A Unified Approach to Likelihood Inference on Stochastic Orderings in a Nonparametric Context

Valentino DARDANONI and Antonio FORCINA

For data in a two-way contingency table with ordered margins, we consider various hypotheses of stochastic orders among the conditional distributions considered by rows and show that each is equivalent to requiring that an invertible transformation of the vectors of conditional row probabilities satisfies an appropriate set of linear inequalities. This leads to the construction of a general algorithm for maximum likelihood estimation under multinomial sampling and provides a simple framework for deriving the asymptotic distribution of log-likelihood ratio tests. The usual stochastic ordering and the so called *uniform* and *likelihood ratio* orderings are considered as special cases. In particular, for each of these three orderings we determine the transformation required to apply the estimation algorithm; we then consider testing the hypothesis that the rows are identically distributed against the alternative that they are stochastically ordered, as well as testing each stochastic order against an unrestricted alternative. We show that in all cases the test statistics are asymptotically distributed as a mixture of chi-squared distributions, with weights determined by the information matrix. By exploiting the special structure of this matrix in these three cases, we find tight upper and lower bounds to the distribution of all test statistics. These bounding distributions are free of nuisance parameters and relatively easy to compute. Two examples are presented to illustrate the methodology and the required computations needed to apply these techniques.

KEY WORDS: Chi-bar-squared distribution; Likelihood inference; Likelihood ratio ordering; Order-restricted inference; Stochastic ordering; Uniform ordering.

1. INTRODUCTION

Stochastic orderings provide, in several applied contexts, the appropriate tools for formalizing the idea that one distribution in some sense attaches more probability to larger values than another. Methods based on comparing location parameters, in contrast, are limited in at least two ways. First, the functional defining a given location parameter defines a notion of magnitude that usually is too specific and need not be implied by the underlying theory, which specifies that one distribution should be *larger* than another. It may also be the case that proper location parameters cannot be computed because observations are available in terms of an ordered categorical variable representing a discretized version of a latent continuous variable. Moreover, in contrast to other nonparametric methods for testing for the equality of one or more distributions against an unrestricted alternative, stochastic orderings allow consideration of different specific one-sided alternatives; these alternatives are defined by simple inequalities on the probability distributions being compared, and these inequalities usually have a meaningful interpretation in various applied contexts. These considerations underscore the importance of statistical procedures designed to detect the occurrence of such orderings on the basis of random samples.

Suppose that observations are available on an ordered discrete variable Y at various settings of an ordered discrete explanatory variable X; the nature of their dependence may be specified by requiring that the conditional distributions of Y given $X = x_i$ satisfy an appropriate stochastic order. There is a large body of literature on the various notions of stochastic orderings and their properties; we refer the reader to the book by Shaked and Shantikumar (1994) for an exhaustive survey. The usual definition of a random variable V being stochastically larger than U, denoted by $V \succeq_s U$, is called *stochastic dominance* and requires that for every real c,

$$\Pr(V \leq c) \leq \Pr(U \leq c).$$

Sometimes it may be appropriate to require that this ordering holds also conditional to U and V belonging to some subset, for all possible subsets within some class of interest. For example, in reliability theory one is interested in knowing whether V has a greater chance of lasting longer than U whatever the amount of time they have already survived; this leads to the notion of *uniform* stochastic ordering. An even stronger restriction is in requiring that V be greater than U when they are known to belong to any given subset of neighboring categories; this leads to the so-called *likelihood ratio* ordering.

If X indexes the rows and Y the columns of a contingency table and we let $U = (Y \mid X = x_{i-1})$ and $V = (Y \mid X = x_i)$, then the stochastic orderings just considered correspond to successively stronger notions of *positive dependence* of Y on X. In particular, the hypothesis that the conditional distribution within each row is *larger* than that in the row above it is equivalent to the hypothesis of *row regression dependence* if we use the stochastic dominance order, to the hypothesis of *hazard rate dependence* if we use the uniform order, and to likelihood ratio dependence if we use the likelihood ratio ordering. Note, however, that these hypotheses often go under different names (for further discussion see Barlow and Proschan 1981; Block, Samp-

Valentino Dardanoni is Associate Professor of Public Finance, Istituto di Scienze Finanziarie, Facoltà di Economia, Università di Palermo, 90132 Palermo, Italy (E-mail: vdardano@weber.ucsd.edu). Antonio Forcina is Professor of Statistics, Dipartimento di Scienze Statistiche, Facoltà di Economia, Università di Perugia, 06100 Perugia, Italy (E-mail: forcina@fisher.stat.unipg.it). The authors thank Ministero dell' Università e della Ricerca Scientifica e Tecnologica for financial support and Valentino Dardanoni also thanks the Department of Economics at the University of California, San Diego for hospitality. The authors also thank an associate editor and two referees for insightful comments.

Journal of the American Statistical Association
September 1998, Vol. 93, No. 443, Theory and Methods

son, and Savits 1990; Jogdeo 1982; Lehmann 1966; Shaked 1977).

In this article we study maximum likelihood estimation and hypothesis testing for a general family of stochastic orderings of which the three orderings considered here are perhaps the main instances. The family is defined by requiring that an invertible transformation of the vector of parameters that determine the conditional distributions of $Y \mid X$ satisfy an appropriate system of linear inequalities. Several results concerning maximum likelihood estimation and hypothesis testing when two or more discrete distributions satisfy a given stochastic order are noted in the literature. For the case of two populations under the usual stochastic order, results were obtained by Grove (1980) and considerably extended by Robertson and Wright (1981). Recently, Wang (1996) has extended these results to the case of comparing more than two distributions and has characterized the asymptotic distribution of the likelihood ratio statistic for testing equality against the stochastic ordering. We show that the actual form of this distribution follows as a special case from our Theorem 2. Dykstra, Kochar, and Robertson (1991) studied the case of several populations satisfying the uniform order, and later Dykstra et al. (1995) considered inference for the likelihood ratio order in two populations. Our results can be considered as the natural extension and unification of these seminal papers.

In Section 2 we show that, after a suitable reparameterization, each of the basic orderings mentioned earlier holds if and only if a certain system of linear inequalities is satisfied. This result leads to the construction, in Section 3, of a general algorithm for maximum likelihood estimation. The proposed algorithm is a constrained version of the so-called *Fisher scoring* algorithm, a quasi-Newton method based on the expected information matrix that also determines the asymptotic distribution of the log-likelihood ratio test statistics. As we show in Section 4, the asymptotic distribution of the log-likelihood ratios to test for or against any of the stochastic orderings belonging to the family just described is of the *chi-bar-squared* type, which involves mixtures of central chi-squared distributions. Because the hypotheses of interest are usually composite, we determine tight bounds to the null distribution of the various test statistics. These bounds are often computationally quite easy to calculate and should make these inference procedures readily accessible. To clarify the computational aspects of the methods discussed in this article and their implications in an applied contest, in Section 5 we briefly discuss two examples.

2. STOCHASTICALLY ORDERED DISTRIBUTIONS

Let $A_1, \ldots, A_{m+1}$ be $m+1$ independent random variables, each taking values in the same set of outcomes $o_1, \ldots, o_{k+1}$ assumed to be completely ordered with

$$p_{ij} = \Pr(A_i = o_j), \qquad i = 1, \ldots, m+1, \qquad j = 1, \ldots, k+1.$$

The probability distribution of A_i is completely determined by the vector $\mathbf{p}_i = (p_{i1}, \cdots, p_{ik})'$, because $p_{i,k+1}$ is equal to $1 - \mathbf{p}_i'\mathbf{1}_k$, where $\mathbf{1}_k$ denotes the $k \times 1$ vector of 1s.

Several stochastic orderings have been proposed to compare the order of magnitude of two or more distributions. We recall the formal definition of three well-known orderings in our discrete setting (see Shaked and Shantikumar 1994 for alternative definitions and further details).

Definition 1. The random variable A_{i+1} is said to dominate A_i according to the simple stochastic ordering, written as $A_{i+1} \succeq_s A_i$, if

$$\sum_{s=1}^{j} p_{is} \geq \sum_{s=1}^{j} p_{i+1,s}, \qquad j = 1, \ldots, k. \tag{1}$$

Definition 2. The random variable A_{i+1} is said to dominate A_i according to the uniform stochastic ordering, written as $A_{i+1} \succeq_u A_i$, if

$$\frac{G_{i+1,j}}{G_{i+1,j-1}} \geq \frac{G_{i,j}}{G_{i,j-1}} \quad \text{for} \quad j = 1, \ldots, k, \tag{2}$$

where $G_{i,j} = 1 - \sum_{s=1}^{j} p_{is}$ denotes the survival function for the ith distribution and $G_{i,0} = 1$.

Definition 3. The random variable A_{i+1} is said to dominate A_i according to the likelihood ratio ordering, written as $A_{i+1} \succeq_r A_i$, if

$$\frac{p_{i,j}}{p_{i,j+1}} \geq \frac{p_{i+1,j}}{p_{i+1,j+1}} \quad \text{for} \quad j = 1, \ldots, k. \tag{3}$$

Denote by H_h the assumption that $[A_1 \preceq_h \cdots \preceq_h A_{m+1}]$ for $h = s, u, r$. In addition, it will be convenient to consider the two extreme situations: H_0: $[A_1 = \cdots = A_{m+1}]$ and H_2: $[A_1, \cdots, A_{m+1}]$ unrestricted. It is worth noting that the following relationship holds among these hypotheses:

$$H_0 \subset H_r \subset H_u \subset H_s \subset H_2.$$

In a contingency table with ordered margins, H_0 corresponds to the case of independence, whereas each hypothesis H_h is equivalent to the requirement that all *generalized odds ratios* of an appropriate type are not smaller than 1 (see, e.g., Agresti 1984, p. 113; Douglas et al. 1990). In particular, it can be easily verified that H_s corresponds to the hypothesis that all *global–local* odds ratios are greater than 1; H_u corresponds to the hypothesis that all *continuation* odds ratios (Fienberg 1980, p. 86) are greater than 1; and H_r corresponds to the hypothesis that all local odds ratios are greater than 1.

We now show that it is possible to express each of H_h, $h = s, u, r$, as a set of linear constraints on a vector of parameters $\boldsymbol{\beta}$, obtained by an appropriate transformation of the vector $\mathbf{p} = (\mathbf{p}_1', \ldots, \mathbf{p}_{m+1}')'$,

$$g(\mathbf{p}) = \mathbf{X}\boldsymbol{\beta}, \quad \text{with} \quad \mathbf{K}\boldsymbol{\beta} \geq \mathbf{0}, \tag{4}$$

where $\mathbf{X}$ and $\mathbf{K}$ are suitable matrices of known constants and the function $g(\mathbf{p})$ is an invertible transformation having continuous first and second derivative and is known as the *link function* in the terminology of McCullagh and Nelder (1989). Note that when an inequality symbol involves vectors or matrices, we mean that the relation is satisfied elementwise.

We first introduce a useful linear operator. Let $\mathbf{T}_s$ be the $s \times s$ upper triangular matrix of 1s; its inverse $\mathbf{T}_s^{-1}$ has 1s on the main diagonal, -1s on the first superdiagonal, and 0s elsewhere. Note that $\mathbf{x}'\mathbf{T}_s$ contains the cumulative sums of the $s \times 1$ vector $\mathbf{x}$, $\mathbf{T}_s^{-1}\mathbf{x}$ contains the differences of each element from the next (with the last element kept unchanged), and $(\mathbf{T}_s^{-1})'\mathbf{x}$ is the vector of differences of each element from the previous one (with the first element unchanged). Finally, we recall the *canonical parameterization* of the multinomial distribution, and so let $\theta_{ij} = \ln(p_{ij}/p_{i,k+1})$ denote the multivariate logistic transformation; this is an invertible mapping with $p_{ij} = \exp(\theta_{ij})/[1 + \sum_j \exp(\theta_{ij})]$.

