enhance the degree of evidence against independence. For the S alternative, two of the 12 sample cumulative odds ratios violate the S order restriction, but barely, and $G^2(I|S) = 27.7$. For the L alternative, five of the sample odds ratios violate it, and its fit is somewhat different from the observed data, giving $G^2(I|L) = 16.1$. Nonetheless, as mentioned above, its null distribution is stochastically lower, and this value is also sufficient to provide strong evidence of association ($P = 0.002$).

In using these test statistics, one should realize that a small P-value does not suggest that the order restriction truly holds, but merely that strong evidence exists against the null hypothesis of independence, based on that test criterion. In fact, it is possible to obtain small P-values even if some of the sample odds ratios violate the inequality constraints, as the results just given for $G^2(I|S)$ and $G^2(I|L)$ illustrate.

3.4. Anomalies with likelihood-ratio tests?

Recent evidence suggests that certain ML order-restricted fits and corresponding LR tests can behave in a counterintuitive manner. Even if all the sample log odds ratios are negative, many of the fitted log odds ratios may be strictly positive. A relatively small P-value may occur even though many, perhaps all, of the sample odds ratios contradict the order restriction. For instance, in Table 1 suppose the alternative of interest is G for the reverse ordering of the rows. All sample global log odds ratios are then negative, but the constrained fit for an alternative of nonnegative associations is not the fit of the independence model. In fact, 5 of the 12 fitted log odds ratios for the G constraints are then strictly positive and $G^2(I|G) = 11.8$ with $P = 0.26$, even though each sample odds ratio contradicts G.

Of the tests discussed in this section, in our experience the one based on $G^2(I|L)$ is least likely to provide seemingly anomalous results when the data strongly contradict the order restriction. To illustrate, suppose that every sample local log odds ratio violates L. Then, using properties of related loglinear models, Agresti and Coull (1998b) argued that the L fit is identical to the independence fit; thus, $G^2(I|L) = 0.0$ and $P = 1.0$, which seems sensible.

Cohen and Sackrowitz (1998) and Cohen et al. (2000) claimed that the LR test is inappropriate for various constrained alternatives, including $G^2(I|S)$. Cohen et al. argued that LR tests for testing equality of means for independent samples from an exponential family distribution are adequate for the simple order alternative of form $\mu_1 \leqslant \mu_2 \leqslant \cdots \leqslant \mu_k$ but can show anomalous behavior for most other inequality-constrained alternatives.

Perlman and Wu (1999) discussed in detail several papers, including those just mentioned, that have noted apparent anomalies of LR tests for inequality-constrained alternatives in various contexts. They argued that the examples used by Cohen and colleagues are not truly anomalous. Specifically, they argued that seemingly counter-intuitive results that occur when the sample data fall outside the parameter spaces described by the null and alternative hypotheses may partly reflect a null that is too sharp. They claimed that such results would not occur if testing were instead done with a null hypothesis consisting of the entire complement of the order-restricted space, rather than a small subset such as independence. Their arguments seem reasonable, but tests for such broader nulls have not yet been developed in the categorical case. This is an interesting and important area for future research. Meanwhile, order-restricted inference with sharper (or nonconvex) nulls should be used with caution when the sample data seem strongly inconsistent with both hypotheses. Again, we emphasize that a small P-value in an order-restricted test does not suggest that the order restriction truly holds, but merely that criterion provides strong evidence against the null. An analogy occurs in regression analysis, in which a test based on the slope may provide strong evidence against independence, even though we would not want to conclude that the true relationship is therefore exactly linear.

4. Inequality constraints on parameters in parametric models

Each monotonicity condition described in the previous section generalizes a particular logit or loglinear model that satisfies the condition. This section discusses these models. Unless specified otherwise, the models treat both X and Y as ordinal.

4.1. Multinomial logit models

The *cumulative logit model*:

$$\text{logit}[\gamma_{j|i}] = \alpha_j - \beta x_i \tag{4}$$

uses logits of cumulative probabilities to provide a stochastic ordering of levels of X. It has the *proportional odds* property whereby the effect β is the same for each cumulative probability. The ordinary linear logit model (1) is the special case of this model with $c = 2$. The usual approach is to assign monotone increasing scores $\{x_i\}$. Then, when $\beta > 0$, this model directly implies condition S; it has been pointed out to us by A. Forcina that this case also satisfies condition L (personal communication). For the equally spaced scores $\{x_i = i\}$, $\exp(\beta)$ is the cumulative odds ratio for each of the $(r-1)$ $(c-1)$ 2×2 tables formed using each pair of adjacent rows and a collapsing of the response above and below any particular cutpoint j. The independence model is the special case of (4) with $\beta = 0$.

ML fitting of model (4) (which is available in standard software, as discussed later) results in cell fitted values that can be compared to $\{n_{ij}\}$ with standard chi-squared statistics as a way of checking model goodness-of-fit. In addition, when we allow all real values for β, inference for β follows standard methods, such as the LR, score, or Wald-based tests and confidence intervals. One can use the sign of $\hat{\beta}$ times the square roots of the test statistics as approximately standard normal test statistics (under the null) for one-sided alternatives such as $\beta > 0$. If the model holds, more parsimonious and more powerful inferences result than with the model replacing $\{\beta x_i\}$ by unordered parameters $\{\beta_i\}$ (hence treating X as nominal), since the same chi-squared noncentrality is focused on a smaller df value. For the restricted parameter case of testing $\beta = 0$ against $\beta > 0$ without permitting negative values of β, under the null the usual statistics have large-sample distribution that is an equal mixture of degenerate at 0 (when the ordinary ML estimate $\hat{\beta}$ would be negative) and chi-squared with df$=$1. The asymptotic power advantage when truly $\beta > 0$ is the same as from doing a one-sided test with unrestricted β.

The cumulative logit form of model (4) is currently a popular one for modeling ordinal data. Although the use of a common effect β for each cumulative probability seems restrictive, it has connections with regression modeling of underlying continuous variables (McCullagh, 1980), and the more general model with effects $\{\beta_j\}$ has the structural problem that cumulative probabilities are misordered for some values of x. One could alternatively use transformations of the cumulative probability other than the logit, such as the probit and nonsymmetric ones (e.g., log–log and complementary log–log) for which the cumulative probability approaches 0 at a different rate than it approaches 1 (McCullagh, 1980). These alternative link functions maintain the stochastic order restriction but without maintaining identical values of the odds ratios for different j. Other alternative models include ones that apply logits or other transforms to probabilities other than the cumulative probabilities. For instance, the adjacent-categories logit model

$$\log(\pi_{j+1|i}/\pi_{j|i}) = \alpha_j + \beta x_i,$$

or generalizations of it that treat predictors as nominal factors, are related to loglinear models described in the next subsection (Agresti, 1984, 1996). This model form, which has natural connections with structures for local odds ratios, also implies stochastic orderings and usually fits well in the same situations that model (4) does.

