Ann. Inst. Statist. Math.
Vol. 55, No. 1, 121–136 (2003)

TESTING FOR INCREASING CONVEX ORDER IN SEVERAL POPULATIONS

XINSHENG LIU* AND JINDE WANG

Department of Mathematics, Nanjing University, Nanjing 210093, China

(Received April 2, 2001; revised March 20, 2002)

Abstract. Increasing convex order is one of important stochastic orderings. It is very often used in queueing theory, reliability, operations research and economics. This paper is devoted to studying the likelihood ratio test for increasing convex order in several populations against an unrestricted alternative. We derive the null asymptotic distribution of the likelihood ratio test statistic, which is precisely the chi-bar-squared distribution. The methodology for computing critical values for the test is also discussed. The test is applied to an example involving data for survival time for carcinoma of the oropharynx.

Key words and phrases: Increasing convex order, chi-bar-squared distribution, likelihood ratio test, asymptotic distribution.

1. Introduction

Stochastic ordering of distributions is an important concept in applied probability and the theory of statistical inference. It arises in many situations and has useful applications in practice. Many types of stochastic ordering have been defined in the literature, for example, in Stoyan (1983).

Statistical inference concerning stochastic ordering has been studied extensively. Because it is often easy to make value judgments when such orderings exist, and because incorporating these orderings into an inference increases its statistical efficiency, it is desirable to recognize the occurrence of such orderings and to model distributional structure under them. Brunk *et al.* (1966) obtained a closed form for nonparametric maximum likelihood estimates (MLE) of F and G under the assumption that $F \leq_{st} G$ (F is smaller than G in the (usual) stochastic order). Dykstra (1982) considered a similar problem with censored data and gave the MLE's in the form of Kaplan-Meier product-limit estimators. Testing procedures based on MLE's of two stochastically ordered distributions have been discussed by Franck (1984), Lee and Wolfe (1976), and Robertson and Wright (1981), among others. For more than two stochastically ordered distributions, Dykstra and Feltz (1989) and Feltz and Dykstra (1985) obtained MLE's by using an iterative algorithm. Y. Wang (1996) has characterized the asymptotic distribution of the likelihood ratio statistic. The inferences involving the uniform stochastic ordering and the likelihood ratio ordering have been discussed by Dykstra *et al.* (1991, 1995).

In this paper we derive the null asymptotic distribution of the likelihood ratio test

*Now at Department of Physics, Nanjing University, Nanjing 210093, China, e-mail: liuxs@nju.edu.cn

statistic for hypothesis testing problem involving increasing convex order constraints in the null hypothesis. We write the formal definition of the increasing convex order as follows (see Shaked and Shanthikumar (1994) for a general reference).

DEFINITION 1. The random variable (r.v.) X is smaller than the r.v. Y in the increasing convex order, written as $X \leq_{icx} Y$, or equivalently, their respective distribution functions (d.f.s) F and G satisfy $F \leq_{icx} G$, if

$$E(X-x)^+ = \int_x^\infty (t-x)dF(t) = \int_x^\infty (1-F(t))dt \leq \int_x^\infty (1-G(t))dt = E(Y-x)^+ (\text{all real} x) \tag{1.1}$$

provided these expectations (equivalently, integrals) are finite.

An important result about increasing convex order is that $X \leq_{icx} Y$ if and only if $Ef(X) \leq Ef(Y)$ holds for all non-decreasing convex real functions for which the expectations are defined, and if $X \leq_{icx} Y$ and $EX = EY$, then $Ef(X) \leq Ef(Y)$ holds for all convex functions f (see Stoyan (1983), p. 9).

There are interpretations of this ordering concept in applied research. For example, if non-negative random variables X, Y are the lives of two machines A, B respectively, $X \leq_{icx} Y$ means that the mean residual life of the machine A is smaller than that of the machine B (e.g., see Definition 1.3.1 of Stoyan (1983)). Ross (1983) has given another kind of interpretations for this ordering. If $X \leq_{icx} Y$ and $E(X) = E(Y)$, then from the result above we have that $\text{Var}(X) \leq \text{Var}(Y)$ (since $f(x) = x^2$ is convex). For this reason it is said intuitively that X is less variable than Y provided $X \leq_{icx} Y$.

Many applications of the increasing convex order have been found in queueing theory, reliability, operations research, economics and so on. For example, Stoyan (1983) has used this ordering to find the optimal sample size in experimental design. Ross (1983) has presented some applications of this ordering in comparison of queues and stochastic processes.

Because the increasing convex order has so many applications and theoretical implications, statistical inference concerning this ordering is certainly worthy of study. However, it is surprising that very little attention has been given to the problem of developing inference procedures for distributions ordered in this ordering. We do not find any tests in the literature specifically designed for the testing problem that we will discuss.

As in Robertson and Wright (1981), we assume the underlying populations are discrete in this article. However, because there is no closed-form expression for the MLE under increasing convex order constraints, we could not invoke the theory of isotonic regression to derive the asymptotic distribution of the likelihood ratio statistic. We overcome this difficulty by transforming increasing convex order constraints into a polyhedral cone constraint and then characterizing the likelihood ratio test statistic by an optimization problem. We will get desired asymptotic distribution by using the limit problem of the primal optimization problem.

This article is organized as follows. The definitions and preliminary results concerning the chi-bar-squared distribution, which we will need in our development, are summarized in Section 2. Section 3 derives the asymptotic distribution of the likelihood ratio test (LRT) statistic. Section 4 gives the methods to find critical values of the test. Section 5 presents an example to illustrate the developed theory.