2.1 The Simple Stochastic Ordering

Let $\mathbf{P} = (\mathbf{p}_1, \ldots, \mathbf{p}_{m+1})'$; then from (1), it follows that H_s holds if and only if all elements in each row of $\mathbf{PT}_k$ are not smaller than the corresponding elements in the row below it or, equivalently, if and only if the first m rows of the matrix $\mathbf{B} = \mathbf{T}_{m+1}^{-1}\mathbf{PT}_k$ are nonnegative (the last row contains $\mathbf{p}_{m+1}$'s cumulative sum). This shows that it is convenient to express the model directly in terms of the $\mathbf{B}$ parameters, so that by applying the row vec operator, (4) takes the form

$$\boldsymbol{\beta} = \text{vec}(\mathbf{B}) = [\mathbf{T}_{m+1}^{-1} \otimes \mathbf{T}_k']\mathbf{p}, \quad \text{with} \quad [\mathbf{I}_{mk}, \mathbf{0}_{mk,k}]\boldsymbol{\beta} \geq \mathbf{0};$$

where $\otimes$ denotes the Kronecker product, $\mathbf{I}_{mk}$ is the mk identity matrix, and $\mathbf{0}_{mk,k}$ is the $mk \times k$ matrix of 0s; note that here $g(\cdot)$ is the identity function.

2.2 The Uniform Stochastic Ordering

We first collect the $m+1$ survival functions into the $m+1 \times k$ matrix $\mathbf{G}$; this is related to $\mathbf{P}$ by the identity

$$\mathbf{G} = \mathbf{1}_{m+1}\mathbf{1}_k' - \mathbf{PT}_k.$$

Note from (2) that we want second-order differences of $\ln(\mathbf{G})$ from the previous row and the previous column and it can be verified that H_u holds if and only if the last m rows of the matrix $\mathbf{B} = (\mathbf{T}_{m+1}^{-1})' \ln(\mathbf{G})\mathbf{T}_k^{-1}$ are nonnegative. By applying again the row vec operator, we obtain

$$\boldsymbol{\beta} = \text{vec}(\mathbf{B}) = (\mathbf{T}_{m+1}^{-1} \otimes \mathbf{T}_k^{-1})' \text{ vec}[\ln(\mathbf{G})],$$

with

$$[\mathbf{0}_{mk,k}, \mathbf{I}_{mk}]\boldsymbol{\beta} \geq \mathbf{0}.$$

To write down the link function, note that $\text{vec}(\mathbf{G}) = \mathbf{1}_{(m+1)k} - (\mathbf{I}_{m+1} \otimes \mathbf{T}_k')\mathbf{p}$, so that $g(\mathbf{p}) = \ln[\mathbf{1}_{(m+1)k} - (\mathbf{I}_{m+1} \otimes \mathbf{T}_k')\mathbf{p}]$.

2.3 The Likelihood Ratio Ordering

First, note that $\ln(p_{i,j}/p_{i,j+1}) = \theta_{i,j} - \theta_{i,j+1}$ for $j < k$ and to $\theta_{i,k}$ for $j = k$. If we then collect the θ_{ij} parameters into the $m+1 \times k$ matrix $\boldsymbol{\Theta}$, from (3) it follows that we need the second-order differences by columns and by row of each element from the next so that H_r holds if and only if the first m rows of the matrix $\mathbf{B} = \mathbf{T}_{m+1}^{-1}\boldsymbol{\Theta}(\mathbf{T}_k^{-1})'$ are nonnegative. By applying the usual row vec operator, we obtain

$$\boldsymbol{\beta} = \text{vec}(\mathbf{B}) = [\mathbf{T}_{m+1}^{-1} \otimes \mathbf{T}_k^{-1}] \text{ vec}(\boldsymbol{\Theta}),$$

with

$$[\mathbf{I}_{mk}, \mathbf{0}_{mk,k}]\boldsymbol{\beta} \geq \mathbf{0}.$$

3. MAXIMUM LIKELIHOOD ESTIMATION

Assume that we have a random sample of n_i independent observations on each of the A_i variables, $i = 1, \ldots, m+1$ and let $n = \sum_i n_i$. Let also $r_i = n_i/n$ be the proportion of observations coming from the ith sample and $\mathbf{r} = (r_1, \ldots, r_m)$. Often the vector $\mathbf{r}$ is a constant fixed by the sample design; alternatively, if the n observations have been sampled independently from the overall set of $m+1$ populations, then $\mathbf{r}$ is an estimate of the relative size of each population and as such represents a vector of *ancillary statistics*. In deriving the asymptotic results, we assume that $\lim_{n\to\infty} r_i > 0$ for all i.

Let n_{ij} be the number of observations sampled from distribution i with outcome o_j. Then $\hat{\mathbf{p}}_i$, the vector of relative frequencies with $\hat{p}_{ij} = n_{ij}/n_i$, is also the unconstrained maximum likelihood (ML) estimate of $\mathbf{p}_i$. From now on, assume that

$$n_i\hat{\mathbf{p}}_i \mid r_i \sim \text{ multinomial}(n_i, \mathbf{p}_i), \qquad i = 1, \cdots, m+1.$$

Under H_0, we have that $\mathbf{p}_1 = \cdots = \mathbf{p}_{m+1} = \mathbf{q}$ (say) and the ML estimate of $\mathbf{q}$ is obtained by simply pooling together the $m+1$ samples,

$$\hat{\mathbf{q}} = \frac{1}{n}\sum_i n_i\hat{\mathbf{p}}_i = \sum_i r_i\hat{\mathbf{p}}_i. \tag{5}$$

The main result of this section is that the ML estimate of $\mathbf{p}$ under each hypothesis H_h can be obtained by first translating the problem into the appropriate $\boldsymbol{\beta}$ parameterization and then solving iteratively a weighted least squares equation with linear inequality constraints.

Remark 1. Because estimates are updated automatically, we must restrict the parameter space so that the elements of $\mathbf{P}$ are nonnegative and the row sums are less than 1. It turns out that we can always write the set of feasible estimates as $\mathcal{B} = \{\boldsymbol{\beta}: \mathbf{R}\boldsymbol{\beta} \geq \mathbf{b}\}$, where $\mathbf{R}$ and $\mathbf{b}$ will depend on the hypothesis H_h under consideration, and incorporate the appropriate ordering constraints.

A similar algorithm for univariate exponential families has been considered by Fahrmeir and Klinger (1994). Both approaches depend on the likelihood function being strictly concave and are closely related to the basic algorithm used in generalized linear models for obtaining unconstrained ML estimates (e.g., McCullagh and Nelder 1989, p. 42).

The block diagonal matrix $n\mathbf{U}(\boldsymbol{\beta}) = n \text{ diag}(r_1\mathbf{W}_1, \ldots, r_{m+1}\mathbf{W}_{m+1})$, where

$$\mathbf{W}_i = \frac{\partial\boldsymbol{\theta}i'}{\partial\mathbf{p}_i} = [\text{diag}(\mathbf{p}_i)^{-1} + \mathbf{11}'/p_{i,k+1}],$$

as shown in the Appendix, is the Fisher information with respect to $\mathbf{p}$. Now we define the invertible matrix $\mathbf{H}(\boldsymbol{\beta})$ by

$$\frac{\partial \mathbf{p}}{\partial \boldsymbol{\beta}'} = \mathbf{H}(\boldsymbol{\beta})\mathbf{X}$$

and let $\mathbf{y}(\boldsymbol{\beta}) = \boldsymbol{\beta} + \mathbf{X}^{-1}\mathbf{H}(\boldsymbol{\beta})^{-1}(\hat{\mathbf{p}} - \mathbf{p})$ denote the *working dependent variable.*

We have the following.

Definition 4. The constrained version of the Fisher scoring algorithm, denoted by CFS, comprises the following steps:

1. Set the starting point $\boldsymbol{\beta}^1 = \mathbf{X}^{-1}\, g(\hat{\mathbf{p}})$.
2. In the sth step, maximize the quadratic function

$$Q_s(\boldsymbol{\beta}) = -\frac{n}{2}[\mathbf{y}(\boldsymbol{\beta}^s) - \boldsymbol{\beta}]'\mathbf{X}'\mathbf{H}(\boldsymbol{\beta}^s)'\mathbf{U}(\boldsymbol{\beta}^s) \times \mathbf{H}(\boldsymbol{\beta}^s)\mathbf{X}[\mathbf{y}(\boldsymbol{\beta}^s) - \boldsymbol{\beta}]$$

   subject to $\boldsymbol{\beta} \in \mathcal{B}$,

3. Iterate until convergence.

Theorem 1. The CFS algorithm converges to $\hat{\boldsymbol{\beta}}_h$, the ML estimate under H_h, $h = s, u, r$.

Proof. This result says that CFS stops if and only if $\boldsymbol{\beta}^s = \hat{\boldsymbol{\beta}}_h$; a detailed proof is given in the Appendix.

The CFS algorithm reduces the estimation of $\boldsymbol{\beta}$ under each of the hypotheses H_h to a quadratic programming problem that can be solved with several efficient algorithms (see, Dykstra 1983; Goldman and Ruud 1993 and references therein). For the actual implementation of the algorithm under the three orderings, the following elements need to be specified: $\mathbf{X}$, $\mathbf{H}(\boldsymbol{\beta})$, $\mathbf{R}$, and $\mathbf{b}$; this is done in the Appendix. Note that the matrix $\mathbf{U}(\boldsymbol{\beta})$ does not depend on the particular ordering being considered.

Theorem 1 provides a unifying framework for ML estimation of the relevant probabilities under the constraints implied by the various orderings and sets the ground for the straightforward derivation of results on hypotheses testing. Specific procedures for ML estimation under H_u and also under H_s and H_r, with $m = 1$ having been proposed by Dykstra et al. (1991 and 1995) and Robertson and Wright (1981); these are obviously more efficient than CFS. Note also that Dykstra and Feltz (1989) and Feltz and Dykstra (1985) (see also Wang 1986) have proposed an iterative algorithm for ML estimation under H_s, and it would be interesting to compare the efficiency properties of these algorithms with CFS. However, efficiency does not seem to be an important issue here because in our experience, even with large tables, CFS is usually very fast.

4. HYPOTHESIS TESTING

The problem of testing whether the observed sample of discrete distributions conforms to a stochastic ordering is an instance of testing inequality constraints; such problems are termed *order-restricted inference, one-sided testing, isotonic regression*, and so on, and the corresponding methods are perhaps not so widely known.

The distribution of the log-likelihood ratio for testing inequality constraints was first obtained by Bartholomew (1959) in the special case of an analysis of variance with ordered alternatives. Kudô (1963) and Perlman (1969) extended these results to a very general context. Shapiro (1988) has given a concise presentation of the general case. Robertson, Wright, and Dykstra (1988) presented a very systematic exposition of results on order-restricted statistical inference. We summarize some key results here.