Just as the linear logit model (1) generalizes to a model (2) with ordered effects of a factor, these models can do the same by replacing βx_i by β_i with $\beta_1 \geqslant \beta_2 \geqslant \cdots \geqslant \beta_r$. This generalization, which does not seem to have been addressed in the literature, provides greater flexibility but with a loss in parsimony and a potential loss in power.

4.2. Loglinear and association models

The cumulative logit model distinguishes between the response and explanatory variables. The *linear-by-linear association model*

$$\log[P(X = i, Y = j)] = \lambda + \lambda_i^X + \lambda_j^Y + \beta x_i y_j \tag{5}$$

is a loglinear model that has increasing row scores $\{x_i\}$ and column scores $\{y_j\}$ and treats the variables symmetrically (For identifiability, one needs constraints such as $\lambda_1^X = \lambda_1^Y = 0$). With the equally spaced scores $\{x_i = i\}$ and $\{y_j = j\}$, all $(r-1)(c-1)$ of the local odds ratios are identical to $\exp(\beta)$. When $\beta > 0$, this model implies condition L. To test H_0: $\beta = 0$ against H_a: $\beta \neq 0$, one can use standard LR, score, or Wald statistics, all of which have asymptotic chi-squared distributions with df $= 1$. For the one-sided alternative, their signed square root is approximately standard normal under the null; or, restricting the parameter to be nonnegative results in the squared statistics having asymptotic distributions that are an equal mixture of degenerate at 0 and chi-squared with df $= 1$. For small samples, one can use conditional methods to conduct an exact small-sample one-sided test. This test is available in the software StatXact (Cytel Software, 1995).

More general models in which $\{\mu_i = \beta x_i\}$ or $\{v_j = \beta y_j\}$ are parameters treat those classifications as nominal. In some applications, it is reasonable to use the model assuming ordered scores but without specifying values for those scores. Agresti et al. (1987) considered ML fitting of this model when $\{v_j\}$ are fixed but $\{\mu_i\}$ are known only to be monotone increasing. Dykstra and Lemke (1988) provided an alternative fitting procedure for this case, and Ritov and Gilula (1991) extended this to the case where both $\{\mu_i\}$ and $\{v_j\}$ are unknown but assumed to be increasing. These generalized models with inequality-constrained scores all are special cases of condition L.

Cumulative logit models can be described with cumulative odds ratios and linear-by-linear association models and other ordinal loglinear models can be described with local odds ratios. Relatively less attention has been paid to models for global odds ratios. For two-way tables, Plackett (1965) proposed a quadrant dependence model in which each global odds ratio is identical (see also Wahrendorf, 1980; Dale, 1986). This model, together with a positive sign for those log odds ratios, implies condition G. See Molenberghs and Lesaffre (1994) for an extension to multivariate responses with covariates. Of these three types of models for odds ratios, the cumulative logit models generalize most naturally to standard regression settings with a single response variable and several predictors, since they explicitly distinguish between response and explanatory variables whereas loglinear and association models do not. That model also generalizes to the mixed model setting in which correlations among repeated responses are generated using random effects in the model; see, for instance, Hedeker and Gibbons (1994).

The ordinal models with either fixed scores or qualitative factors for ordinal predictors have become quite popular in applications. By contrast, inequality-constrained models with ordered but unspecified parameter scores have not seen much application. Although they have the advantage of greater generality compared to models assuming a linear trend for fixed scores, there is the potential of loss of power from the addition of parameters (e.g., the linear trend models have only a single association parameter), and the user can no longer rely on standard chi-squared tests and confidence intervals for the effects. Moreover, there are also potential pitfalls that need investigation. For instance, Section 3.4 alluded to possible anomalies with LR tests for inequality constraints on some log odds ratios. It is of interest to study this in more detail and investigate the suitability or unsuitability of order-restricted inferences for related models that are described by such odds ratios. In particular, following Perlman and Wu (1999), it may be of interest to develop tests in which the null hypothesis is the entire complement of the model's inequality-constrained space.

Finally, there are less-commonly used classes of models, not discussed here, for which parameters have no connection with odds ratios. For instance, Goodman (1985, 1996) noted similarities between association models such as (5) and correlation models and correspondence analysis models. These other types of models also have score parameters for which order restrictions may be useful. See Böckenholt and Böckenholt (1990), Parsa and Smith (1993), Ritov and Gilula (1993), and Takane et al. (1991) for work in this direction.

4.3. Dose–response example revisited

One can fit cumulative logit models (such as model (4)) with various commercially available software; for instance, it is the default option for multinomial modeling with PROC LOGISTIC in SAS. One can fit loglinear models such as the linear-by-linear association model (5) with any software for generalized linear models, such as PROC GENMOD in SAS. See, for instance, Agresti (1996) for details.

Using dose scores $(1,2,3,4)$ with Table 1, the cumulative logit model has $\hat{\beta}=0.176$ (standard error=0.056). The LR statistic for testing that $\beta=0$ equals 9.6, and the signed square root of 3.1 has a one-sided P-value of 0.001 for the alternative $\beta>0$. Similar substantive results occur for other ordinal models using a single parameter to describe a linear trend. For instance, the linear-by-linear association model (5) has signed square root of the LR statistic equal to 3.1 ($P=0.001$). In summary, ordinal

LR test statistics provide strong evidence against independence, whether we measure the evidence using fits with inequality constraints assuming nonnegative log odds ratios or using a common nonnegative value for certain log odds ratios.

5. Extensions to multi-way tables and multivariate responses

In practice these days, multiway contingency tables are more common than two-way tables. In some cases the extra dimensions refer to additional covariates, such as in a biomedical study that stratifies for several age groups or medical centers. In other cases the many dimensions result from the categorical response being multivariate; a common example is when a categorical response is measured repeatedly, such as in a longitudinal study. There does not seem to be much literature on inequality-constrained methods for such cases, although some of the methods of the previous three sections apparently generalize fairly directly.