2. The chi-bar-squared distribution

Let $Y \sim N(0, V)$ be an m-dimensional normal vector, C be a convex cone and define

$$\bar{\chi}^2 = Y'V^{-1}Y - \min_{\beta \in C}(Y - \beta)'V^{-1}(Y - \beta). \tag{2.1}$$

Suppose that $\hat{\beta}$ is the optimal solution to the optimization problem (2.1), that is, $\hat{\beta}$ is the projection of Y onto C in the V^{-1} metric. Then we have $\bar{\chi}^2 = \|\hat{\beta}\|^2$, where the norm and inner product are taken with respect to the matrix V^{-1}. The basic distributional result concerning $\bar{\chi}^2$ is that it is distributed as a mixture of chi-squared distribution, i.e.,

$$P(\bar{\chi}^2 \geq c) = \sum_{i=0}^{m} \omega_i P(\chi_i^2 \geq c), \tag{2.2}$$

where χ_i^2 is a chi-squared random variable with i degrees of freedom, $\chi_0^2 \equiv 0$ and ω_i's are nonnegative weights such that $\sum_{i=0}^{m} \omega_i = 1$. The distribution of $\bar{\chi}^2$ is determined by V and C, and we write $\bar{\chi}^2 \sim \bar{\chi}^2(V, C)$.

Particular cases of the distributional result (2.2) first appeared in Bartholomew (1959). His results were extended to a very general context by Kudô (1963) and independently by Nüesch (1966) for known covariance matrix, by Perlman (1969) for unknown covariance matrix. Shapiro (1988) has given a presentation of the general case.

Using properties of projections onto convex cones and their dual, we can write $\bar{\chi}^2$ as follows

$$\bar{\chi}^2 = \min_{\beta \in C^0}(Y - \beta)'V^{-1}(Y - \beta), \tag{2.3}$$

where C^0 is the dual cone of C.

The weights $\omega_i = \omega_i(m, V, C)$ which appeared in (2.2) depend on V and C, and the computation of the probability weights ω_i is a difficult numerical problem. In the case of $C = R_+^m = \{x : x \geq 0\}$ Kudô (1963) proposed a formula for the weights $\omega_i(m, V, R_+^m)$, denoted subsequently by $\omega_i(m, V)$. An expression in a closed form of $\omega_i(m, V)$ for $m \leq 4$ is available. For $m > 4$, reasonably accurate estimates of the weights can be easily obtained by Monte Carlo simulations.

3. Distribution theory for the test

3.1 *Hypothesis testing problem*

In this subsection we transform increasing convex order constraints into a polyhedral cone constraint and then we can formulate our hypothesis testing problem in a tractable form.

Let $X_i (i = 1, \ldots, m)$ be independent random variables, each taking value in the same set $\{b_1, \ldots, b_{k+1}\}$ (assume $b_1 < \cdots < b_{k+1}$) with $p_{ij} = P(X_i = b_j)$, $i = 1, \ldots, m, j = 1, \ldots, k+1$. Set $p_i = (p_{i1}, \ldots, p_{i,k+1})'$, $i = 1, \ldots, m$. Assume we have a random sample of size n_i from the population X_i, and n_{ij} observations sampled from X_i with the outcome $b_j (n_i = \sum_{j=1}^{k+1} n_{ij}, i = 1, \ldots, m)$. Then $\tilde{p}_i$, the vector of relative frequencies with $\tilde{p}_{ij} = n_{ij}/n_i$, is also the unconstrained maximum likelihood estimate of p_i, and $n_i\tilde{p}_i$ has a multinomial distribution with parameters n_i and p_i. Let $n = \sum_{i=1}^m n_i$. In deriving the asymptotic results, we assume that $\lim_{n\to\infty} n_i/n = r_i > 0$, $i = 1, \ldots, m$.

Consider the hypothesis $H_1 : p_1 \leq_{icx} \cdots \leq_{icx} p_m$. By Definition 1 we have

$$H_1 : \sum_{j=l+1}^{k+1} (b_j - b_l)p_{ij} \leq \sum_{j=l+1}^{k+1} (b_j - b_l)p_{i+1,j}, \qquad l = 1, \ldots, k, i = 1, \ldots, m-1. \tag{3.1}$$

Because

$$p_{i,k+1} = 1 - \sum_{j=1}^{k} p_{ij}, \qquad i = 1, \ldots, m, \tag{3.2}$$

the probability distribution of X_i is completely determined by the vector $\theta_i = (p_{i1}, \ldots, p_{ik})'$. Combining (3.1) and (3.2), we have

$$\begin{aligned} H_1 : (b_l - b_{k+1})\sum_{j=1}^{l} p_{ij} &+ \sum_{j=l+1}^{k} (b_j - b_{k+1})p_{ij} \\ \leq (b_l - b_{k+1})\sum_{j=1}^{l} p_{i+1,j} &+ \sum_{j=l+1}^{k} (b_j - b_{k+1})p_{i+1,j}, \\ & l = 1, \ldots, k, i = 1, \ldots, m-1. \end{aligned} \tag{3.3}$$