4.1 The Chi-Bar-Squared Distribution

Let $\mathcal{C}$ be a closed convex cone in $\mathcal{R}^t$, $\mathbf{V}$ a $t \times t$ symmetric and positive definite matrix, and $\hat{\mathbf{y}}_{V,C}$ the projection of a vector $\hat{\mathbf{y}} \in \mathcal{R}^t$ onto $\mathcal{C}$ in the $\mathbf{V}^{-1}$ metric; that is, $\hat{\mathbf{y}}_{V,C}$ is the solution to the problem

$$\min_{\mathbf{y} \in \mathcal{C}} (\hat{\mathbf{y}} - \mathbf{y})'\mathbf{V}^{-1}(\hat{\mathbf{y}} - \mathbf{y}).$$

Using standard properties of projections onto convex cones and their duals, one can show that

$$\|\hat{\mathbf{y}}\|^2 = \|\hat{\mathbf{y}}_{V,C}\|^2 + \|\hat{\mathbf{y}}_{V,C^o}\|^2, \tag{6}$$

where $\mathcal{C}^o$, the dual of $\mathcal{C}$ in the $\mathbf{V}^{-1}$ metric, is defined as the set $\mathcal{C}^o = \{\mathbf{y}^o\colon \mathbf{y}'\mathbf{V}^{-1}\mathbf{y}^o \le 0, \text{ for all } \mathbf{y} \in \mathcal{C}\}$.

Under the assumption that $\hat{\mathbf{y}} \sim \mathrm{N}(\mathbf{0}, \mathbf{V})$, the distribution of the random variable

$$\bar{\chi}^2(\mathbf{V}, \mathcal{C}) = \hat{\mathbf{y}}'_{V,C}\mathbf{V}^{-1}\hat{\mathbf{y}}_{V,C}$$

is well known and depends on the cone $\mathcal{C}$ and the matrix $\mathbf{V}$. Several presentations of basic results on the chi-bar-squared distribution are available in the literature (e.g., Gourieroux, Holly, and Monfort 1982; Raubertas, Nordheim, and Lee 1986; and Shapiro 1988). Here we recall only the notions essential in the statement of our main theorems. Other results on the chi-bar-squared distribution used in the proofs are stated in the Appendix.

The survival function of $\bar{\chi}^2(\mathbf{V}, \mathcal{C})$ is given by

$$\Pr(\bar{\chi}^2(\mathbf{V}, \mathcal{C}) \ge x) = \sum_0^t w_i(\mathbf{V}, \mathcal{C}) \Pr(\chi_i^2 \ge x) \tag{7}$$

where χ_i^2 denotes a chi-squared random variable with i df, $\Pr(\chi_0^2 \ge x) = 0$ for $x > 0$, and $w_i(\mathbf{V}, \mathcal{C})$, $i = 0, 1, \cdots, t$ are nonnegative weights depending on the matrix $\mathbf{V}$ and the cone $\mathcal{C}$ and sum to 1. Though computation of the probability weights $w_i(\mathbf{V}, \mathcal{C})$ is a difficult numerical problem unless t is less than 4 (e.g., Shapiro 1985), reasonably accurate estimates can be easily obtained by Monte Carlo simulations. Note, however, that in the important special case where $\mathbf{V} = \mathbf{I}_t$ and $\mathcal{C} = \mathcal{O}_t$, the positive orthant, the probability weights are distributed as in the symmetric binomial distribution; that is, $w_i(\mathbf{I}_t, \mathcal{O}_t) = 2^{-t}t!/[i!(t-i)!]$, $i = 0, \cdots, t$.

4.2 Asymptotic Distribution of the Test Statistics

A pathbreaking contribution by Chernoff (1954) investigated the asymptotic distribution of the log-likelihood ratio statistics for testing a set of smooth nonlinear constraints when the true value of the parameters to be tested is a boundary point of the sets defining the null and alternative hypotheses. Significant extensions of his results relevant for our problem have been obtained by Kodde and Palm (1986), Perlman (1969), Shapiro (1985), and Wolak (1991). Wolak in particular derived the asymptotic distribution of the test statistics for testing linear and nonlinear inequality constraints in a nonlinear model, and his results are directly applicable to our context.

Let β^0 denote the true population value of β and let $\hat{\beta} = \mathbf{X}^{-1}g(\hat{\mathbf{p}})$ denote its unrestricted ML estimate. A standard application of the central limit theorem and the delta method ensures that the asymptotic distribution of $\sqrt{n}(\hat{\beta}-\beta^0)$ is $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}(\beta^0))$, where $\boldsymbol{\Sigma}(\beta)^{-1} = \mathbf{X'H'UHX}$ is the average Fisher information, with the expectation being taken conditionally to $\mathbf{r}$, the vector of sample proportions.

Let

$$T_{h2} = 2[L(\hat{\beta}) - L(\hat{\beta}_h)]$$

denote the log-likelihood ratio for testing H_h against H_2 for $h = s, u, r$. The true value of β under H_h is contained in the cone defined by the set of inequalities $\mathbf{K}\beta \geq \mathbf{0}$, and the jth constraint is said to be *active* if the jth element of β^0 is 0. Because only the boundary elements of β^0 are relevant to the asymptotic distribution of T_{h2}, let $\mathbf{J}$ be the submatrix of $\mathbf{K}$ obtained by deleting the rows corresponding to nonactive constraints. Thus, $\mathbf{J}$ is an $s \times (m+1)k$ matrix that selects the *active constraints*, and β^0 is a boundary point of $\mathcal{O}_{mk}$ whenever $s > 0$, so that $\mathbf{J}\beta^0 = \mathbf{0}_s$. Then, from the results of Wolak (1991), it follows that

$$\lim_{n\to\infty} \Pr(T_{h2} \geq x \mid \mathbf{J}\beta^0 = \mathbf{0}_s) = \sum_{0}^{s} w_i(\mathbf{J}\boldsymbol{\Sigma}(\beta^0)\mathbf{J}', \mathcal{O}_s^o)\Pr(\chi_i^2 \geq x). \quad (8)$$

Recall now that under H_0, $\hat{\beta}_0 = \mathbf{X}^{-1}g((\mathbf{1}\otimes\hat{\mathbf{q}}))$, and that the asymptotic distribution of $T_{02} = 2[L(\hat{\beta}) - L(\hat{\beta}_0)]$ is χ^2_{mk}, and consider the log-likelihood ratio for testing H_0 against H_h, $h = s, u, r$ given by $T_{0h} = 2[L(\hat{\beta}_h) - L(\hat{\beta}_0)]$. Because of equations (8) and (6), and noting that under H_0, $\mathbf{J} = \mathbf{K}$, it follows that the asymptotic distribution of T_{0h} is given by

$$\lim_{n\to\infty} \Pr(T_{0h} \geq x \mid \mathbf{K}\beta^0 = \mathbf{0}_{mk}) = \sum_{0}^{mk} w_i(\mathbf{K}\boldsymbol{\Sigma}(\beta^0)\mathbf{K}', \mathcal{O}_{mk})\Pr(\chi_i^2 \geq x). \quad (9)$$

4.3 Testing Equality Against a Stochastic Order

When testing the hypothesis of equality of the $m+1$ distributions against a given stochastic ordering, the situation is somewhat complicated by the null hypothesis being composite, with $\mathbf{q}$, the "true" value of $\mathbf{p}_1 = \cdots = \mathbf{p}_{m+1}$, being a nuisance parameter. The main result stated here provides upper and lower bounds to the distribution of T_{0h}, $h = s, u, r$, under the assumption that $\mathbf{q}$ is free to vary within H_0. We also show that in general it is not possible to improve on these bounds; in particular, these bounds are *tight* in the sense that there exist sequences of $\mathbf{q}$ vectors such that the limiting distribution of T_{0h} converge to these bounds. These results make it possible to compute the critical values under the *least favorable distribution* of the test statistics; that is, the distribution with the smallest rejection region.

Theorem 2. Under H_0 and within the set of all possible vectors of positive probabilities $\mathbf{q}$, the asymptotic distributions of the likelihood ratio test statistics T_{0s}, T_{0u}, and T_{0r}, satisfy

$$\bar{\chi}^2(\mathbf{S}_r^{-1}, \mathcal{O}_m) \preceq_s T_{0r} \sim \bar{\chi}^2(\mathbf{S}_r^{-1}\otimes\mathbf{S}_q^{-1}, \mathcal{O}_{mk}) \preceq_s \sum_{i=1}^{k}\bar{\chi}_i^2(\mathbf{S}_r^{-1}, \mathcal{O}_m)$$

$$T_{0u} \sim \sum_{i=1}^{k}\bar{\chi}_i^2(\mathbf{S}_r^{-1}, \mathcal{O}_m)$$

$$\sum_{i=1}^{k}\bar{\chi}_i^2(\mathbf{S}_r^{-1}, \mathcal{O}_m) \preceq_s T_{0s} \sim \bar{\chi}^2(\mathbf{S}_r^{-1}\otimes\mathbf{S}_q, \mathcal{O}_{mk}) \preceq_s \chi^2_{m(k-1)} + \bar{\chi}^2(\mathbf{S}_r^{-1}, \mathcal{O}_m)$$

where $\mathbf{S}_r = \mathbf{T}'_m[\mathrm{diag}(\mathbf{r}) - \mathbf{rr}']\mathbf{T}_m$, $\mathbf{S}_q = \mathbf{T}'_k[\mathrm{diag}(\mathbf{q}) - \mathbf{qq}']\mathbf{T}_k$ and $\bar{\chi}_i^2(\mathbf{S}_r^{-1}, \mathcal{O}_m)$, $i = 1, \cdots, k$, denote k iid $\bar{\chi}^2(\mathbf{S}_r^{-1}, \mathcal{O}_m)$ random variables.

Proof. See the Appendix.

In the Appendix we also show that the weights $w_j(\mathbf{S}_r^{-1}, \mathcal{O}_m)$, which determine the limiting distributions in theorem 2, are the same as the *level probabilities* used in the context of order restricted inference (for an extensive discussion see Robertson et al. 1988). This connection is useful to recognize that the asymptotic distribution of T_{0u} given above is indeed identical to that derived by Dykstra et al. (1991, thm. 3.1). Note also that for fixed $\mathbf{r}$ (the proportion in the row totals), T_{0u} is an asymptotically similar test statistic, contrary to T_{0s} and T_{0r}. It is also worth noticing the following facts stemming from the theorem:

a. By setting $m = 1$, the upper bounds given by Dykstra et al. (1995, eq. 36) and Robertson and Wright (1981, sec. 4.1) in the two-sample problem emerge as special cases of theorem 2,
b. Wang's (1996) conjecture that the asymptotic distribution of T_{0s} is actually chi-bar-squared also in the multisample problem is confirmed,
c. When $k = 1$, the asymptotic distribution of the three test statistics collapses to the same chi-bar-squared distribution $\bar{\chi}^2(\mathbf{S}_r^{-1}, \mathcal{O}_m)$,
d. The lower and upper bounds to T_{0r} and T_{0s} tend to grow further apart as k grows larger.

The weights $w_j(\mathbf{S}_r^{-1}, \mathcal{O}_m)$ depend on $\mathbf{r}$ and when $m > 4$ are generally difficult to evaluate. However, in the special case when the vector $\mathbf{r} = \mathbf{1}_m/(m+1)$, known as *uniform* *margins*, the resulting weights may be easily calculated by a recursive formula discussed by Dykstra et al. (1991, thm. 4) and Robertson et al. (1988, p. 82), among others. In this case exact upper and lower bounds for the critical values are easily computed. Robertson et al. (1988, sec. 3.1) also argued that the critical values computed under the assumption of uniform margins may be used as a rather accurate approximation as long as the ratio between the largest to the smallest margin is not too large relative to the amount of accuracy required.