5.1. Stratified data

To illustrate, let Z denote a control variable, and let $\theta_{ij(k)}$ denote an odds ratio computed between X and Y in partial table k of Z, $k = 1,\ldots,K$. If Y alone is the response, the cumulative odds ratio may be most natural, whereas if X and Y are responses then the local or global odds ratio may be preferred. For any definition of positive association, one could estimate cell probabilities within each partial table, subject to that constraint. In addition, one could test conditional independence of X and Y, given Z, against the alternative

$$\log\theta_{ij(k)} \geqslant 0, \quad i = 1,\ldots,r-1, \;\; j = 1,\ldots,c-1, \;\; k = 1,\ldots,K.$$

The LR statistic is the sum of the LR statistics from applying the test of the previous section separately to each partial table. This stratified test refers to a large number of parameters, particularly if K is large, which is disadvantageous to power.

When one expects the true pattern of association to be similar in each partial table, more powerful tests can take advantage of this similarity. One can do this, in the spirit of the Cochran–Mantel–Haenszel test for stratified 2×2 tables, by basing the test on the common odds ratio structure

$$\theta_{ij(1)} = \cdots = \theta_{ij(K)}$$

for all i and j. One maximizes the likelihood for the three-way table subject to the constraint that the logs of these $(r-1)(c-1)$ parameters simultaneously are at least 0. In practice, such a test can have a substantial power advantage (especially for large K) unless there is actually severe heterogeneity. See Agresti and Coull (1998b) for details.

A more fully parametric approach generalizes basic logit and loglinear models. For instance, the extension of the linear-by-linear association model (5) to

$$\log[P(X = i, Y = j, Z = k)] = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \beta x_i y_j \tag{6}$$

assumes that each conditional local odds ratio between X and Y is homogeneous across levels of Z. The pattern is determined by the scores $\{x_i\}$ and $\{y_j\}$ and the strength is determined by β; specifically, $\log\theta_{ij(k)} = \beta(u_{i+1} - u_i)(v_{j+1} - v_j)$. Similarly, it is straightforward to add a stratification factor to the cumulative logit model (4).

Another order-restricted approach for stratified tables specifies an ordering among strata. For $2 \times 2 \times K$ tables, denote the odds ratio for the kth stratum by θ_k. Anraku (1994) obtained conditional ML and Mantel–Haenszel estimates of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$ under the restriction $\theta_1 \leqslant \cdots \leqslant \theta_K$; see El Barmi (1997) for related work. Alternatively, one could use logit or loglinear models specifying alternatives such as a linear trend in log odds ratios across the strata. For stratified $r \times c$ contingency tables, one can consider a heterogeneous association version of (6) that replaces β by β_k, and then use alternatives such as $\beta_1 \leqslant \cdots \leqslant \beta_K$ or a linear trend in $\{\beta_k\}$. Sampson and Whitaker (1989) obtained ML estimates of cell probabilities when one stratum is assumed to be stochastically larger than a second one. Specifically, in each of two layers of an $r \times c \times 2$ table, this stochastic ordering alternative is

$$\sum_{(i,j)\in U} P(X=i, Y=j|Z=1) \geqslant \sum_{(i,j)\in U} P(X=i, Y=j|Z=2)$$

for all sets U satisfying $(i,j) \in U$ and $i \leqslant i'$, $j \leqslant j' \Rightarrow (i',j') \in U$.

As in the two-way case, it is unclear whether LR statistics with inequality constraints can have undesirable properties. Presumably analyses based on the local odds ratio behave reasonably, because of the connection with loglinear models. Here there is also much room for work on interval estimation based on inequality-constrained models. Likewise, as the number of predictors increases, there is scope for addressing issues of model comparison and selection.

5.2. Marginal homogeneity for multivariate responses

A somewhat different multi-way contingency table problem results from repeated measurement of a categorical response. Specifically, suppose a multivariate response has the same ordered categorical scale for each component. One such application is when subjects in a survey provide their opinion about several related items, such as whether government spending should (increase, stay the same, decrease) for defense, health, education, the environment, and so forth. Another application is when a longitudinal study measures the same variable on the same subjects at several time points. In such cases, it is often of interest to compare the marginal distributions; for instance, to analyze whether subjects tend to prefer higher government spending on health or on the environment.

El Barmi and Dykstra (1995) gave general results for tests pertaining to multinomial probabilities that, as a special case, can provide direct comparisons of marginal distributions. For instance, consider the case of $r = 2$ repeated responses. One can then express models in terms of a square $c \times c$ contingency table, with the same categories in each dimension, that cross classifies the first and the second response. For the cell probabilities $\{\pi_{ij}\}$ in such a square table, one might be interested in the stochastic ordering condition,

$$\pi_{1+} + \cdots + \pi_{j+} \leqslant \pi_{+1} + \cdots + \pi_{+j} \text{ for } j = 1, \ldots, c-1.$$

Robertson et al. (1988, p. 290) discussed projections that provide fitted values for this alternative. Using an algorithm based on Fenchel duality, El Barmi and Dykstra (1995) tested for this stochastic ordering against the unrestricted alternative. Agresti and Coull (1998b) described alternative ways of performing tests of marginal homogeneity against various order-restricted alternatives.

Other order-restricted approaches for square tables apply to departures from symmetry of cell probabilities, the strong condition by which $\pi_{ij} = \pi_{ji}$ for all i and j. El Barmi and Kochar (1994) tested symmetry against H_a: $\pi_{ij} \geqslant \pi_{ji}$, for $i \geqslant j$. Robertson (1986) tested symmetry and a type of unimodality around the main diagonal against alternatives where one of these does not hold. Goodman (1985) described a model that specifies the odds π_{ij}/π_{ji} to be a monotone function in $(i-j)$. One can obtain this solution by fitting the log-linear model of diagonal asymmetry,

$$\log(\pi_{ij}) = \lambda_{ij} + \delta_{i-j} I(i > j), \; \lambda_{ij} = \lambda_{ji},$$

where I is an indicator function, using the pool-adjacent-violators algorithm to constrain $0 \leqslant \delta_1 \leqslant \cdots \leqslant \delta_{c-1}$. Since the model has concave log likelihood, when the unconstrained estimates violate the constraint the ML solution occurs on the boundary of the order-restricted parameter space. Alternatively, one could fit models that have common values (Bhattacharya, 1998) or linear trends (Agresti, 1983) in such terms, for instance $\log(\pi_{ij}/\pi_{ji}) = \delta(i-j)$ for $i > j$.