For $l = 1, \ldots, k$, letting $a_l = b_l - b_{k+1}$,

$$B_l = (\underbrace{a_l, \ldots, a_l}_{l}, a_{l+1}, \ldots, a_k)', \qquad C_l = (-B_l', B_l')', \qquad D_{il} = (\underbrace{0_k', \ldots, 0_k'}_{i-1}, C_l', 0_k', \ldots, 0_k')',$$

where 0_k is a $k \times 1$ zero vector, D_{il} is a $mk \times 1$ vector, $\theta = (\theta_1', \ldots, \theta_m')'$, (3.3) can be rewritten as

$$\begin{aligned} H_1 : \theta \in S &= \{\theta : D_{il}'\theta \geq 0, l = 1, \ldots, k, i = 1, \ldots, m-1\} \\ &= \{\theta : A_j'\theta \geq 0, j = 1, \ldots, (m-1)k\} = \{\theta : A\theta \geq 0\}, \end{aligned} \tag{3.4}$$

where A is a $(m-1)k$-by-mk matrix, and $\text{rank}(A) = (m-1)k$. S is a polyhedral cone. On the other hand, the potential constraints are that $p_i \in \{(x_1, \ldots, x_{k+1})' : x_j > 0, \sum_{j=1}^{k+1} x_j = 1\}$, $i = 1, \ldots, m$. Equivalently, we can write these constraints as $\theta \in E = \{\theta : p_{ij} > 0, \sum_{j=1}^{k} p_{ij} < 1, i = 1, \ldots, m\}$. In this paper, we consider likelihood ratio statistic for testing problem involving two hypotheses, namely, H_1 and H_2: no restriction among $p_i (i = 1, \ldots, m)$. With the notation above, our hypothesis testing problem can be written as

$$H_1 : \theta \in S \cap E \quad \text{versus} \quad H_2 - H_1 : \theta \in R^{mk} - S \cap E. \tag{3.5}$$

The hypothesis $H_1 : p_1 \leq_{icx} \cdots \leq_{icx} p_m$ is implied by the hypothesis $H_1 : p_1 \leq_{st} \cdots \leq_{st} p_m$, but not conversely. Hence the test discussed here has a less restriction than the one considered by Robertson and Wright (1981).

3.2 *The asymptotic distribution*

In this subsection we give the null asymptotic distribution of the LRT statistic for our testing problem. We begin by expressing the likelihood function of $(p_1, \ldots, p_m)$ as

$$L(p_1, \ldots, p_m) \propto \prod_{i=1}^{m} \prod_{j=1}^{k+1} p_{ij}^{n_{ij}}.$$

For the hypothesis testing problem (3.5), the likelihood ratio statistic is

$$\begin{aligned} \text{(3.6)} \qquad T_{12} &= 2\left\{\max \log L(p_1, \ldots, p_m) - \max_{p_1 \leq_{icx} \cdots \leq_{icx} p_m} \log L(p_1, \ldots, p_m)\right\} \\ &= 2\{\log L(\tilde{p}_1, \ldots, \tilde{p}_m) - \log L(\hat{p}_1, \ldots, \hat{p}_m)\} \\ &= 2\sum_{i=1}^{m}\sum_{j=1}^{k+1}[n_{ij}(\log \tilde{p}_{ij} - \log \hat{p}_{ij})] \\ &= 2\left\{\sum_{i=1}^{m}\sum_{j=1}^{k}[n_{ij}(\log \tilde{p}_{ij} - \log \hat{p}_{ij})] \right. \\ &\qquad \left. + \sum_{i=1}^{m}\left[n_{i,k+1}\left(\log\left(1 - \sum_{t=1}^{k}\tilde{p}_{it}\right) - \log\left(1 - \sum_{t=1}^{k}\hat{p}_{it}\right)\right)\right]\right\} \\ &= 2\{\log L(\tilde{\theta}) - \log L(\hat{\theta})\} \\ &= 2\left\{\max_{\theta \in R^{mk}} \log L(\theta) - \max_{\theta \in S \cap E} \log L(\theta)\right\}, \end{aligned}$$

where $\hat{p}_i (i = 1, \ldots, m)$ and $\hat{\theta}$ are the restricted MLE's of $p_i (i = 1, \ldots, m)$ and θ under H_1 respectively, $\tilde{\theta}$ is the unrestricted MLE of θ. Under H_1, for $i = 1, \ldots, m$ we denote the unknown true value of p_i by $p_i^{(0)} = (p_{i1}^{(0)}, \ldots, p_{i,k+1}^{(0)})'$. Thus the unknown true value of θ is $\theta_0 = ((\theta_1^{(0)})', \ldots, (\theta_m^{(0)})')'$, where $\theta_i^{(0)} = (p_{i1}^{(0)}, \ldots, p_{ik}^{(0)})'$. To derive the asymptotic distribution of T_{12}, we first give the following lemma.

LEMMA 1. *$n^{1/2}(\hat{\theta} - \theta_0)$ is bounded in probability, that is, $\hat{\theta} - \theta_0 = O_p(n^{-1/2})$. Here $O_p(\cdot)$ is used in the sense of Mann and Wald (1943) (that is, for a sequence of k-dimensional chance variables $\{a_n\}$ and a sequence of positive numbers $\{b_n\}$, we write $a_n = O_p(b_n)$ if for each $\epsilon > 0$, there is an M_ϵ such that $P(|a_n| < M_\epsilon b_n) > 1 - \epsilon$).*

PROOF. We first show that $\hat{\theta}$ is a consistent estimate of θ_0. Note that $\theta_0 \in S$. Let

$$g_n(\theta) = \frac{1}{n}\log L(\theta) = \sum_{i=1}^{m}\left\{\frac{n_i}{n}\left[\sum_{j=1}^{k}\tilde{p}_{ij}\log p_{ij} + \tilde{p}_{i,k+1}\log\left(1 - \sum_{t=1}^{k}p_{it}\right)\right]\right\}.$$

Then

$$g(\theta) = E\left[\frac{1}{n}\log L(\theta)\right] = \sum_{i=1}^{m}\left\{\frac{n_i}{n}\left[\sum_{j=1}^{k}p_{ij}^{(0)}\log p_{ij} + p_{i,k+1}^{(0)}\log\left(1 - \sum_{t=1}^{k}p_{it}\right)\right]\right\}.$$