Global upper and lower bounds are obtained in the following theorem. The distribution in the upper bounds allows the computation of conservative critical values that depend only on the number of populations being compared and on the number of outcomes being considered and are readily calculated.

Theorem 3. Under H_0 and for any pair of vectors of positive probabilities $\mathbf{r}$ and $\mathbf{q}$, the asymptotic distribution of the likelihood ratio test statistics T_{0s}, T_{0u}, and $T_{0,r}$, satisfy

$$\bar{\chi}^2(1, \mathcal{O}_1) \preceq_s T_{0r} \preceq_s \bar{\chi}^2(\mathbf{I}_{mk}, \mathcal{O}_{mk}),$$
$$\bar{\chi}^2(\mathbf{I}_k, \mathcal{O}_k) \preceq_s T_{0u} \preceq_s \bar{\chi}^2(\mathbf{I}_{mk}, \mathcal{O}_{mk}),$$

and

$$\bar{\chi}^2(\mathbf{I}_k, \mathcal{O}_k) \preceq_s T_{0s} \preceq_s \chi^2_{mk-1} + \bar{\chi}^2(1, \mathcal{O}_1).$$

Proof. See the Appendix.

As expected, conditioning on the vector of sample proportions $\mathbf{r}$ (Theorem 2), is better than not (Theorem 3), in terms of sharper bounds. Note that the improvement is greater the greater the number of populations, for $m = 1$, the two sets of bounds coincide.

It is interesting to note that the bounds presented in Theorem 3 involve only the chi-bar-squared distribution with symmetric binomial weights, so that the computation of these bounds is rather trivial. Moreover, because both the binomial and the chi-squared distributions converge to the normal, it follows that if at least one of m, k is large enough, then a normal approximation to the distribution of these test statistics can be used (see Dykstra 1991).

4.4 Testing a Stochastic Order Against No Restriction

The problem of testing for any of the stochastic dominance hypotheses against H_2 entails an even more difficult problem due to the fact that under H_h, $h = s, u, r$, all of the $(m+1)k$ conditional probabilities represented in the vector $\mathbf{p}$ are nuisance parameters. However, it turns out that within the set of distributions allowed by H_h, when $\mathbf{r}$ is allowed to vary a unique least favorable distribution exists, which allows computation of conservative critical values.

Theorem 4. For any $\mathbf{r}, \mathbf{p} \in H_h$, $h = s, u, r$, the asymptotic distribution of the likelihood ratio test statistics $T_{s,2}$, $T_{u,2}$, and $T_{r,2}$, satisfy

$$T_{r2} \preceq_s \chi^2_{mk-1} + \bar{\chi}^2(1, \mathcal{O}_1),$$
$$T_{u2} \preceq_s \chi^2_{(m-1)k} + \bar{\chi}^2(\mathbf{I}_k, \mathcal{O}_k),$$

and

$$T_{s2} \preceq_s \chi^2_{(m-1)k} + \bar{\chi}^2(\mathbf{I}_k, \mathcal{O}_k).$$

Proof. See the Appendix.

Note that theorem 4.2 of Robertson and Wright (1981) emerges as a special case of Theorem 4 for the simple ordering with two populations. There seem to be no results in the literature for testing either the uniform or the likelihood ratio orderings against the unrestricted alternative.

4.5 The Practical Implementation of the Test Statistics

There are two kinds of problems when applying the testing procedures described in this section: computation of the probability weights and handling of nuisance parameters. Whenever exact computations are not possible, the probability weights of the chi-bar-squared distribution must be estimated by a Monte Carlo technique by projecting a reasonable number (say r) of (pseudo) random vectors $\mathbf{x} \sim \mathrm{N}(\mathbf{0}, \mathbf{V})$ (where $\mathbf{V}$ is the appropriate covariance matrix) onto the positive orthant. Let Y have a chi-bar-squared distribution with weights $\mathbf{w}$ and let $\Pr(Y \geq c) = \alpha$ and $\Pr(\chi^2_i \geq c) = c_i$; then by a central limit approximation, $\hat{\mathbf{w}} \sim \mathrm{N}(\mathbf{w}, [\mathrm{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}']/r)$. Thus, starting from a preliminary estimate of $\mathbf{w}$, we can choose r so as to achieve the required level of precision by imposing that $\Pr(|\, \mathbf{c}'\hat{\mathbf{w}} - \alpha \mid /\alpha \leq \lambda)$ be close enough to 1, where λ is a small fraction measuring the relative error that can be allowed.

The presence of nuisance parameters poses problems of a rather different nature depending on whether we are testing equality against a given stochastic ordering (Sec. 4.3) or a given stochastic order against the unrestricted alternative (Sec. 4.4). In the first case, recall that from Theorems 2 and 3 we can derive ranges, bounded by a lower and an upper critical value, with the range provided by Theorem 2 being nested inside the other. Hence a possible strategy would be to use the two sets of critical values in succession and decide that no further analysis is necessary whenever the observed value of the T_{0h} statistic falls outside either range of critical values. This is also computationally simple, because the limiting distributions in Theorem 3 depend on binomial weights that are easily computed. Although the computations in Theorem 2 are not so trivial, the simplified covariance structure allows the weights to be computed either exactly if $m \leq 4$ or by the uniform row approximation if appropriate, or by Monte Carlo estimation as discussed earlier.

We recall that T_{0u} is asymptotically similar, and hence a final conclusion concerning the uniform ordering may be based on the unique critical value provided by Theorem 2. On the other hand, the observed value of T_{0s} and/or T_{0r} could still fall within the bounds derived from Theorem 2. In such a case a conclusion based on the upper bound, though formally correct, may be rather conservative. The structure of the column totals implied by the least favorable distribution requires that certain elements of $\mathbf{q}$ be extremely small relative to others while the evidence against this possibility may be quite strong. If so, one could perform the test *locally*; that is, by replacing the unknown parameters with their ML estimates. The conclusions from such local tests would be asymptotically valid, as explained by, for example, Wolak (1991). In practice one can replace $\mathbf{q}$ with $\hat{\mathbf{q}}$ in (9) and use the expression for the covariance matrix under H_0 for the chosen stochastic order, as given explicitly in the Appendix.

A small simulation experiment was conducted to assess the performance of the local test under H_0. We set $\mathbf{r}$ and $\mathbf{q}$ as in the data set given by Agresti (1984, p. 13), which we analyze in Section 5; 2,000 sets of $m = 4$ independent multinomial samples with probability vector $\mathbf{q}$ (each having size proportional to $\mathbf{r}$) were drawn with replacement. For each observed value of T_{0s} and T_{0r}, two p values based on the asymptotic distribution (9) were computed: one using the true value of $\mathbf{q}$ and the other using $\hat{\mathbf{q}}$. Then each nominal size of interest α was compared with the proportion of p values not exceeding α. Results for α between .005 and .5 are summarized in Table 1. The reassuring result is that even when the overall sample size is not too large, estimation of the nuisance parameters introduces no appreciable distortion. On the other hand, the rate of convergence to the asymptotic distribution, even under the true value $\mathbf{q}$, is possibly of concern; if the sample size is quite small, it would be wise to complement the procedures described in this article with a simple Monte Carlo test.

When testing a given stochastic order against no restrictions, all conditional probabilities in $\mathbf{P}$ are nuisance parameters. Computation of critical values from Theorem 4 involves again binomial weights and thus is straightforward. However, this least favorable distribution is calculated under H_0 (implying that no inequality defining the stochastic order is strict) and by letting the vectors $\mathbf{r}$ and $\mathbf{q}$ approach a very unbalanced structure. Hence the procedure is very conservative. Critical values could be computed again under H_0 but with the observed value of $\mathbf{r}$ and $\hat{\mathbf{q}}$ as discussed earlier. This procedure is still conservative, especially if the true table $\mathbf{P}$ is far from H_0 so that only few elements of $\boldsymbol{\beta}^0$ are on the boundary.

The alternative is a local test obtained by replacing $\boldsymbol{\beta}^0$ with $\hat{\boldsymbol{\beta}}_h$, its ML estimate under H_h. However, the true asymptotic distribution [see (8)] depends heavily on the number of inequalities that hold as equalities in the population while, even with a very large sample size, it is likely that this number will be underestimated; if, say, β_i^0 was 0, there is approximately a 50% chance that its ML estimate be positive. This is why the p values computed from the local test will tend to be larger and the procedure too liberal. This may be a useful feature, however, if the specific structure implied by the null is an appealing conclusion. These features of the local test emerge clearly from Table 2, which summarizes the results of a simulation study similar to the one reported in Table 1. Here, however, the actual significance levels are always much larger than the nominal ones.

Table 1. Absolute Deviations Between Nominal and Actual Error Rates of the Local Test of Equality Against H_h

	n	True test		Local test	
		T_{0s}	T_{0r}	T_{0s}	T_{0r}
Maximum	200	.0255	.0200	.0260	.0190
	4,000	.0095	.0125	.0085	.0110
Average	200	.0056	.0046	.0057	.0045
	4,000	.0026	.0028	.0026	.0028

Table 2. Ratio Between Actual and Nominal Error Rates in the Local Test of H_h Against H_2.

n	Averages H_s	Averages H_u	Averages H_r	Minimum H_s	Minimum H_u	Minimum H_r
200	1.9165	1.9457	1.9245	1.5930	1.5390	1.5290
4,000	1.5997	1.7046	1.5004	1.4000	1.5286	1.2300

4.6 Testing Against a Stochastic Ordering

An intrinsic weakness of the testing procedures described in this section is that if the stochastic order under examination, except for a few violations, held in most cells of a given table, then the chance would be high that H_0 would be rejected against the stochastic order. Moreover, it would also be quite likely that the stochastic order assumption would be retained when testing against no restriction.

A radical solution to this difficulty would be to take as null H_n: *the given stochastic order does not hold.* More precisely, let $\boldsymbol{\gamma} = \mathbf{K}\boldsymbol{\beta}$ be the subset of $\boldsymbol{\beta}$, which needs to be nonnegative for the stochastic order to hold, and define $H_n = \{\boldsymbol{\gamma}\colon \min(\boldsymbol{\gamma}) \leq 0\}$ whose complement, H_a, implies that the stochastic order holds strictly. From the results in Section 4.2, it follows that the unrestricted ML estimate of $\boldsymbol{\gamma}$, $\hat{\boldsymbol{\gamma}} = \mathbf{K}\hat{\boldsymbol{\beta}}$, is such that asymptotically, $\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0) \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Delta})$, where $\boldsymbol{\gamma}^0 = \mathbf{K}\boldsymbol{\beta}^0$ and $\boldsymbol{\Delta} = \mathbf{K}\boldsymbol{\Sigma}(\boldsymbol{\beta}^0)\mathbf{K}'$. Then the results of Sasabuchi (1980) imply that the likelihood ratio procedure for testing H_n against H_a reduces to

$$\text{reject if } \min(\hat{z}_i) > z_\alpha,$$

where $\hat{z}_i = \hat{\gamma}_i/\sqrt{\Delta_{ii}}$, and z_α is the α% critical value from the standard $\mathrm{N}(0, 1)$ distribution.