Again, open questions concern whether LR tests with inequality constraints may, in certain situations, provide anomalous results, and whether improved behavior results from using broader null hypotheses. Also, in practice hypotheses of symmetry and marginal homogeneity are usually not plausible, and it is of greater interest to model the components of the multivariate response in terms of predictors while adjusting for the dependence among repeated measurements. A challenging problem would be isotonic model-building of each component of the multivariate response in terms of predictors, perhaps adapting ideas from the literature on generalized estimating equations to obtain reasonable marginal fits even when the joint association structure is misspecified.

6. Comparison and practical application of order-restricted methods

This final section discusses the need for research on comparing different approaches for handling order restrictions and the need for literature and software to make inequality-constrained methods more accessible to data analysts.

6.1. Order restriction: inequality constraints vs. linear trends

In this review, we have seen that it is straightforward to conduct statistical inference using models that incorporate ordering using linear trends for a fixed set of scores; however, it can be complex even for large samples to conduct such inference using models that express trends purely in terms of inequality constraints, whether the constraints apply to probabilities, odds ratios, or model parameters for levels of a factor. Thus, a natural question to ask is whether it is worth the extra effort to conduct inference based on inequality constraints. For instance, if the inequality constraints (2) on binomial probabilities truly holds but the simpler linear logit model (1) does not, is there the potential of a substantive power gain by performing a test of independence against the inequality-constrained alternative instead of the linear logit test? In terms of estimation performance, when do the inequality-constrained estimates of the binomial probabilities tend to be better than the linear logit estimates?

If the linear logit model (1) holds or nearly holds, it is natural to use inference based on it. In practice, though, one would often expect the binomial parameters to change monotonically without satisfying, even approximately, the linear logit model. Yet, there may be no a priori reason to choose a particular alternative link or structural form for the relationship. Is the inequality-constrained inference superior as one moves sufficiently far away from the linear logit model? One would expect so, but that inference must counteract the parsimonious benefit the linear logit-based inference has of focusing inference on a single parameter.

Questions of these types have not received much attention for categorical data analyses. Poon (1980) and Collings et al. (1981) compared the performance of the one-sided Armitage test for trend, the test of Barlow et al. (1972) based on the inverse sine transformation of the proportions, and test statistics based on the isotonic regression of the sample proportions under ordered alternatives, umbrella alternatives of the form $p_1 \leqslant p_2 \leqslant \cdots \leqslant p_i > p_{i+1} \geqslant \cdots \geqslant p_r$, and a probit model linear in log dose. These authors noted that the trend test and inverse sine test suffer appreciably under the umbrella alternative with respect to power, and concluded that the isotonic tests are preferable in this case. Agresti and Coull (1998a) performed some comparisons of models (1) and (2). For the moderate-sized values of r commonly encountered in practice (say, around 4 or 5), they noted that if the true probabilities are strictly monotone but deviate somewhat from the linear logit model, it is still better to use logit-based probability estimates unless the sample size is quite large. There may be scope here to develop monotone models that are smoother than the inequality-constrained methods.

6.2. The future of order-restricted methods for contingency tables

A substantial literature exists on order-restricted methods of inference, both for continuous and categorical responses. Generalized linear models that incorporate ordering of response categories or factor levels with linear trends on some scale are popular in practice. Despite the extensive literature on methods that require only an inequality constraint rather than structural trends such as linearity, such methods seem to be rarely used. They do not typically appear in applied statistics texts or in the major software packages. There are undoubtedly a variety of reasons for this. First, the methods are not simple to apply, requiring algorithms and/or nonstandard limiting distributions even for simple problems such as the one-way layout with normal responses. This should not be a major hurdle, however, in the modern computing age. Second, much of the research literature is not easily readable by applied statisticians and other data analysts. There is a definite need for a book on inequality-constrained inference having an applied orientation.

Surprisingly, procedures do not seem to be available in popular software packages for even some basic problems such as testing equality of parameters from an exponential family against a simply-ordered alternative. Easy availability of such software to users would be an important first step in increasing the usage of such methods. For categorical data problems, some specialized programs are available and others undoubtedly exist of which we are unfamiliar. Agresti and Coull (1996) provided software for simulating P-values of exact LR tests comparing several binomial parameters against the ordered alternative and comparing two multinomial distributions against the stochastic ordering alternative. Dardanoni and Forcina (1998) provided general MATLAB programs for large-sample inequality constrained testing in $r \times c$ tables with S, L, and C alternatives.

Besides the reasons just cited, the lack of application may reflect the rather limited development of inequality-constrained methods, which for the most part relate to fairly elementary problems. For multinomial data, for instance, the major emphasis has been on comparing two groups or inference about probabilities for a single multinomial distribution. Moreover, most of this work has focused on hypothesis testing, for which there may indeed usually be little benefit to using order-restricted methods. For instance, a test based on a linear trend parameter in a model may perform adequately and even have greater power than the inequality-constrained test unless reality departs quite markedly from the model.

There seems to be considerable scope, on the other hand, for the further development of inequality-constrained estimation and modeling with generalized linear models. For instance, does the good performance of inequality-constrained estimates when a trend in binomial proportions is monotone but irregular or with potential slight departures from monotonicity extend to other generalized linear models? How can one reduce the bias in the constrained ML estimates? How can one obtain isotonic regression fits that are smoother than the jagged ones provided by ML? Is it straightforward to include additional variables as stratification factors? How can one summarize effects in models with inequality constraints on odds ratios? How can one select a model form and the relevant predictors for that model? How can one judge whether a model with inequality constraints fits better than one with even more restrictive inequality constraints or even one with linear effects? Especially important, how can one construct confidence intervals for relevant parameters in the context of such models? For examples of some recent uses of order-restricted methods in estimation and modeling, see Disch (1981), Wolfe et al. (1981), Schmoyer (1984), Agresti et al. (1987), Morris (1988), McDonald and Diamond (1990), Piegorsch (1990), Wax and Gilula (1990), Bacchetti (1989), Gelfand and Kuo (1991), Geyer (1991), Ritov and Gilula (1991), Silvapulle (1994), Eddy et al. (1995), Paula and Sen (1995), Nandram et al. (1997), and Park (1998).