The gradient vector and the Hessian matrix of $g(\theta)$ are given respectively by

$$\nabla g(\theta) = \left(\frac{\partial g(\theta)}{\partial p_{11}}, \ldots, \frac{\partial g(\theta)}{\partial p_{1k}}, \ldots, \frac{\partial g(\theta)}{\partial p_{m1}}, \ldots, \frac{\partial g(\theta)}{\partial p_{mk}}\right)',$$
$$H(\theta) = \left(\frac{\partial^2 g(\theta)}{\partial \theta_i \partial \theta_j}\right)_{(mk)\times(mk)} = \begin{pmatrix} H_1 & & 0 \\ & \ddots & \\ 0 & & H_m \end{pmatrix}, \tag{3.7}$$

where $\partial g(\theta)/\partial p_{ij} = (n_i/n)[p_{ij}^{(0)} p_{ij}^{-1} - p_{i,k+1}^{(0)}(1 - \sum_{t=1}^k p_{it})^{-1}]$; $H_i = (h_{jl}^i)_{k\times k}$ with $h_{jj}^i = -(n_i/n)[p_{ij}^{(0)} p_{ij}^{-2} + p_{i,k+1}^{(0)}(1 - \sum_{t=1}^k p_{it})^{-2}]$ and $h_{jl}^i = -(n_i/n) p_{i,k+1}^{(0)}(1 - \sum_{t=1}^k p_{it})^{-2} (j \neq l), j, l = 1, \ldots, k, i = 1, \ldots, m$. It is easily seen that $H(\theta) < 0$ (negative definite) and $\nabla g(\theta_0) = 0$. Hence $g(\theta)$ is a strictly concave function and θ_0 maximizes $g(\theta)$ over S.

By the central limit theorem, we obtain

$$n_i^{1/2}(\tilde{p}_i - p_i^{(0)}) \xrightarrow{L} U_i, \quad i = 1, \ldots, m, \tag{3.8}$$

where $U_i = (U_{i1}, \ldots, U_{i,k+1})'$, $i = 1, \ldots, m$, are independent, and U_i follows a multivariate normal distribution with the covariance matrix satisfying

$$\operatorname{var}(U_{ij}) = p_{ij}^{(0)}(1 - p_{ij}^{(0)}), \quad \operatorname{cov}(U_{ij}, U_{il}) = -p_{ij}^{(0)} p_{il}^{(0)} (j \neq l), \quad i = 1, \ldots, m, \tag{3.9}$$

and "$\rightarrow^L$" stands for convergence in distribution. Thus, appealing to Theorem 5.1 of Billingsley (1968), we have that

$$g_n(\theta) = g(\theta) + n^{-1/2} \sum_{i=1}^m \left\{ r_i^{1/2} \left[\sum_{j=1}^k U_{ij} \log p_{ij} + U_{i,k+1} \log\left(1 - \sum_{t=1}^k p_{it}\right)\right]\right\} \cdot \{1 + o_p(1)\}. \tag{3.10}$$

$g_n(\theta)$ and $g(\theta)$ achieve their maxima on S at $\hat{\theta}$ and θ_0, so (3.10) implies that $\hat{\theta}$ is a consistent estimate of θ_0.

Next by Taylor's theorem we have

$$g_n(\theta) - g_n(\theta_0) = \nabla g_n(\theta_0)'(\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)' H_n(\theta_0)(\theta - \theta_0) + \|\theta - \theta_0\|^3 O_p(1). \tag{3.11}$$

$\nabla g_n(\theta_0)$ and $H_n(\theta_0)$ can be obtained by replacing $p_{ij}^{(0)}$ with $\tilde{p}_{ij}$ and p_{ij} with $p_{ij}^{(0)}$ in (3.7). Again using (3.8) and the weak convergence result mentioned earlier we get that

$$[n^{1/2} \nabla g_n(\theta_0)]'(\theta - \theta_0) \xrightarrow{L} f'(\theta - \theta_0), \tag{3.12}$$

where $f = (f_{11}, \ldots, f_{1k}, \ldots, f_{m1}, \ldots, f_{mk})'$, $f_{ij} = r_i^{1/2}((p_{ij}^{(0)})^{-1} U_{ij} - (p_{i,k+1}^{(0)})^{-1} U_{i,k+1})$ $(i = 1, \ldots, m, j = 1, \ldots, k)$. Since $\tilde{p}_i \to p_i^{(0)}$ in probability, $H_n(\theta_0) \to -V$ in probability, where V is defined in Theorem 1. Hence

$$\frac{1}{2}(\theta - \theta_0)' H_n(\theta_0)(\theta - \theta_0) = -\frac{1}{2}(\theta - \theta_0)' V(\theta - \theta_0) + o_p(1)\|\theta - \theta_0\|^2. \tag{3.13}$$

Refer to the equation (3.11) with θ replaced by $\hat{\theta}$. Note that $\hat{\theta}$ is a consistent estimate of θ_0. Then for any $\epsilon > 0$, there is a constant $C_\epsilon(> 0)$, and a sequence $c_n(\epsilon) \to 0$ such that with probability greater than $1 - \epsilon$

$$(3.14)\quad 0 \leq g_n(\hat{\theta}) - g_n(\theta_0) \leq -\frac{1}{2}(\hat{\theta} - \theta_0)'V(\hat{\theta} - \theta_0) + n^{-1/2}C_\epsilon\|\hat{\theta} - \theta_0\| + c_n(\epsilon)\|\hat{\theta} - \theta_0\|^2.$$

Multiplying (3.14) by $-n$ we obtain $0 \geq \frac{1}{2}\hat{\gamma}'V\hat{\gamma} - C_\epsilon\|\hat{\gamma}\| - c_n(\epsilon)\|\hat{\gamma}\|^2$, where $\hat{\gamma} = n^{1/2}(\hat{\theta} - \theta_0)$. Then one could find a constant M_ϵ such that $\|\hat{\gamma}\| \leq M_\epsilon$ by the positive definiteness of V. The lemma follows. □