Although this testing procedure is very easy to apply in practice and gives very good protection against the error of accepting a given stochastic order when it is violated, the actual size of the test is typically much smaller than its nominal level at H_0 and consequently its power in detecting a given stochastic order in a neighborhood of H_0 can be extremely low. This happens expecially when the size of $\boldsymbol{\gamma}$ is large and negative correlations are present in $\boldsymbol{\Delta}$.

Table 3. Hospital Operations According to Severity and Extent of Side Effects

Severity	Side effects: None	Slight	Moderate
A	61	28	7
B	68	23	13
C	58	40	12
D	53	38	16

Table 4. Log-Likelihood Ratio Tests of H_0 Against a Stochastic Order and Corresponding 5% Critical Values

		Critical values				
		Lower bounds			Upper bounds	
Alternatives	T_{0h}	Thm. 3	Thm. 2	Local	Thm. 2	Thm. 3
H_s	10.61	4.2296	6.9124	7.6122	9.8392	11.9091
H_u	9.04	4.2296	6.9124	6.9124	6.9124	8.4088
H_r	8.75	2.7045	4.5115	6.1748	6.9124	8.4088

5. EXAMPLES

As an illustration, we briefly reanalyze two known datasets. In the first dataset, 417 duodenal ulcer patients are classified according to the severity of the operation to which they have been subjected (increasing from A to D) and the extent of side effects.

As noted by Agresti (1984, p. 13), H_0 cannot be rejected, as the p value corresponding to the T_{02} statistics is 10.88. However, as the results from Table 4 seem to indicate, the evidence against H_0 becomes much stronger if we look at specific alternatives such as any of H_s, H_u, or H_r which in this context are the natural explanation for the association in the data.

Note that only for H_s would further computations beyond those of Theorem 3 be necessary to reject the null H_0 at the 5% significance level. The example seems to suggest that when the nature of the data is such that testing H_0 against a stochastic order rather than against H_2 is deemed appropriate, one can achieve a substantial increase in power.

As indicated by the positive value of the T_{h2} statistics in Table 5, the table does not conform exactly to any of the three orderings. However, this is probably due to random fluctuation, because the observed values are always well below the 5% critical values computed locally, which, as argued in Section 4.5, are unfavorable to the hypothesized ordering.

As a second example, consider a well known dataset analyzed by Goodman (1991, p. 1086), among others. The data (Table 6) refer to a sample of British males cross-classified according to the father's occupational status (row categories) and to the son's occupational status (column categories). We have collapsed the first two categories in each classification to have more even margins.

It is easily verified that Table 6 conforms to H_s, so that $T_{s2} = 0$ and $T_{0s} = 839.14$ is equal to the chi-squared statistic. The hypothesis that the simple ordering holds would be accepted at any significance level and irrespective of its actual specification. These results are rather obvious because of the strong association between row and column categories. However, the data do not conform exactly to either H_u or H_r, although $T_{0u} = 831.35$ and $T_{0r} = 829.75$ are well beyond the 1% critical value of 28.59 computed under the least favorable distribution of Theorem 3.

Table 5. Log-Likelihood Ratio Tests of Stochastic Order Against H_2

		5% Critical values	
Null	T_{h2}	Local	Under H_0
H_s	.27	4.41	9.07
H_u	1.84	4.23	9.64
H_r	2.13	4.02	10.14

Table 6. Father (A_j) and Son (O_j) Occupational Status for a Sample of 3,488 British Males

	O_1	O_2	O_3	O_4	O_5	O_6
A_1	125	60	26	49	14	5
A_2	47	65	66	123	23	21
A_3	31	58	110	223	64	32
A_4	50	114	185	715	258	189
A_5	6	19	40	179	143	71
A_6	3	14	32	141	91	106

The value of the test statistic $T_{u2} = 7.79$ indicates that H_u is also accepted against H_2, given that the 5% critical values are equal to 34.72 and 9.29 under the least favorable distribution and under the local hypothesis. Hence sons coming from a better family have a better chance of success not only in general, but also conditional to having already had a certain amount of success. On the other hand, the value of $T_{r2} = 9.39$ is below the 5% critical value computed from the least favorable distribution (37.07), although not below the critical value computed locally (9.31). Hence if we are interested in comparing chances of success conditional on remaining within any given subset of neighboring classes, then there does not seem to be strong evidence in favor of this stronger hypothesis in the table.

5.1 Software Implementation

All of the foregoing results have been computed by a system of Matlab functions that are available to the interested reader. These functions perform the following tasks:

- For a given H_h and a rectangular table, compute the ML estimate $\hat{\boldsymbol{\beta}}_h$ and the test statistics T_{0h} and T_{h2}; our experience with several real and simulated samples shows that this is very fast and usually takes fewer than six iterations to converge.
- Given a $t \times t$ covariance matrix $\mathbf{V}$, estimate the weights $w_j(\mathbf{V}, \mathcal{O}_t)$, $j = 0, \ldots, t$ by projecting onto the positive orthant observations drawn from a $\mathrm{N}(\mathbf{0}, \mathbf{V})$, and counting the proportions of binding constraints. The technique described in Section 4.5 indicates that 2,000 replicates are enough for a reasonable level of accuracy: with a 166 MHz 586 processor this takes about 15 seconds for $t = 8$, and a few minutes for $t = 25$ (this corresponds to a 6×6 table). In the two examples given here, most of the required distributions either had exact binomial weights or the largest t for which weights had to be estimated was 8 (number of active constraints in the local hypothesis).
- Given a vector of weights corresponding to the asymptotic distribution of T_{0h} or T_{h2} and a significance value, search for the corresponding critical value. This function uses the secant method and numerical integration and is usually very fast.

APPENDIX: PROOFS

A.1 Proof of Theorem 1

To show that the first derivatives of $L(\beta)$ and $Q_s(\beta)$ computed at $\beta = \beta^s$ are equal, first express the log-likelihood function in terms of the canonical parameters $\theta_{ij} = \ln(p_{ij}/p_{i,k+1})$, recalling that $n_{ij} = nr_i\hat{p}_{ij}$. Expand the derivative of L by the chain rule and substitute for the expression of $\mathbf{y}(\beta)$ in the derivative of Q_s as follows:

$$L(\beta) = \sum_i \left[\sum_j n_{ij} \ln(p_{ij}) + \left(n_i - \sum_j n_{ij}\right) \ln\left(1 - \sum_j p_{ij}\right)\right] + \text{const}$$

$$= n \sum_i r_i \left\{ \mathbf{p}_i'\boldsymbol{\theta}_i - \ln\left[1 + \sum_j \exp(\theta_{ij})\right]\right\} + \text{const};$$

$$\frac{\partial L}{\partial \beta} = \sum_i \frac{\partial \mathbf{p}_i'}{\partial \beta}\frac{\partial \boldsymbol{\theta}_i'}{\partial \mathbf{p}_i}\frac{\partial L}{\partial \boldsymbol{\theta}_i} = \sum_i \frac{\partial \mathbf{p}_i'}{\partial \beta} n(r_i\mathbf{W}_i)(\hat{\mathbf{p}}_i - \mathbf{p}_i)$$

$$= n\mathbf{X}'\mathbf{H}'\mathbf{U}(\hat{\mathbf{p}} - \mathbf{p});$$

and

$$\frac{\partial Q_s(\beta)}{\partial \beta} = n\mathbf{X}'\mathbf{H}(\beta^s)'\mathbf{U}(\beta^s)\mathbf{H}(\beta^s) \times \mathbf{X}\{\beta^s + \mathbf{X}^{-1}\mathbf{H}(\beta^s)^{-1}[\hat{\mathbf{p}} - \mathbf{p}(\beta^s)] - \beta\}.$$

The result follows by replacing β with β^s. To compute the second derivative of L, let $\mathbf{u}_j$ be a $k \times 1$ vector having 1 in the jth position and 0s elsewhere and proceed step by step:

$$\frac{\partial}{\partial p_{ij}}\left(\frac{\partial L}{\partial \mathbf{p}_i}\right)' = nr_i \left\{ \left[-\frac{1}{p_{ij}^2}\mathbf{u}_j\mathbf{u}_j' + \frac{1}{p_{i,k+1}^2}\mathbf{1}\mathbf{1}'\right](\hat{\mathbf{p}}_i - \mathbf{p}_i) + \mathbf{W}_i(-\mathbf{u}_j)\right\}$$

$$= -nr_i \left[\frac{\hat{p}_{ij}}{p_{ij}^2}\mathbf{u}_j + \frac{\hat{p}_{i,k+1}}{p_{i,k+1}^2}\mathbf{1}\right],$$

$$\frac{\partial}{\partial \mathbf{p}_i}\left(\frac{\partial L}{\partial \mathbf{p}_i}\right)' = -nr_i\left[\operatorname{diag}(\hat{\mathbf{p}}_i)\operatorname{diag}(\mathbf{p}_i)^{-2} + \mathbf{1}\mathbf{1}'\frac{\hat{p}_{i,k+1}}{p_{i,k+1}^2}\right]$$

$$= nr_i\mathbf{V}_i \text{ (say)},$$

and

$$\frac{\partial}{\partial \beta}\left(\frac{\partial L}{\partial \beta}\right)' = -n\mathbf{X}'\mathbf{H}'\operatorname{diag}(r_1\mathbf{V}_1, \ldots, r_{m+1}\mathbf{V}_{m+1})\mathbf{H}\mathbf{X}.$$

It is easily verified that $E(\mathbf{V}_i) = \mathbf{W}_i$; thus Q_s has also the same Fisher information as L.

Under the assumption that the elements of $\mathbf{p}$ are strictly positive, both $\mathbf{U}$ and each $\mathbf{V}_i$ are positive definite, unless any $\hat{\mathbf{p}}_i$ has two or more elements equal to 0, in which case $\mathbf{V}_i$ is nonnegative definite. Thus both L (except in degenerate cases) and Q are strictly concave and have a unique maximum belonging to the convex set $\mathcal{B}$. Because $\mathbf{X}^{-1}g(\hat{\mathbf{p}})$ is also the unconstrained maximum of Q, if this point is contained in $\mathcal{B}$, then the algorithm stops at the first step. Otherwise, β^{s+1} will be the projection onto the nearest face of $\mathcal{B}$ according to the metric defined by $\mathbf{X}'\mathbf{H}(\beta^s)'\mathbf{U}(\beta^s)\mathbf{H}(\beta^s)\mathbf{X}$. Now, unless $\mathbf{X}^{-1}g(\hat{\mathbf{p}})$ belongs to $\mathcal{B}$, $\hat{\beta}_h$ must be on the boundary and such that any face of $\mathcal{B}$ containing $\hat{\beta}_h$ is orthogonal to the direction of steepest ascent determined by the first derivative of L. Recalling that Q has the same first derivative as L, it follows that CFS will not stop until $\hat{\beta}_h$ has been reached.

A.2 The Practical Implementation of Theorem 1

A.2.1 Estimation Under H_S

In this case $\mathbf{X} = \mathbf{T}_{m+1}\otimes(\mathbf{T}_k^{-1})'$ and, because $g(\cdot)$ is the identity, $\mathbf{H}(\beta) = \mathbf{I}_{(m+1)k}$. As concerns $\mathbf{R}$ and $\mathbf{b}$, we have

$$\mathbf{R} = \begin{pmatrix} \mathbf{I}_{mk}\otimes\mathbf{0}_{mk,k} \\ \mathbf{X} \\ -(\mathbf{I}_{m+1}\otimes\mathbf{1}_k{}')\mathbf{X} \end{pmatrix}$$

and

$$\mathbf{b} = \begin{pmatrix} \mathbf{0}_{mk,1} \\ \mathbf{0}_{(m+1)k,1} \\ -\mathbf{1}_{m+1} \end{pmatrix}$$

where the last two blocks correspond to the constraints $p_{ij} \geq 0$ for $i = 1, \cdots, m+1$, $j = 1, \cdots, k$ and $\sum_j p_{ij} \leq 1$.