As computations become increasingly feasible with modern simulation methods, the Bayesian approach to inference is becoming more popular, even for relatively complex, hierarchical models. There seems to be considerable scope for developing order-restricted methods in a Bayesian framework. When it is sensible to assume an order restriction $\beta_1 \leqslant \beta_2 \leqslant \cdots \leqslant \beta_r$ for a set of parameters, one way to ensure that the posterior distribution recognizes this constraint is by specifying the prior in terms of parameters such as $\log(\beta_j - \beta_{j-1})$. When sample estimates violate the order, ML estimates commonly fall on the boundary of the constrained parameter space; an appealing aspect of posterior means in a Bayesian approach is that they often fall in the interior of the parameter space. Fahrmeir and Klinger (1994), Agresti and Chuang (1986), Gelfand et al. (1992), and McDonald and Prevost (1997) reviewed existing approaches to fitting parameter-constrained models, including Bayesian ones in the latter three papers.

Finally, order-restricted methods could be extended to additional types of applied problems. For instance, there has been little work on order-restricted methods for repeated measurement data, as opposed to independent samples. See Mukerjee (1988) and Shin et al. (1996) for examples in the normal-theory context and George and Kodell (1996) for an example in a multivariate binary response setting. In addition, model-based inference for inequality-constrained parameters could be extended to more complex models for multi-way tables. A variety of applications may benefit from such analyses in a generalized linear mixed model or latent variable format. This poses a tremendous challenge, as log-likelihoods can be poorly behaved and it can be difficult to construct proper inferences even in ordinary mixed models without constraints. See Hoijtink and Molenaar (1997) for an inequality constrained analysis with item response models. For an even greater challenge, such methods could be extended to generalized additive models (Hastie and Tibshirani, 1990). For instance, one might want to predict a binomial probability as an unspecified smooth function of a dose level for each of two groups, under the constraint that one response curve always falls above the other. For all such models, the development of influence diagnostics would be useful; see Paula (1993) for some work in this direction for generalized linear models.

In summary, the surface seems to have been barely scratched in developing practically useful methods with inequality constraints for contingency tables. Many challenges remain, but the usefulness of such methods for the applied statistician is also in doubt. In our opinion, much of the future work should be devoted to (1) evaluating theoretical properties of LR inference and alternatives with reference to when the methods may have inappropriate behavior, (2) investigating issues of interval estimation and model selection and building, (3) evaluating when it is worth the effort to use such methods instead of simpler models with linear trends, (4) preparing survey articles or monographs showing data analysts how to use the methods, and (5) preparing user-friendly software for conducting the methods. This work should keep those interested in researching such methods busy well into the new millenium.

Acknowledgements

This work was supported by grants from the National Institutes of Health and the National Science Foundation (A. Agresti) and NIEHS grant ESO7142 (B.A. Coull). The authors are appreciative of helpful comments by two referees and by Antonio Forcina, Arthur Cohen, and Harold Sackrowitz.

References

Agresti, A., 1983. A simple diagonals-parameter symmetry and quasi-symmetry model. Statist. Prob. Lett. 1, 313–316.

Agresti, A., 1984. Analysis of Ordinal Categorical Data. Wiley, New York.

Agresti, A., 1996. An Introduction to Categorical Data Analysis. Wiley, New York.

Agresti, A., Chuang, C., 1986. Bayesian and maximum likelihood approaches to order-restricted inference for models for ordinal categorical data. In: Dykstra, R., Robertson, T., Wright, F.T. (Eds.), Advances in Order Restricted Statistical Inference. Springer, New York, pp. 6–27.

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

Agresti, A., Coull, B.A., 1998a. An empirical comparison of inference using order-restricted and linear logit models for a binary response. Comm. Statist. Simulation Comput. 27, 147–166.

Agresti, A., Coull, B.A., 1998b. Order-restricted inference for monotone trend alternatives in contingency tables. Comput. Statist. Data Anal. 28, 139–155.

Agresti, A., Chuang, C., Kezouh, A., 1987. Order-restricted score parameters in association models for contingency tables. J. Amer. Statist. Assoc. 82, 619–623.

Amundsen, H.T., Ljøgodt, H., 1979. Small sample tests against an ordered set of binomial probabilities. Scand. J. Statist. 6, 81–85.

Anraku, K., 1994. Estimation of odds ratios under order restrictions. Comm. Statist. Theor. Methods 23, 3257–3272.

Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W.T., Silverman, E., 1955. An empirical distribution for sampling with incomplete information. Ann. Math. Statist. 26, 641–647.

Bacchetti, P., 1989. Additive isotonic models. J. Amer. Statist. Assoc. 84, 289–293.

Baras, M., 1983a. Testing equality of probabilities of k mutually exclusive events against ordered alternatives. Biometrika 70, 473–478.

Baras, M., 1983b. Testing randomness in a multinomial experiment with grouped frequencies. Comm. Statist. Theor. Methods 12, 2575–2580.

Barlow, R.E., Bartholomew, D.J., Bremner, J.M., Brunk, H.D., 1972. Statistical Inference Under Order Restrictions. Wiley, New York.

Barnhart, H.X., Sampson, A.R., 1994. Overview of multinomial models for ordinal data. Comm. Statist. Theor. Methods 23, 3395–3416.

Bartholomew, D.J., 1959. A test of homogeneity for ordered alternatives. Biometrika 46, 36–48.

Bennett, B.M., 1962. On an exact test for trend in binomial trials and its power function. Metrika 5, 49–53.

Berger, V., Sackrowitz, H., 1997. Improving tests for superior treatment in contingency tables. J. Amer. Statist. Assoc. 92, 700–705.

Bhattacharya, B., 1995. Restricted tests for and against the increasing failure rate ordering on multinomial parameters. Statist. Probab. Lett. 25, 309–316.

Bhattacharya, B., 1997. Testing multinomial parameters under order restrictions. Comm. Statist. Theor. Methods 26, 1839–1865.

Bhattacharya, B., 1998. Testing conditional symmetry against one-sided alternatives in square contingency tables. Metrika 47, 71–84.

Bhattacharya, B., Dykstra, R.L., 1994. Statistical inference for stochastic ordering. In: Shaked, M., Shanthikumar, J.G. (Eds.), Stochastic Orders and their Applications. Academic Press, Boston, pp. 221–249.

Bhattacharya, B., Nandram, B., 1996. Bayesian inference for multinomial populations under stochastic ordering. J. Statist. Comput. Simulation 54, 145–163.