It can be easily seen that the equality (3.6) is equivalent to

$$(3.15)\qquad \begin{aligned} T_{12} &= \min_{p_1 \leq_{icx} \cdots \leq_{icx} p_m} \{2[\log L(\tilde{p}_1, \ldots, \tilde{p}_m) - \log L(p_1, \ldots, p_m)]\} \\ &= \min_{\theta \in S \cap E} \{2[\log L(\tilde{\theta}) - \log L(\theta)]\}. \end{aligned}$$

Let $F_n(\theta) = 2[\log L(\tilde{\theta}) - \log L(\theta)]$(then $F_n(p_1, \ldots, p_m) = 2[\log L(\tilde{p}_1, \ldots, \tilde{p}_m) - \log L(p_1, \ldots, p_m)]$). From Lemma 1 we can use $\beta = n^{1/2}(\theta - \theta_0)$ as the optimization variable. The variable β is often used in the statistical literature, for example, in Prakasa Rao (1987), and J. Wang (1996). Because $\beta = n^{1/2}(\theta - \theta_0)$ could be any real number, the constraint set E will be slackened in this case. Substituting β into the problem (3.15), we have

$$(3.16)\qquad T_{12} = \min_{\beta \in S_n} G_n(\beta) = G_n(\hat{\beta}_n),$$

where $G_n(\beta) = F_n(n^{-1/2}\beta + \theta_0)$, $S_n = \{\beta : A_l'(n^{-1/2}\beta + \theta_0) \geq 0, l = 1, \ldots, (m-1)k\}$ and $\hat{\beta}_n$ is the optimal solution. To find the limit form of problem (3.16), we first give the limit form of the objective function $G_n(\beta)$.

THEOREM 1.

$$(3.17)\qquad G_n(\beta) \xrightarrow{L} G(\beta) = (Z - \beta)'V(Z - \beta),$$

where $Z \sim N(0, V^{-1})$ *and the block diagonal matrix* $V = \mathrm{diag}(V_1, \ldots, V_m)$, $V_i = r_i M_i, M_i = (m_{jl}^{(i)})_{k\times k}$, $m_{jj}^{(i)} = (p_{ij}^{(0)})^{-1} + (p_{i,k+1}^{(0)})^{-1} (j = 1, \ldots, k)$, $m_{jl}^{(i)} = (p_{i,k+1}^{(0)})^{-1} (j \neq l), i = 1, \ldots, m$.

PROOF.

$$\begin{aligned} F_n(p_1, \ldots, p_m) &= 2[\log L(\tilde{p}_1, \ldots, \tilde{p}_m) - \log L(p_1, \ldots, p_m)] \\ &= 2\sum_{i=1}^{m}\sum_{j=1}^{k+1} \{n_i[\tilde{p}_{ij}(\log \tilde{p}_{ij} - \log p_{ij})]\}. \end{aligned}$$

Writing a second order Taylor's expansion for $\log p_{ij}$ about $\tilde{p}_{ij}$, we obtain

$$\begin{aligned} F_n(p_1, \ldots, p_m) &= \sum_{i=1}^{m}\sum_{j=1}^{k+1}[n_i \tilde{p}_{ij}\alpha_{ij}^{-2}(p_{ij} - \tilde{p}_{ij})^2] \\ &= \sum_{i=1}^{m}\left\{\frac{n_i}{n}\sum_{j=1}^{k+1}[\tilde{p}_{ij}\alpha_{ij}^{-2}(n^{1/2}(p_{ij} - p_{ij}^{(0)}) - n^{1/2}(\tilde{p}_{ij} - p_{ij}^{(0)}))^2]\right\}, \end{aligned}$$

where α_{ij} is between p_{ij} and $\tilde{p}_{ij}$. Since $\tilde{p}_i$ is a consistent estimate of $p_i^{(0)}$ and θ is in a $n^{-1/2}$-shrinking neighborhood of θ_0, with probability 1 $\alpha_i = (\alpha_{i1}, \ldots, \alpha_{i,k+1})' \to p_i^{(0)}$ and $\tilde{p}_i \to p_i^{(0)}, i = 1, \ldots, m$. From (3.8) it follows that $n_i^{1/2}(\tilde{\theta}_i - \theta_i^{(0)}) \xrightarrow{L} U_i, i = 1, \ldots, m$, where $U_i = (U_{i1}, \ldots, U_{ik})'$, $i = 1, \ldots, m$, are independent, U_i follows a multivariate normal distribution with mean zero and covariance matrix satisfying the equality (3.9), and $\tilde{\theta}_i = (\tilde{p}_{i1}, \ldots, \tilde{p}_{ik})'$, $\theta_i^{(0)} = (p_{i1}^{(0)}, \ldots, p_{ik}^{(0)})'$ for $i = 1, \ldots, m$. Hence, appealing to Theorem 4.4 of Billingsley (1968), we have that

$$(3.18) \qquad \begin{aligned} &(n_1^{1/2}(\tilde{\theta}_1 - \theta_1^{(0)})', \ldots, n_m^{1/2}(\tilde{\theta}_m - \theta_m^{(0)})', \tilde{p}_1', \ldots, \tilde{p}_m', \alpha_1', \ldots, \alpha_m')' \\ &\xrightarrow{L} (U_1', \ldots, U_m', (p_1^{(0)})', \ldots, (p_m^{(0)})', (p_1^{(0)})', \ldots, (p_m^{(0)})')' \end{aligned}$$