A.2.2 Estimation Under H_U

Here we have $\mathbf{X} = (\mathbf{T}_{m+1}\otimes\mathbf{T}_k)'$, and the derivative of the link function gives $\mathbf{H}(\beta) = -[\mathbf{I}_{m+1}\otimes(\mathbf{T}_k^{-1})']\operatorname{diag}[\operatorname{vec}(\mathbf{G})]$. Note that because $\operatorname{vec}(\mathbf{G}) = \exp[\mathbf{X}\beta]$, the constraint $1 - \sum_s p_{i,s} = G_{ik} \geq 0$ is always satisfied. Thus, to ensure that the rows of $\mathbf{G}$ are nonincreasing and that $G_{i1} \leq 1$ for all i, we need only set $\ln(\mathbf{G})\mathbf{T}_k^{-1} \leq 0$. After taking the row vec operator, substituting for β, and simplifying, we get

$$\mathbf{R} = \begin{pmatrix} \mathbf{0}_{mk,k}, \mathbf{I}_{mk} \\ -\mathbf{T}_{m+1}'\otimes\mathbf{I}_k \end{pmatrix}$$

and

$$\mathbf{b} = \begin{pmatrix} \mathbf{0}_{mk,1} \\ \mathbf{0}_{(m+1)k,1} \end{pmatrix}.$$

A.2.3 Estimation Under H_r

In this case $\mathbf{X} = \mathbf{T}_{m+1}\otimes\mathbf{T}_k$, and because the link function is simply the canonical parameterization, $\mathbf{H}(\beta)$ is block diagonal, with the ith block equal to $\mathbf{W}_i^{-1}$. This parameterization has the special property that if all elements of $\hat{\boldsymbol{\Theta}}$ are finite, then the corresponding estimates $\hat{\mathbf{P}}$ are always admissible, so $\mathcal{B}$ is defined simply by $[\mathbf{I}_{mk}, \mathbf{0}_{mk,k}]\beta \geq \mathbf{0}$.

A.3 Some Results on the Chi-Bar-Squared Distribution

In the proofs of the theorems herein we use several known properties of the chi-bar-squared distribution. To make these results easily accessible with self-consistent and more compact notation, they are restated here.

For a $r \times t$ matrix $\mathbf{A}$ of rank r, we define $\mathbf{N}$, a $(t-r) \times t$ matrix whose rows span the space orthogonal to that spanned by the rows of $\mathbf{A}$; $\mathcal{C}(\mathbf{A})$, the cone obtained by the set of linear inequalities $\{\mathbf{y}\colon\ \mathbf{A}\mathbf{y} \geq 0\}$; $\bar{\mathcal{C}}(\mathbf{A})$, the cone obtained by the set of linear inequalities involving $\mathbf{A}$ and equalities involving $\mathbf{N}$, $\bar{\mathcal{C}}(\mathbf{A}) = \{\mathbf{y}\colon \mathbf{A}\mathbf{y} \geq \mathbf{0}, \mathbf{N}\mathbf{y} = \mathbf{0}\}$; and the dual of $\mathcal{C}(\mathbf{A})$ in the $\mathbf{V}^{-1}$ metric, $\mathcal{C}^o(\mathbf{A}) = \{\mathbf{y}\colon\ \mathbf{y} = -\mathbf{V}^{-1}\mathbf{A}'\mathbf{z}, \mathbf{z} \geq \mathbf{0}\}$.

Lemma A.1. Let $\mathbf{V}$ be a $t \times t$ positive definite matrix and let $\mathbf{A}$ a $r \times t$ matrix of rank r. The following results hold:

a. $\bar{\chi}^2(\mathbf{V}, \mathcal{C}(\mathbf{A})) = \bar{\chi}^2(\mathbf{V}, \mathcal{C}(-\mathbf{A}))$
b. $\bar{\chi}^2(\mathbf{V}, \mathcal{O}_t) = \bar{\chi}^2(\mathbf{V}^{-1}, \mathcal{O}_t^o)$
c. $\mathcal{C}_1 \subseteq \mathcal{C}_2 \Rightarrow \bar{\chi}^2(\mathbf{V}, \mathcal{C}_1) \preceq_s \bar{\chi}^2(\mathbf{V}, \mathcal{C}_2)$
d. $\bar{\chi}^2(1, \mathcal{O}_1) \preceq_s \bar{\chi}^2(\mathbf{V}, \mathcal{C}) \preceq_s \chi^2_{t-1} + \bar{\chi}^2(1, \mathcal{O}_1)$

e. $\bar{\chi}^2(\mathbf{V}\otimes\mathbf{I}_q, \mathcal{C}(\mathbf{A}\otimes\mathbf{I}_q)) = \bar{\chi}^2(\mathbf{I}_q\otimes\mathbf{V}, \mathcal{C}(\mathbf{I}_q\otimes\mathbf{A})) = \sum_{i=1}^q \bar{\chi}_i^2(\mathbf{V}, \mathcal{C}(\mathbf{A}))$
f. $\bar{\chi}^2(\mathbf{V}, \mathcal{C}(\mathbf{A})) = \chi^2_{t-r} + \bar{\chi}^2(\mathbf{AVA}', \mathcal{O}_r)$
g. $\bar{\chi}^2(\mathbf{V}, \mathcal{C}^o(\mathbf{A})) = \bar{\chi}^2(\mathbf{AVA}', \mathcal{O}_r^o)$
h. $\bar{\chi}^2(\mathbf{V}, \bar{\mathcal{C}}(\mathbf{A})) = \bar{\chi}^2(\mathbf{AVA}', \mathcal{O}_{t-r})$.

Proof. (a) and (c) are trivial; (d) was proved by Wolak (1991). All of the other results can be easily derived from those discussed by Shapiro (1988). Dobler (1994) presented a useful discussion of polyhedral cones and their duals that may aid derivation of some of these results.

A.4 Preliminaries to the Proofs of Theorems 2, 3, and 4

A.4.1 The Covariance Matrix $\mathbf{S}_f$

For a given discrete probability distribution of dimension $t+1$, denote by $\mathbf{f}$ the vector containing the first t elements, let $\mathbf{c}_f = \mathbf{T}_t'\mathbf{f}$ be the cumulative distribution and let $\boldsymbol{\rho}_f$ be the $t\times 1$ vector with $\rho_1 = c_1/(1-c_1)$ and $\rho_i = c_i/(1-c_i) - c_{i-1}/(1-c_{i-1})$, $i = 2, \cdots, t$. Note that if the elements of $\mathbf{f}$ are strictly positive, then $\mathbf{c}_f$ is strictly increasing and $\boldsymbol{\rho}_f > \mathbf{0}$.

Let $\mathbf{S}_f$ be the covariance matrix of $\mathbf{c}_f$; this matrix has interesting properties that are crucial for deriving our results. It can be verified that

$$\mathbf{S}_f = \mathbf{T}_t'\boldsymbol{\Omega}_f\mathbf{T}_t = \text{diag}(\mathbf{1}_t - \mathbf{c}_f)\mathbf{T}_t'\text{diag}(\boldsymbol{\rho}_f)\mathbf{T}_t\text{diag}(\mathbf{1}_t - \mathbf{c}_f),$$

where $\boldsymbol{\Omega}_f$ denotes $[\text{diag}(\mathbf{f}) - \mathbf{ff}']$. The second expression provides a useful identity for $\text{diag}(\boldsymbol{\rho}_f)$ and an explicit form for the Choleski decomposition: $\mathbf{S}_f = \mathbf{L}_f\mathbf{L}_f'$, where $L_f(ij) = (1-c_i)\sqrt{\rho_j} > 0$ for $i \geq j$ and $L_f(ij) = 0$ otherwise. Thus for any vector $\mathbf{f}$, $\mathbf{S}_f$ will have positive correlations, and we can also show that there exist sequences of probability vectors such that the two extreme cases of zero and unitary correlations are obtained in the limit.

Let $\mathbf{e}_t$ be the $t \times 1$ vector with 1 as the first element and 0s otherwise, so that $\mathcal{C}(\mathbf{e}_t')$ defines the half space $\{\mathbf{y}: y_1 \geq 0\}$. An important consequence of the properties of the Cholesky decomposition $\mathbf{L}_f$ is

$$\mathcal{O}_t \subset \mathcal{C}(\mathbf{L}_f) \subset \mathcal{C}(\mathbf{e}_t'), \tag{A.1}$$

which follows from the fact that $\mathbf{L}_f$ is lower triangular with positive elements. Moreover, we can show that there exist sequences of $\mathbf{f}$ vectors such that the cone $\mathcal{C}(\mathbf{L}_f)$ converges to either $\mathcal{O}_t$ or $\mathcal{C}(\mathbf{e}_t')$.

To show convergence to $\mathcal{O}_t$, let $\mathbf{f}_o = (1-\varepsilon, \varepsilon - \varepsilon^2, \ldots, \varepsilon^{t-1} - \varepsilon^t)'$. Then $\boldsymbol{\rho}_{fo} = (1/\varepsilon - 1, 1/\varepsilon^2 - 1, \ldots, 1/\varepsilon^t - 1)'$ and for $i > j$, $L_{ij}/L_{ii} = \sqrt{(1/\varepsilon^j - 1)/(1/\varepsilon^i - 1)} \to 0$ as $\varepsilon \to 0$, and thus $\mathcal{C}(\mathbf{L}_{fo}) \to \mathcal{O}_t$.

For the second part, let $\mathbf{f}_1 = (1/2)[1-(t-1)\varepsilon, \varepsilon, \ldots, \varepsilon]'$; then $\mathbf{c}_{f1} \to (1/2)\mathbf{1}_t$, $\boldsymbol{\rho}_{f1} \to \mathbf{e}_t$ and $\mathbf{L}_{f1} \to (1/2)\mathbf{1}_t\mathbf{e}_t'$, so that $\mathcal{C}(\mathbf{L}_{f1} \to \mathcal{C}(\mathbf{e}_t')$.

Finally, using a well-known formula for inverting the covariance matrix of the multinomial distribution (e.g., Graybill 1983, p. 189) and some additional algebra, the following useful result may also be established:

$$\mathbf{S}_f^{-1} = \mathbf{T}_t^{-1}[\text{diag}(\mathbf{f})^{-1} + \mathbf{11}'/f_{k+1}]\mathbf{T}_t^{-1'} = \mathbf{D}_t\text{diag}(\dot{\mathbf{f}})^{-1}\mathbf{D}_t',$$

where $\mathbf{D}_t = (\mathbf{I}_t\colon \mathbf{0})\mathbf{T}_{t+1}^{-1} = (\mathbf{0}\colon \mathbf{I}_t)(\mathbf{T}_{t+1}^{-1})'$ is a $t \times t+1$ difference operator and $\dot{\mathbf{f}} = (f_1, \ldots, f_{t+1})$.