Block, H., Qian, S., Sampson, A.R., 1994. Structure algorithms for partially ordered isotonic regressions. J. Computat. Graph. Statist. 3, 285–300.

Böckenholt, U., Böckenholt, I., 1990. Canonical analysis of contingency tables with linear constraints. Psychometrika 55, 633–639.

Brunk, H.D., Franck, W.E., Hanson, D.L., Hogg, R.V., 1966. Maximum likelihood estimation of the distributions of two stochastically ordered random variables. J. Amer. Statist. Assoc. 61, 1067–1080.

Chacko, V.J., 1966. Modified chi-square test for ordered alternatives. Sankhya, Ser. B 28, 185–190.

Chuang-Stein, C., Agresti, A., 1997. A review of tests for detecting a monotone dose-response relationship with ordinal response data. Statist. Med. 16, 2599–2618.

Cohen, A., Sackrowitz, H.B., 1992. An evaluation of some tests of trend in contingency tables. J. Amer. Statist. Assoc. 87, 470–475.

Cohen, A., Sackrowitz, H.B., 1998. Directional tests for one-sided alternatives in multivariate models. Ann. Statist. 26, 2321–2378.

Cohen, A., Kemperman, J.H.B., Sackrowitz, H.B., 2000. Properties of likelihood inference for order restricted models. J. Multivariate Analysis 72, 50–77.

Cohen, A., Sackrowitz, H.B., Sackrowitz, M., 2000. Testing whether treatments is "better" than control with ordered categorical data: An evaluation of new methodology. Statistics in Medicine 2699–2712.

Collings, B.J., Margolin, B.H., Oehlert, G.W., 1981. Analyses for binomial data, with application to the fluctuation test for mutagenicity. Biometrics 37, 775–794.

Cytel Software, 1995. StatXact, Version 3. Cytel, Cambridge, MA.

Dale, J.R., 1984. Local versus global association for bivariate ordered responses. Biometrika 71, 507–514.

Dale, J.R., 1986. Global cross-ratio models for bivariate, discrete, ordered responses. Biometrics 42, 909–917.

Dardanoni, V., Forcina, A., 1998. A unified approach to likelihood inference on stochastic orderings in a non-parametric context. J. Amer. Statist. Assoc. 93, 1112–1123.

Dinh, K.T., Nguyen, T.T., 1994. Maximum likelihood estimators of binomial parameters under an order restriction. Amer. Statistician 48, 29–30.

Disch, D., 1981. Bayesian nonparametric inference for effective doses in a quantal-response experiment. Biometrics 37, 713–722.

Douglas, R., Fienberg, S.E., 1990. An overview of dependency models for cross-classified categorical data involving ordinal variables. In: Block, H.W., Sampson, A.R., Savits, T.S. (Eds.), Topics in Statistical Dependence, Institute of Mathematical Statistics Lecture Notes, Hayward, CA, pp. 167–188.

Douglas, R., Fienberg, S.E., Lee, M.-L.T., Sampson, A.R., Whitaker, L.R., 1990. Positive dependence concepts for ordinal contingency tables. In: Block, H.W., Sampson, A.R., Savits, T.H. (Eds.), Topics in Statistical Dependence. Institute of Mathematical Statistics Lecture Notes, Hayward, CA, pp. 189–202.

Dykstra, R.L., Lee, C.C., 1991. Multinomial estimation procedures for isotonic cones. Stat. Prob. Lett. 11, 155–160.

Dykstra, R.L., Kochar, S., Robertson, T., 1991. Statistical inference for uniform stochastic ordering in several populations. Ann. Statist. 19, 870–888.

Dykstra, R.L., Kochar, S., Robertson, T., 1995. Inference for likelihood ratio ordering in the two-sample problem. J. Amer. Statist. Assoc. 90, 1034–1040.

Dykstra, R.L., Lee, C.C., Yan, X., 1996. Multinomial estimation procedures for two stochastically ordered distributions. Statist. Probab. Lett. 30, 353–361.

Dykstra, R.L., Lemke, J., 1988. Duality of I projections and maximum likelihood estimation for log-linear models under cone constraints. J. Amer. Statist. Assoc. 83, 546–554.

Dykstra, R.L., Robertson, T., 1982a. An algorithm for isotonic regression for two or more independent variables. Ann. Statist 10, 708–711.

Dykstra, R.L., Robertson, T., 1982b. Order restricted statistical tests on multinomial and Poisson parameters: The starshaped restriction. Ann. Statist. 10, 1246–1252.

Eddy, W.F., Quan, S., Sampson, A., 1995. Isotonic probability modeling with multiple covariates. Computer Science and Statistics: Proceedings of the 27th Symposium on the Interface, Vol. 27, pp. 500–505.

Eeden, C.V., 1956. Maximum likelihood estimation of ordered probabilities. Indag. Math. 18, 444–455.

El Barmi, H., 1997. Testing for or against a trend in the odds ratios in k 2×2 contingency tables. Commun. Statist. Theory Methods 26, 1877–1891.

El Barmi, H., Dykstra, R.L., 1994. Restricted multinomial maximum likelihood estimation based upon Fenchel duality. Statist. Probab. Lett. 21, 121–130.

El Barmi, H., Dykstra, R.L., 1995. Testing for and against a set of linear inequality constraints in a multinomial setting. Canad. J. Statist. 23, 131–143.

El Barmi, H., Dykstra, R.L., 1996. Restricted product multinomial and product Poisson maximum likelihood estimation based upon Fenchel duality. Statist. Probab. Lett. 29, 117–123.

El Barmi, H., Dykstra, R.L., 1998. Maximum likelihood estimates via duality for log-convex models when cell probabilities are subject to convex constraints. Ann. Statist. 26, 1878–1893.

El Barmi, H., Kochar, S., 1994. Likelihood ratio tests for bivariate symmetry against ordered alternatives in a square contingency tables. Statist. Probab. Lett. 22, 167–173.

Evans, M., Gilula, Z., Guttman, I., Swartz, T., 1997. Bayesian analysis of stochastically ordered distributions of categorical variables. J. Amer. Statist. Assoc. 92, 208–214.

Fahrmeir, L., Klinger, J., 1994. Estimating and testing generalized linear models under inequality restrictions. Statist. Papers 35, 211–229.