as $n_i (i = 1, \ldots, m)$ simultaneously approach ∞. Since $F_n(p_1, \ldots, p_m)$ are continuous function of the vector in the left side of (3.18), we may apply the weak convergence result mentioned in the proof of Lemma 1. Substituting $p_{i,k+1}^{(0)} = 1 - \sum_{t=1}^k p_{it}^{(0)}, p_{i,k+1} = 1 - \sum_{t=1}^k p_{it}, \tilde{p}_{i,k+1} = 1 - \sum_{t=1}^k \tilde{p}_{it}$, and $\beta = n^{1/2}(\theta - \theta_0)$ into $F_n(p_1, \ldots, p_m)$ we obtain

$$\begin{aligned} F_n(\theta) \xrightarrow{L} \sum_{i=1}^m \Bigg\{ r_i \Bigg[&\sum_{j=1}^k (p_{ij}^{(0)})^{-1} (r_i^{-1/2} U_{ij} - \beta_{ij})^2 \\ &+ \left(1 - \sum_{t=1}^k p_{it}^{(0)}\right)^{-1} \left(\sum_{t=1}^k (r_i^{-1/2} U_{it} - \beta_{it})\right)^2 \Bigg] \Bigg\} \\ = \sum_{i=1}^m (Z_i - \beta_i)' V_i (Z_i - \beta_i), \end{aligned}$$

where $Z_i = (r_i^{-1/2} U_{i1}, \ldots, r_i^{-1/2} U_{ik})', \beta_i = (\beta_{i1}, \ldots, \beta_{ik})'$ with $\beta_{ij} = n^{1/2}(p_{ij} - p_{ij}^{(0)})$ for $j = 1, \ldots, k$, V_i is defined in Theorem 1, $i = 1, \ldots, m$. It is easily seen that $Z_i \sim N(0, V_i^{-1})$, and $V_i^{-1} = r_i^{-1} M_i^{-1}$, $M_i^{-1} = (w_{jl}^{(i)})_{k \times k}$ with $w_{jj}^{(i)} = p_{ij}^{(0)}(1 - p_{ij}^{(0)})(j = 1, \ldots, k)$ and $w_{jl}^{(i)} = -p_{ij}^{(0)} p_{il}^{(0)} (j \neq l)$.

Let $Z = (Z_1', \ldots, Z_m')'$, $V = \text{diag}(V_1, \ldots, V_m)$, $\beta = (\beta_1', \ldots, \beta_m')'$. Since $Z_i, i = 1, \ldots, m$, are independent, $Z \sim N(0, V^{-1})$. Thus (3.17) follows. The proof is complete. □

Next we study the limit of the feasible solution set S_n. Let S_0 be the interior of S. Suppose that S^i is the set that exactly $(m-1)k - i$ of the $(m-1)k$ inequalities of S are strict (without loss of generality suppose that the last $(m-1)k - i$ inequalities of S are strict), that is,

$$\begin{aligned} &S_0 = \{\theta : A_j'\theta > 0, j = 1, \ldots, (m-1)k\} \\ &S^i = \{\theta : A_n'\theta = 0, n = 1, \ldots, i, A_l'\theta > 0, l = i+1, \ldots, (m-1)k\} (0 < i \leq (m-1)k). \end{aligned}$$

Then we have the following result.

THEOREM 2. *Suppose that θ_0 is the unknown true value of θ. Then as $n \to \infty$ we have*

(1) *If $\theta_0 \in S_0$, then $S_n \uparrow R^{mk}$.*

(2) *If* $\theta_0 \in S^i$, *then* $S_n \uparrow T^i = \{\beta : A_j'\beta \geq 0, j = 1, \ldots, i\}(0 < i < (m-1)k)$.

(3) *If* $\theta_0 \in S^{(m-1)k}$, *then* $S_n \uparrow S$.

Here "$A_n \uparrow B$*" means that* $A_n \subseteq A_{n+1}$ *for any* n *and* $\lim_{n\to\infty} A_n = B$.

PROOF. We have that

$$\begin{aligned} S_n &= \{\beta : A_j'(n^{-1/2}\beta + \theta_0) \geq 0, j = 1, \ldots, (m-1)k\} \\ &= \{\beta : A_j'\beta \geq n^{1/2}(-A_j'\theta_0), j = 1, \ldots, (m-1)k\}. \end{aligned}$$

Then $S_1 \subset S_2 \subset S_3 \subset \ldots$ since $A_j'\theta_0 \geq 0$ for $j = 1, \ldots, (m-1)k$. Furthermore, if $A_j'\theta_0 = 0$ for some j, $A_j'\beta \geq 0$. If $A_j'\theta_0 > 0$, $A_j'\beta \geq n^{1/2}(-A_j'\theta_0) \to -\infty(n \to \infty)$, that is, this inequality is slack. Thus the result of the theorem follows immediately. □

Remark 1. Because $\{S_n\}$ is monotone and closed, the convergence in Theorem 2 is also the convergence of sets in Kuratowski's sense. We write $S = (K)\lim S_n$, if for any $z \in S$ there is a sequence $\{z_n\}$ such that $z_n \in S_n$ and $z_n \to z$ and for any sequence $\{z_n\}$ with $z_n \in S_n$ any accumulation point of $\{z_n\}$ must belong to S (e.g., see Attouch (1985)). This kind of convergence of sets will lead to convergence of optimal solution of the related optimization problems, as shown later.