A.5 Covariance Matrices Under H_0

Given $\mathbf{p} = \mathbf{1}\otimes\mathbf{q}$, the matrices $\mathbf{U}$ and $\mathbf{H}$ under H_0 take a considerably simpler form:

$$\mathbf{U} = \text{diag}(\dot{\mathbf{r}})\otimes\boldsymbol{\Omega}_q^{-1},$$

$$\mathbf{H}(\text{under } H_u) = -\mathbf{I}_{m+1}\otimes[(\mathbf{T}_k^{-1})'\text{diag}(\mathbf{1}_k - \mathbf{c}_q)],$$

and

$$\mathbf{H}(\text{under } H_r) = \mathbf{I}_{m+1}\otimes\boldsymbol{\Omega}_q.$$

Using these tools, we compute the covariance matrix $\mathbf{K\Sigma}(\beta^0)\mathbf{K}'$ for each hypothesis H_h, $h = s, u, r$:

- Simple ordering:
$$[(\mathbf{I}_m\colon \mathbf{0})\otimes\mathbf{I}_k][\mathbf{T}_{m+1}^{-1}\otimes\mathbf{T}_k'][\text{diag}(\dot{\mathbf{r}})^{-1}\otimes\boldsymbol{\Omega}_q][(\mathbf{T}_{m+1}^{-1})'\otimes\mathbf{T}_k][(\mathbf{I}_m\colon \mathbf{0})\otimes\mathbf{I}_k]' = [(\mathbf{I}_m\colon \mathbf{0})\mathbf{T}_{m+1}^{-1}\text{diag}(\dot{\mathbf{r}})^{-1}(\mathbf{T}_{m+1}^{-1})'(\mathbf{I}_m\colon \mathbf{0})'] \otimes[\mathbf{T}_k'\boldsymbol{\Omega}_q\mathbf{T}_k] = \mathbf{S}_r^{-1}\otimes\mathbf{S}_q$$
- Uniform ordering:
$$[(\mathbf{0}\colon \mathbf{I}_m)\otimes\mathbf{I}_k][\mathbf{T}_{m+1}^{-1}\otimes\mathbf{T}_k^{-1}]'[\mathbf{I}_{m+1}\otimes(\text{diag}(\mathbf{1}-\mathbf{c}_q)^{-1}\mathbf{T}_k')] [\text{diag}(\dot{\mathbf{r}})^{-1}\otimes\boldsymbol{\Omega}_q][\mathbf{I}_{m+1}\otimes(\mathbf{T}_k\text{diag}(\mathbf{1}-\mathbf{c}_q)^{-1})][\mathbf{T}_{m+1}^{-1}\otimes\mathbf{T}_k^{-1}] [(\mathbf{0}\colon \mathbf{I}_m)'\otimes\mathbf{I}_k] = [(\mathbf{0}\colon \mathbf{I}_m)(\mathbf{T}_{m+1}^{-1})'\text{diag}(\dot{\mathbf{r}})^{-1}\mathbf{T}_{m+1}^{-1}(\mathbf{0}\colon \mathbf{I}_m)'] \otimes[(\mathbf{T}_k^{-1})'\text{diag}(\mathbf{1}-\mathbf{c}_q)^{-1}\mathbf{T}_k'\boldsymbol{\Omega}_q\mathbf{T}_k\text{diag}(\mathbf{1}-\mathbf{c}_q)^{-1}\mathbf{T}_k^{-1}] = \mathbf{S}_r^{-1}\otimes\text{diag}(\boldsymbol{\rho}_q)$$
- Likelihood ratio ordering:
$$[(\mathbf{I}_m\colon \mathbf{0})\otimes\mathbf{I}_k][\mathbf{T}_{m+1}^{-1}\otimes\mathbf{T}_k^{-1}][\mathbf{I}_{m+1}\otimes\boldsymbol{\Omega}_q^{-1}][\text{diag}(\dot{\mathbf{r}})^{-1}\otimes\boldsymbol{\Omega}_q] [\mathbf{I}_{m+1}\otimes\boldsymbol{\Omega}_q^{-1}][\mathbf{T}_{m+1}^{-1}\otimes\mathbf{T}_k^{-1}]'[(\mathbf{I}_m\colon \mathbf{0})'\otimes\mathbf{I}_k] = [(\mathbf{I}_m\colon \mathbf{0})\mathbf{T}_{m+1}^{-1}\text{diag}(\dot{\mathbf{r}})^{-1}(\mathbf{T}_{m+1}^{-1})'(\mathbf{I}_m\colon \mathbf{0})'] \otimes[\mathbf{T}_k^{-1}\boldsymbol{\Omega}_q^{-1}(\mathbf{T}_k^{-1})'] = \mathbf{S}_r^{-1}\otimes\mathbf{S}_q^{-1}.$$

A.6 Proof of Theorem 2

Start from (9) and substitute the covariance matrix $\mathbf{K\Sigma}(\beta^0)\mathbf{K}'$ in each hypothesis. For the simple stochastic ordering, after applying Lemma A.1f we get

$$T_{0s} \sim \bar{\chi}^2(\mathbf{S}_r^{-1}\otimes\mathbf{S}_q, \mathcal{O}_{mk}) = \bar{\chi}^2(\mathbf{S}_r^{-1}\otimes\mathbf{I}_k, \mathcal{C}(\mathbf{I}_m\otimes\mathbf{L}_q)).$$

Using equation (A.1), by Lemma A.1c we have

$$\bar{\chi}^2(\mathbf{S}_r^{-1}\otimes\mathbf{I}_k, \mathcal{O}_{mk}) \preceq_s T_{0s} \preceq_s \bar{\chi}^2(\mathbf{S}_r^{-1}\otimes\mathbf{I}_k, \mathcal{C}(\mathbf{I}_m\otimes\mathbf{e}_k')).$$

By Lemma A.1e, the lower limit collapses to $\sum_1^k \bar{\chi}_i^2(\mathbf{S}_r^{-1}\mathcal{O}_m)$. Again apply Lemma A.1f to the upper limit and note that

$$(\mathbf{I}_m\otimes\mathbf{e}_k')(\mathbf{S}_r^{-1}\otimes\mathbf{I}_k)(\mathbf{I}_m\otimes\mathbf{e}_k) = \mathbf{S}_r^{-1},$$

thus establishing the bounds for T_{0s}.

For T_{0u} apply Lemma A.1f to get

$$\begin{aligned} T_{0u} &\sim \bar{\chi}^2(\mathbf{S}_r^{-1}\otimes\text{diag}(\boldsymbol{\rho}_q), \mathcal{O}_{mk}) \\ &= \bar{\chi}^2(\mathbf{S}_r^{-1}\otimes\mathbf{I}_k, \mathcal{C}(\mathbf{I}_m\otimes\text{diag}(\boldsymbol{\rho}_q^{1/2}))) \\ &= \bar{\chi}^2(\mathbf{S}_r^{-1}\otimes\mathbf{I}_k, \mathcal{O}_{mk}), \end{aligned}$$

because for any $\mathbf{a} \geq \mathbf{0}_k$ the cone $\mathcal{C}(\text{diag}(\mathbf{a})) = \mathcal{O}_k$; the result follows again by Lemma A.1e.

As regards T_{0r}, apply Lemma A.1f to get

$$T_{0r} \sim \bar{\chi}^2(\mathbf{S}_r^{-1}\otimes\mathbf{S}_q^{-1}, \mathcal{O}_{mk}) = \bar{\chi}^2(\mathbf{S}_r^{-1}\otimes\mathbf{I}_k, \mathcal{C}(\mathbf{I}_m\otimes(\mathbf{L}_q^{-1})')).$$

From standard properties of cones and their duals, equation (A.1) implies $\bar{\mathcal{C}}(\mathbf{e}_k') \subset \mathcal{C}((\mathbf{L}_q^{-1})') \subset \mathcal{O}_k$, with $\bar{\mathcal{C}}(\mathbf{e}_k')$ defining the half line $\{\mathbf{y}: y_1 \geq 0, y_i = 0,\ i = 2, \cdots, k\}$. Thus, by Lemma A.1c,

$$\bar{\chi}^2(\mathbf{S}_r^{-1}\otimes\mathbf{I}_k, \bar{\mathcal{C}}(\mathbf{I}_m\otimes\mathbf{e}_k')) \preceq_s T_{0r} \preceq_s \bar{\chi}^2(\mathbf{S}_r^{-1}\otimes\mathbf{I}_k, \mathcal{O}_{mk}).$$

For the lower bound, apply Lemma A.1h, whereas the upper bound reduces, as before, to T_{0u}.

Finally, use the vectors $\mathbf{f}_o$ and $\mathbf{f}_1$ defined earlier in place of $\mathbf{q}$, to establish the tightness of the lower and upper bound to T_{0s} and of the upper and lower bound to T_{0r}.

Connection Between Probability Weights and Level Probabilities. Recall that $\mathbf{S}_r^{-1}$ can be written as $\mathbf{D}_m \text{diag}(\dot{\mathbf{r}})^{-1}\mathbf{D}_m'$, and apply Lemma A.1f as

$$\chi_1^2 + \bar{\chi}^2(\mathbf{S}_r^{-1}, \mathcal{O}_m) = \bar{\chi}^2(\text{diag}(\dot{\mathbf{r}})^{-1}, \mathcal{C}(\mathbf{D}_m)),$$

which implies that $w_j(\mathbf{S}_r^{-1}, \mathcal{O}_m) = w_{j+1}(\text{diag}(\dot{\mathbf{r}})^{-1}, \mathcal{C}(\mathbf{D}_m))$, which is the probability that the projection of a random vector $\mathbf{y} \sim \text{N}(\mathbf{0}_{m+1}, \text{diag}(\dot{\mathbf{r}})^{-1})$ on the cone $\mathcal{C}(\mathbf{D}_m)) = \{\mathbf{x}: x_1 \geq x_2 \cdots \geq x_{m+1}\}$ has exactly j distinct values, so that $w_{j+1}(\text{diag}(\dot{\mathbf{r}})^{-1}, \mathcal{C}(\mathbf{D}_m)) = P(j, m+1)$, the level probabilities as defined by, for instance, in Barlow, Bartholomew, Bremner, and Brunk (1972) and Robertson et al. (1988).

A.7 Proof of Theorem 3

All the bounding distributions in Theorem 2 depend on the random variable $\bar{\chi}^2(\mathbf{S}_r^{-1}, \mathcal{O}_m)$ which, as shown earlier, is related to the distribution $\bar{\chi}^2(\text{diag}(\dot{\mathbf{r}})^{-1}, \mathcal{C}(\mathbf{D}_m))$ studied in detail by Robertson and Wright (1982), so that the bounds in Theorem 3 (and their tightness) can be deduced from their results.

As a simple alternative proof, note that, by Lemmas A.1b and A.1g,

$$\bar{\chi}^2(\mathbf{S}_r^{-1}, \mathcal{O}_m) = \bar{\chi}^2(\mathbf{S}_r, \mathcal{O}_m^o) = \bar{\chi}^2(\mathbf{I}_m, \mathcal{C}^o(\mathbf{L}_r)).$$

Then, using again the fact that $\bar{\mathcal{C}}(\mathbf{e}_m') \subset \mathcal{C}((\mathbf{L}_r^{-1})') \subset \mathcal{O}_m$, by Lemma A.1c, and noting that $\bar{\chi}^2(\mathbf{I}_m, \bar{\mathcal{C}}(\mathbf{e}_m')) = \bar{\chi}^2(1, \mathcal{O}_1)$,

$$\bar{\chi}^2(1, \mathcal{O}_1) \preceq_s \bar{\chi}^2(\mathbf{S}_r^{-1}, \mathcal{O}_m) \preceq_s \bar{\chi}^2(\mathbf{I}_m, \mathcal{O}_m). \tag{A.2}$$

Note also that $\sum_1^k \bar{\chi}_i^2(\mathbf{I}_m, \mathcal{O}_m) = \bar{\chi}^2(\mathbf{I}_{mk}, \mathcal{O}_{mk})$.