Feltz, C.J., Dykstra, R.L., 1985. Maximum likelihood estimation of the survival functions of N stochastically ordered random variables. J. Amer. Statist. Assoc. 80, 1012–1019.

Fygenson, M., 1997. A new approach in modeling a categorical response. I. Binary response. J. Amer. Stat. Assoc. 92, 322–332.

Gautam, S., Sampson, A.R., Singh, H., 1999. Iso-chi-squared testing of $2 \times k$ ordered tables. Unpublished manuscript.

Gebhardt, F., 1970. An algorithm for monotone regression with one or more independent variables. Biometrika 57, 263–271.

Gelfand, A.E., Kuo, L., 1991. Nonparametric Bayesian bioassay including ordered polytomous response. Biometrika 78, 657–666.

Gelfand, A.E., Smith, A.F.M., Lee, T.M., 1992. Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. J. Amer. Statist. Assoc. 87, 523–532.

George, E.O., Kodell, R.L., 1996. Tests of independence, treatment heterogeneity, and dose-related trend with exchangeable binary data. J. Amer. Statist. Assoc. 91, 1602–1610.

Geyer, C.J., 1991. Constrained maximum likelihood exemplified by isotonic convex logistic regression. J. Amer. Statist. Assoc. 86, 717–724.

Goodman, L.A., 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. Ann. Statist. 13, 10–69.

Goodman, L.A., 1996. A simple general method for the analysis of cross-classified data: Reconciliation and synthesis of some methods of Pearson, Yule, and Fisher, and also some methods of correspondence analysis and association analysis. J. Amer. Statist. Assoc. 91, 408–428.

Greenberg, I., 1985. A one-sided goodness-of-fit test for a multinomial population. J. Amer. Statist. Assoc. 80, 558–562.

Grove, D.M., 1980. A test of independence against a class of ordered alternatives in a $2 \times c$ contingency table. J. Amer. Statist. Assoc. 75, 454–459.

Grove, D.M., 1984. Positive association in a two-way contingency table: Likelihood ratio tests. Comm. Statist. Theor. Methods 13, 931–945.

Grove, D.M., 1986. Positive association in a two-way contingency table: A numerical study. Commun Statist. Simul. Comput. 15, 633–648.

Hastie, T., Tibshirani, R., 1990. Generalized Additive Models. Chapman & Hall, London.

Hedeker, D., Gibbons, R.D., 1994. A random-effects ordinal regression model for multilevel analysis. Biometrics 50, 933–944.

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, 171–189.

Kimeldorf, G., Sampson, A.R., Whitaker, L.R., 1992. Min and max scorings for two-sample ordinal data. J. Amer. Statist. Assoc. 87, 241–247.

Kulatunga, D.D.S., Asai, M., Sasabuchi, S., 1996. A simulation study of several tests of the equality of binomial probabilities against ordered alternatives. J. Statist. Comput. Simul. 56, 53–75.

Lancaster, H.O., 1949. The derivation and partition of χ^2 in certain discrete distributions. Biometrika 36, 117–129.

Lancaster, H.O., 1961. Significance tests in discrete distributions. J. Amer. Statist. Assoc. 56, 223–234.

Lee, Y.J., 1977. Maximin tests of randomness against ordered alternatives: The multinomial distribution case. J. Amer. Statist. Assoc. 72, 673–675.

Lee, Y.J., 1980. Test of trend in count data: multinomial distribution case. J. Amer. Statist. Assoc. 75, 1010–1014.

Lee, C.C., 1983. The min-max algorithm and isotonic regression. Ann. Statist. 11, 467–477.

Lee, C.C., 1987a. Chi-squared tests for and against an order restriction on multinomial parameters. J. Amer. Statist. Assoc. 82, 611–618.

Lee, C.C., 1987b. Maximum likelihood estimation for stochastically ordered multinomial populations with fixed and random zeros. In: MacNeill, I.B., Umphrey, G.J. (Eds.), Foundations of Statistical Inference, Reidel, Boston, pp. 189–197.

Lee, C.C., Robertson, T., Wright, F.T., 1993. Bounds on distributions arising in order restricted inferences with restricted weights. Biometrika 80, 405–416.

Lehmann, E., 1966. Some concepts of dependence. Ann Math. Statist. 37, 1137–1153.

Liu, Q., 1998. An order-directed score test for trend in ordered $2 \times K$ tables. Biometrics 54, 1147–1154.

Lucas, L.A., Wright, F.T., 1991. Testing for and against a stochastic ordering between multivariate multinomial populations. J. Multivariate Anal. 38, 167–186.

McCullagh, P., 1980. Regression models for ordinal data (with discussion). J. Roy. Statist. Soc. Ser. B 42, 109–142.

McDonald, J.W., Diamond, I.D., 1990. On the fitting of generalized linear models with nonegativity parameter constraints. Biometrics 46, 201–206.

McDonald, J.W., Prevost, A.T., 1997. The fitting of parameter-constrained demographic models. Math. Comp. Model. 26, 79–88.

Molenberghs, G., Lesaffre, E., 1994. Marginal modeling of correlated ordinal data using a multivariate Plackett distribution. J. Amer. Statist. Assoc. 89, 633–644.

Morris, M.D., 1988. Small-sample confidence limits for parameters under inequality constraints with application to quantal bioassay. Biometrics 44, 1083–1092.

Mukerjee, H., 1988. Order restricted inference in a repeated measures model. Biometrika 75, 616–617.

Nair, V.N., 1987. Chi-squared-type tests for ordered alternatives in contingency tables. J. Amer. Statist. Assoc. 82, 283–291.

Nandram, B., Sedransk, J., Smith, S.J., 1997. Order-restricted Bayesian estimation of the age composition of a population of Atlantic cod. J. Amer. Statist. Assoc. 92, 33–40.

Nguyen, T., Sampson, A.R., 1987. Testing for positive quadrant dependence in ordinal contingency tables. Nav. Res. Log. Quar. 34, 859–877.

Oh, M., 1995. On maximum likelihood estimation of cell probabilities in $2 \times k$ contingency tables under negative dependence restrictions with various sampling schemes. Comm. Statist. Theor. Methods 24, 2127–2143.

Oluyede, B.O., 1993a. A modified chi-square test for testing equality of two multinomial populations against an ordering restricted alternative. Biometrical J. 35, 997–1012.

Oluyede, B.O., 1993b. A modified chi-square test for testing equality of two multinomial populations against an order restricted alternative. Comm. Statist. Theor. Methods 22, 1133–1155.