With Theorems 1 and 2 we can formulate a limit problem of (3.16):

$$T = G(\hat{\beta}) = \min_{\beta \in K} (Z - \beta)'V(Z - \beta), \tag{3.19}$$

where K is one of the $T^i(0 < i \leq (m-1)k)$, R^{mk} and S, and $\hat{\beta}$ is the optimal solution. Although for the objective function we have $G_n(\beta) \to^L G(\beta)$ for any fixed β, it has not been shown that $G_n(\hat{\beta}_n) \to^L G(\hat{\beta})$. When β is varying over some connected set D, $\{G_n(\beta), \beta \in D\}$ and $\{G(\beta), \beta \in D\}$ can be viewed as stochastic processes. We will study the convergence in distribution of the sequence of these stochastic processes in the following lemma.

LEMMA 2. *The stochastic processes* $\{G_n(\beta), \beta \in D\}$ *converge in distribution to* $\{G(\beta), \beta \in D\}$, *that is,*

$$\{G_n(\beta), \beta \in D\} \xrightarrow{L} \{G(\beta), \beta \in D\}, \tag{3.20}$$

where $D = \{\beta : \|\beta\| \leq M, \beta \in R^{mk}\}$.

PROOF. According to the theory of probability (see Prakasa Rao (1975)), $\{G_n(\beta), \beta \in D\}$ converges in distribution to $\{G(\beta), \beta \in D\}$ if and only if the following two conditions are satisfied:

(a) Any finite dimensional distribution of process $\{G_n(\beta), \beta \in D\}$ converges weakly to the corresponding finite-dimensional distribution of $\{G(\beta), \beta \in D\}$;

(b) For any $\epsilon > 0$ it holds that

$$\lim_{n\to\infty} \sup_{h\to 0} P\left\{ \sup_{\|\beta^{(1)} - \beta^{(2)}\| \leq h} |G_n(\beta^{(1)}) - G_n(\beta^{(2)})| > \epsilon, \beta^{(1)}, \beta^{(2)} \in D \right\} = 0.$$

First we check condition (a). By Cramér-Wold theorem, it suffices to show that for any $c_1, \ldots, c_r \in R$ and any $\beta^{(1)}, \ldots, \beta^{(r)} \in D$, we have

$$\sum_{j=1}^{r} c_j G_n(\beta^{(j)}) \xrightarrow{L} \sum_{j=1}^{r} c_j G(\beta^{(j)}).$$

This convergence result can be proved in the same way as Theorem 1. We will not repeat the procedure here.

Next verify the condition (b). For $\beta^{(1)}, \beta^{(2)}$ in D, we have

$$\begin{aligned} &|G_n(\beta^{(1)}) - G_n(\beta^{(2)})| \\ &\quad = 2|\log L(n^{-1/2}\beta^{(2)} + \theta_0) - \log L(n^{-1/2}\beta^{(1)} + \theta_0)| \\ &\quad = 2n|[g_n(n^{-1/2}\beta^{(2)} + \theta_0) - g_n(\theta_0)] - [g_n(n^{-1/2}\beta^{(1)} + \theta_0) - g_n(\theta_0)]|, \end{aligned}$$

where $g_n(\cdot)$ is defined in the proof of Lemma 1. From (3.11) and (3.13) it follows that

$$\begin{aligned} g_n(n^{-1/2}\beta^{(i)} + \theta_0) - g_n(\theta_0) = n^{-1} \Big[& -\frac{1}{2}(\beta^{(i)})' V \beta^{(i)} + o_p(1)\|\beta^{(i)}\|^2 \\ & + (n^{1/2} \nabla g_n(\theta_0))'\beta^{(i)} + n^{-1/2}\|\beta^{(i)}\|^3 O_p(1) \Big]. \end{aligned}$$

Therefore, when $n \to \infty$ and $\|\beta^{(1)} - \beta^{(2)}\| \to 0$, for any given ϵ we have $|(\beta^{(1)})'V\beta^{(1)} - (\beta^{(2)})'V\beta^{(2)}| < \frac{\epsilon}{4}$, and with probability approaching one

$$\begin{aligned} |G_n(\beta^{(1)}) - G_n(\beta^{(2)})| \leq\ & |(\beta^{(1)})'V\beta^{(1)} - (\beta^{(2)})'V\beta^{(2)}| \\ & + 2|(n^{1/2} \nabla g_n(\theta_0))'(\beta^{(1)} - \beta^{(2)})| + \frac{\epsilon}{4}. \end{aligned}$$

Using (3.12) and the fact that D is compact we conclude that

$$\begin{aligned} &\lim_{n\to\infty} \sup_{h\to 0} P\left\{ \sup_{\|\beta^{(1)}-\beta^{(2)}\| \leq h} |G_n(\beta^{(1)}) - G_n(\beta^{(2)})| > \epsilon, \beta^{(1)}, \beta^{(2)} \in D \right\} \\ &\quad \leq \lim_{n\to\infty} \sup_{h\to 0} P\left\{ \sup_{\|\beta^{(1)}-\beta^{(2)}\| \leq h} |(n^{1/2} \nabla g_n(\theta_0))'(\beta^{(1)} - \beta^{(2)})| \geq \frac{1}{4}\epsilon, \beta^{(1)}, \beta^{(2)} \in D \right\} \\ &\quad = 0. \end{aligned}$$

Then both conditions (a) and (b) are satisfied. The assertion of this lemma follows. □

From the convergence result on the sequence of stochastic processes in Lemma 2, the desired convergence will be obtained.