The tightness of the bounds is established by letting $\mathbf{r}$ equal to $\mathbf{f}_o$ and $\mathbf{f}_1$ to approximate the upper and lower bound in equation (A.2).

A.8 Proof of Theorem 4

Start from (8) and note that in the case of T_{s2}, $\mathbf{K\Sigma}(\beta^0)\mathbf{K}'$ may be written as

$$(\mathbf{D}_m \otimes \mathbf{I}_k)(\text{diag}(\dot{\mathbf{r}})^{-1} \otimes \mathbf{I}_k)\text{diag}(\mathbf{S}_{p1}, \ldots \mathbf{S}_{p,m+1})(\mathbf{D}_m' \otimes \mathbf{I}_k) = \mathbf{A}(\text{diag}(\dot{\mathbf{r}})^{-1} \otimes \mathbf{I}_k)\mathbf{A}',$$

where $\mathbf{A} = (\mathbf{D}_m \otimes \mathbf{I}_k)\text{diag}(\mathbf{L}_{p1}, \ldots, \mathbf{L}_{p,m+1})$. Now apply Lemmas A.1g and A.1c in view of the fact that $\mathcal{C}(\mathbf{A}) \subset \mathcal{C}(\mathbf{JA})$ implies $\mathcal{C}^o(\mathbf{JA}) \subset \mathcal{C}^o(\mathbf{A})$:

$$T_{s2} \sim \bar{\chi}^2(\text{diag}(\dot{\mathbf{r}})^{-1} \otimes \mathbf{I}_k, \mathcal{C}^o(\mathbf{JA})) \preceq_s \bar{\chi}^2(\text{diag}(\dot{\mathbf{r}})^{-1} \otimes \mathbf{I}_k, \mathcal{C}^o(\mathbf{A})).$$

The relation $\mathcal{C}^o(\mathbf{e}_m') \subset \mathcal{C}(\mathbf{D}_m)$ implies $\mathcal{C}^o(\mathbf{D}_m) \subset \mathcal{C}(\mathbf{e}_m')$ and this extends to the product of $\mathbf{D}_m \otimes \mathbf{I}_k$ by a block diagonal matrix with nonnegative entries, so, by Lemmas A.1c and A.1f we have

$$T_{s2} \preceq_s \bar{\chi}^2(\text{diag}(\dot{\mathbf{r}})^{-1} \otimes \mathbf{I}_k, \mathcal{C}(\mathbf{e}_m' \otimes \mathbf{I}_k)) = \chi^2_{(m-1)k} + \bar{\chi}^2(\mathbf{I}_k, \mathcal{O}_k).$$

In the case of T_{u2}, write $\mathbf{K\Sigma}(\beta^0)\mathbf{K}'$ as

$$(\mathbf{D}_m \otimes \mathbf{I}_k)(\text{diag}(\dot{\mathbf{r}})^{-1} \otimes \mathbf{I}_k)\text{diag}[\text{diag}(\boldsymbol{\rho}_1), \ldots \text{diag}(\boldsymbol{\rho}_{p,m+1})] \times (\mathbf{D}_m' \otimes \mathbf{I}_k) = \mathbf{A}(\text{diag}(\dot{\mathbf{r}})^{-1} \otimes \mathbf{I}_k)\mathbf{A}',$$

with $\mathbf{A} = (\mathbf{D}_m \otimes \mathbf{I}_k)\text{diag}[\text{diag}(\boldsymbol{\rho}_1), \ldots \text{diag}(\boldsymbol{\rho}_{p,m+1})]^{1/2}$. Because each $\boldsymbol{\rho}_{pi}$ has positive entries, all of the arguments used in the previous case carry through, and thus T_{u2} has the same upper bound as T_{s2}.

Finally, to establish the upper bound to T_{r2}, simply use Lemma A.1d.

The tightness of these bounds can be derived by appealing to Theorem 3, which established the tightness of the lower bounds to T_{0h} for $h = s, u, r$. In particular, the duality equation (6) and equations (8) and (9) imply that under H_0, the following relationship between T_{0h} and T_{2h} holds:

$$T_{2h} \sim \chi^2_{mk} - T_{0h}, \quad \text{for} \quad h = s, u, r.$$

Thus, given $H_0 \in H_h$, $h = s, u, r$, the same choice of the vectors $\mathbf{f}_o$ and $\mathbf{f}_1$, replacing $\mathbf{q}$ and $\mathbf{r}$ appropriately where needed, suffices to show that the bounds considered in this theorem are tight, noting that

$$\chi^2_{mk} - \bar{\chi}^2(\mathbf{I}_k, \mathcal{O}_k) = \chi^2_{(m-1)k} + \bar{\chi}^2(\mathbf{I}_k, \mathcal{O}_k)$$

and

$$\chi^2_{mk} - \bar{\chi}^2(1, \mathcal{O}_1) = \chi^2_{mk-1} + \bar{\chi}^2(1, \mathcal{O}_1).$$

[Received January 1997. Revised January 1998.]

REFERENCES

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

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

Barlow, R. E., and Proschan, F. (1979), *Statistical Theory of Reliability and Life Testing*, New York: Holt, Rinehart and Winston.

Bartholomew, D. J. (1959), "A Test of Homogeneity for Ordered Alternatives," *Biometrika*, 46, 36–48.

Block, H. W., Sampson, A. R., and Savits, T. H. (1990), *Topics in Statistical Dependence*, IMS Lecture Notes-Monograph Series, Vol. 16, Hayward, CA: Institute for Mathematical Statistics.

Chernoff, H. (1954) "On the Distribution of the Likelihood Ratio," *Annals of Mathematical Statistics*, 25, 573–578.

Dobler, C. P. (1994), "A Matrix Approach to Finding a Set of Generators and Finding the Polar (Dual) of a Class of Polyhedral Cones," *SIAM Journal of Matrix Analysis and Applications*, 15, 796–803.

Douglas, R., Fienberg, S. E., Lee, M. T., Sampson, A. R., and Whitaker, L. R. (1990), "Positive Dependence Concepts for Ordinal Contingency Tables," in *Topics in Statistical Dependence*, eds. H. W. Block, A. R. Sampson, and T. H. Savits, pp. 189–202 Hayward, CA: Institute for Mathematical Studies.

Dykstra, R. L. (1983), "An Algorithm for Restricted Least Squares," *Journal of the American Statistical Association*, 78, 837–842.

——— (1991), "Asymptotic Normality for Chi-Bar-Square Distributions," *Canadian Journal of Statistics*, 19, 297–306.

Dykstra, R. L., and Feltz, C. J. (1989), "Nonparametric Maximum Likelihood Estimation of Survival Functions With a General Stochastic Ordering and Its Dual," *Biometrika*, 76, 331–341.

Dykstra, R. L., Kochar, S., and Robertson, T. (1991), "Statistical Inference for Uniform Stochastic Ordering in Several Populations," *The Annals of Statistics*, 19, 870–888.

——— (1995), "Inference for Likelihood Ratio Ordering in the Two-Sample Problem," *Journal of the American Statistical Association*, 90, 1034–1040.

Fahrmeir, L., and J. Klinger (1994), "Estimating and Testing Generalized Linear Models Under Inequality Restrictions," mimeo.

Feltz, C. J., and Dykstra, R. L. (1985), "Maximum Likelihood Estimation of the Survival Function of N Stochastically Ordered Random Variables," *Journal of the American Statistical Association*, 80, 1012–1019.

Fienberg, S. E. (1980), *The Analysis of Cross-Classsified Categorical Data*, (2nd ed.), Cambridge, MA: MIT Press.

Goldman, S. M. and Ruud, P. A. (1993), "Nonparametric Multivariate Regression Subject to Constraint," mimeo.

Goodman, L. (1991), "Measures, Models and Graphical Displays in the Analysis of Cross-Classified Data," *Journal of the American Statistical Association*, 86, 1085–1111.

Gourieroux, C., Holly, A., and Monfort A. (1982), "Likelihood Ratio Test, Wald Test, and Kuhn–Tucker Test in Linear Models With Inequality Constraints in the Regression Parameters," *Econometrica*, 50, 63–80.

Graybill, F. A. (1983), *Matrices With Applications in Statistics*, Belmont, CA: Wadsworth.

Grove, D. M. (1980), "A Test of Independence Against a Class of Ordered Alternatives in a 2 × C Contingency Table," *Journal of the Americal Statistical Association*, 75, 454–459.

Jogdeo, K. (1982), "Concepts of Dependence," in *Encyclopedia of Statistical Sciences*, eds. S. Kotz, and N. L. Johnson, New York: Wiley.

Lehmann, E. L. (1959), *Testing Statistical Hypothesis*, New York: Wiley.

——— (1966), "Some Concepts of Dependence," *Annals of Mathematical Statistics*, 37, 1137–1156.

Kodde, D. A., and Palm, F. C. (1986), "Wald Criteria for Jointly Testing Equality and Inequality Restrictions," *Econometrica*, 54, 1243–1248.

Kudô, A. (1963), "A Multivariate Analogue of the One-Sided Test," *Biometrika*, 50, 403–418.

McCullagh, P., and Nelder, J.A. (1989), *Generalized Linear Models*, New York: Chapman and Hall.

Perlman, M. D. (1969), "One-Sided Testing Problems in Multivariate Analysis," *Annals of Mathematical Statistics*, 40, 549–567.

Raubertas, R. F., Nordheim, E. V., and Lee, C. C. (1986), "Hypothesis Test for Normal Means Constrained by Linear Inequalities," *Communications in Statistics; Part A—Theory and Methods*, 15, 2809–2833.

Robertson, T. and Wright, F. T. (1981), "Likelihood Ratio Tests for and Against Stochastic Ordering Between Multinomial Populations," *The Annals of Statistics*, 9, 1248–1257.

——— (1982), "Bounds on Mixtures of Distributions Arising in Order Restricted Inference," *The Annals of Statistics*, 10, 302–306.

Robertson, T., Wright, F. T., and Dykstra, R. (1988), *Order Restricted Statistical Inference*, New York: Wiley.

Sasabuchi, S. (1980), "A Test of a Multivariate Normal Mean With Composite Hypotheses Determined by Linear Inequalities," *Biometrika*, 67, 429–439.

Shaked, M. (1977), "A Family of Concepts of Dependence for Bivariate Distributions," *Journal of the Americal Statistical Association*, 72, 642–650.

Shaked, M., and Shanthikumar, J.G. (1994), *Stochastic Orders and Their Applications*, San Diego: Academic Press.

Shapiro, A. (1985), "Asymptotic Distribution of Test Statistics in the Analysis of Moment Structures Under Inequality Constraints," *Biometrika*, 72, 133–144.

Shapiro, A. (1988), "Towards a Unified Theory of Inequality Constrained Testing in Multivariate Analysis," *International Statistical Review*, 56, 49–62.

Wang, Y. (1996), "A Likelihood Ratio Test Against Stochastic Ordering in Several Populations," *Journal of the Americal Statistical Association*, 91, 1676–1683.

Wolak, F. A. (1991), "The Local Nature of Hypothesis Tests Involving Inequality Constraints in Nonlinear Models," *Econometrica*, 59, 981–995.