Oluyede, B.O., 1994a. A modified chi-square test of independence against a class of ordered alternatives in an $r \times c$ contingency table. Canad. J. Statist. 22, 75–87.

Oluyede, B.O., 1994b. Tests for equality of several binomial populations against an order restricted alternative and model selection for one-dimensional multinomials. Biometrical J. 36, 17–32.

Park, C.G., 1998. Testing for unimodal dependence in an ordered contingency table with restricted marginal probabilities. Statist. Prob. Lett. 37, 121–129.

Parsa, A.R., Smith, W.B., 1993. Scoring under ordered constraints in contingency tables. Comm. Statist. Theor. Methods 22, 3537–3551.

Patefield, W.M., 1981. An efficient method of generating random $R \times C$ tables with given row and column totals. J. Roy. Statist. Soc. Ser. C 30, 91–97.

Patefield, W.M., 1982. Exact tests for trends in ordered contingency tables. J. Roy. Statist. Soc. Ser. C 31, 32–43.

Paula, G.A., 1993. Assessing local influence in restricted regression models. Comput. Statist. Data Anal. 16, 63–79.

Paula, G.A., Sen, P.K., 1995. One-sided tests in generalized linear models with parallel regression lines. Biometrics 51, 1494–1501.

Perlman, M.D., Wu, L., 1999. The emperor's new tests. Stat. Sci. 14, 355–369.

Piegorsch, W.W., 1990. One-sided significance tests for generalized linear models under dichotomous response. Biometrics 46, 309–316.

Plackett, R.L., 1965. A class of bivariate distributions. J. Amer. Statist. Assoc. 60, 516–522.
Podgor, M.J., Gastwirth, J.L., Mehta, C.R., 1996. Efficiency robust tests of independence in contingency tables with ordered classifications. Statist. Med. 15, 2095–2105.
Poon, A.H., 1980. A Monte Carlo study of the power of some k-sample tests for ordered binomial alternatives. J. Statist. Comput. Simul. 11, 251–259.
Ramsey, F.L., 1972. A Bayesian approach to bioassay. Biometrics 28, 841–858.
Rao, M.B., Krishnaiah, P.R., Subramanyam, K., 1987. A structure theorem on bivariate positive quadrant dependent distributions and tests for independence in two-way contingency tables. J. Multivariate Anal. 23, 93–118.
Ritov, Y., Gilula, Z., 1991. The order-restricted RC model for ordered contingency tables: Estimation and testing for fit. Ann. Statist. 19, 2090–2101.
Ritov, Y., Gilula, Z., 1993. Analysis of contingency tables by correspondence models subject to order constraints. J. Amer. Statist. Assoc. 88, 1380–1387.
Robertson, T., 1978. Testing for and against an order restriction on multinomial parameters. J. Amer. Statist. Assoc. 73, 197–202.
Robertson, T., 1986. On testing symmetry and unimodality. In: Dykstra, R., Robertson, T., Wright, F.T. (Eds.), Advances in Order Restricted Statistical Inference. Springer, New York, pp. 231–248.
Robertson, T., Wegman, E.G., 1978. Likelihood ratio tests for order restrictions in exponential families. Ann. Statist. 6, 485–505.
Robertson, T., Wright, F.T., 1981. Likelihood-ratio tests for and against a stochastic ordering between multinomial populations. Ann. Statist. 9, 1248–1257.
Robertson, T., Wright, F.T., 1985. One-sided comparisons for treatments with a control. Canad. J. Statist. 13, 109–122.
Robertson, T., Wright, F.T., Dykstra, R.L., 1988. Order Restricted Statistical Inference. Wiley, New York.
Sackrowitz, H., 1982. Procedures for improving the MLE for ordered binomial parameters. J. Statist. Plann. Inference 6, 287–296.
Sampson, A.R., Singh, H., 2002. Min and max scorings for two sample partially ordered categorical data. J. Statist. Plann. Inference 107, 219–236.
Sampson, A.R., Whitaker, L.R., 1989. Estimation of multivariate distributions under stochastic ordering. J. Amer. Statist. Assoc. 84, 541–548.
Schervish, M.J., 1985. Comments on a one-sided goodness-of-fit test for a multinomial population. J. Amer. Statist. Assoc. 80, 562–563.
Schmoyer, R.L., 1984. Sigmoidally constrained maximum likelihood estimation in quantal bioassay. J. Amer. Statist. Assoc. 79, 448–453.
Sedransk, J., Monahan, J., Chiu, H.Y., 1985. Bayesian estimation of finite population parameters in categorical data models incorporating order restrictions. J. Roy. Statist. Soc. Ser. B 47, 519–527.
Shi, N.-Z., 1991. A test of homogeneity of odds ratios against order restrictions. J. Amer. Statist. Assoc. 86, 154–158.
Shin, D.W., Park, C.G., Park, T., 1996. Testing for ordered group effects with repeated measurements. Biometrika 83, 688–694.
Silvapulle, M.J., 1994. On tests against one-sided hypotheses in some generalized linear models. Biometrics 50, 853–858.
Soms, A.P., 1985. Permutation tests for k-sample binomial data with comparisons of exact and approximate p-levels. Comm. Statist. Theor. Methods 14, 217–233.
Takane, Y., Yanai, H., Mayekawa, S., 1991. Relationships among several methods of linearly constrained correspondence analysis. Psychometrika 56, 667–684.
Takeuchi, K., Hirotsu, C., 1982. The cumulative chi-squares method against ordered alternatives in two-way contingency tables. Reports Statist. Appl. Res. 29, 1–13.
Tsai, M.M., 1993. UI score tests for some restricted alternatives in exponential families. J. Multivariate Anal. 45, 305–323.
Wahrendorf, F., 1980. Inference in contingency tables with ordered categories using Plackett's coefficient of association for bivariate distributions. Biometrika 67, 15–21.
Wang, Y., 1996. A likelihood ratio test against stochastic ordering in several populations. J. Amer. Statist. Assoc. 91, 1676–1683.

Wax, Y., Gilula, Z., 1990. A score test for monotonic trend in hazard rates for grouped survival data in stratified populations. Comm. Statist. Theor. Methods 19, 3375–3386.

Wolfe, R.A., Roi, L.D., Margosches, E.H., 1981. Monotonic dichotomous regression estimates: A burn care example. Biometrics 37, 157–167.