THEOREM 3. *T_{12} converges in distribution to T, that is,*

$$T_{12} = G_n(\hat{\beta}_n) \xrightarrow{L} G(\hat{\beta}) = T. \tag{3.21}$$

PROOF. Note that the sample functions of the stochastic processes $\{G_n(\beta), \beta \in D\}$ and $\{G(\beta), \beta \in D\}$ are continuous functions on D. Let $C(D)$ be the space of all continuous functions over D whose metric is defined by

$$d(h_1, h_2) = \sup_{\beta \in D} |h_1(\beta) - h_2(\beta)|, \qquad h_1, h_2 \in C(D).$$

Then the stochastic processes $\{G_n(\beta), \beta \in D\}$ and $\{G(\beta), \beta \in D\}$ induce a family of probability measures $\{\mu, \mu_n, n = 1, \ldots\}$. By Lemma 2, the convergence in (3.20) implies $\{\mu_n\}$ converges weakly to μ, written as $\mu_n \Rightarrow \mu$.

Define mappings $H_n(\cdot)$ and $H(\cdot)$ on $C(D)$ such that

$$H_n(f_n) = \min_{\beta \in S_n \cap D} f_n(\beta) = f_n(\hat{\beta}_n^{(D)}) \tag{3.22}$$

and

$$H(f) = \min_{\beta \in K \cap D} f(\beta) = f(\hat{\beta}^{(D)}) \tag{3.23}$$

for $f_n, f \in C(D)$, where K is as in (3.19), and $\hat{\beta}_n^{(D)}$ and $\hat{\beta}^{(D)}$ are optimal solutions. First we are going to show that

$$H_n(G_n) \xrightarrow{L} H(G). \tag{3.24}$$

Since $G(\beta)$ is a strictly convex function and $K \cap D$ is a convex set, the problem $H(G) = \min_{\beta \in K \cap D} G(\beta)$ has a unique optimal solution. To show (3.24), by $\mu_n \Rightarrow \mu$ and an extension of the continuous mapping theorem (cf. Theorem 5.5 of Billingsley (1968)) it suffices to show that

$$\lim_{n \to \infty} H_n(f_n) = H(f) \tag{3.25}$$

for any f_n, f in $C(D)$ with $f_n \to f$ and f is such that (3.23) has a unique optimal solution. Observe the convergence of f_n to f means that $\max |f_n(\beta) - f(\beta)| \to 0$ and this implies $f_n(\beta_n) \to f(\beta)$ for any $\beta_n \to \beta$. Thus, to show (3.25), it suffices to show that $\hat{\beta}_n^{(D)} \to \hat{\beta}^{(D)}$. We first show the following: if $\hat{\beta}_n^{(D)}$, $n = 1, 2, \ldots$ are optimal solutions of problem (3.22) and $\bar{\beta}$ is an accumulation point of $\{\hat{\beta}_n^{(D)}\}$, then $\bar{\beta}$ must be an optimal solution of problem (3.23). Suppose it is not true. Then there is a point β_0 in $K \cap D$ such that $f(\beta_0) < f(\bar{\beta})$. Without loss of generality we assume β_0 is a interior point of D (since f is continuous). On the other hand, by Remark 1 there is a sequence β_n such that $\beta_n \in S_n$ and $\beta_n \to \beta_0$, and then $\beta_n \in D$ when n is large enough. As $\bar{\beta}$ is an accumulation point of $\{\hat{\beta}_n^{(D)}\}$, there must be a subsequence $\{\hat{\beta}_{n_j}^{(D)}\}$ such that $\hat{\beta}_{n_j}^{(D)} \to \bar{\beta}$. Noticing that $f_n \to f$, we obtain

$$f(\beta_0) = \lim f_n(\beta_n) \geq \lim f_n(\hat{\beta}_{n_j}^{(D)}) = f(\bar{\beta}).$$

This contradicts the working assumption $f(\beta_0) < f(\bar{\beta})$. Hence $\bar{\beta}$ must be an optimal solution of problem (3.23).

Since D is compact and S_n, K are closed, $\{\hat{\beta}_n^{(D)}\}$ must have accumulation points. Moreover, by the assumption on f the only possible accumulation point is $\hat{\beta}^{(D)}$. Thus we get (3.25) and then (3.24).

Since $\hat{\beta}$ is the optimal solution of the problem (3.19) and $\beta = 0 \in K$, we have

$$0 \geq (Z - \hat{\beta})'V(Z - \hat{\beta}) - Z'VZ = \hat{\beta}'V\hat{\beta} - 2\hat{\beta}'VZ.$$

Observe that $Z \sim N(0, V^{-1})$ and $V > 0$ (positive definite). Then for any $\epsilon > 0$ there exists a constant M_ϵ such that $\|\hat{\beta}\| \leq M_\epsilon$ with a probability larger than $1 - \epsilon$. Without loss generality, we assume that this M_ϵ is the same as that M_ϵ in the proof of Lemma 1 (otherwise we may choose the larger one as the common M_ϵ). Note that $G_n(\hat{\beta}_n) = H_n(G_n), G(\hat{\beta}) = H(G)$ when $\|\hat{\beta}_n\| \leq M_\epsilon$ and $\|\hat{\beta}\| \leq M_\epsilon$. Therefore

$$P(G_n(\hat{\beta}_n) \neq H_n(G_n)) < \epsilon, \qquad P(G(\hat{\beta}) \neq H(G)) < \epsilon.$$

By the arbitrariness of ϵ and $H_n(G_n) \rightarrow^L H(G)$, we get (3.21). This is the desired result. □

Now we come to the distribution result of the likelihood ratio test statistic.

THEOREM 4. *Let S_0, S^i, T^i and θ_0 be as in Theorem 2. Let $(T^i)^0$ and S^0 be dual cones of T^i and S. Then we have:*
(1) *If $\theta_0 \in S_0$, then $\lim_{n\to\infty} P(T_{12} = 0) = 1$.*
(2) *If $\theta_0 \in S^i$, then $T_{12} \rightarrow^L \bar{\chi}^2(V^{-1}, (T^i)^0)(0 < i < (m-1)k)$.*
(3) *If $\theta_0 \in S^{(m-1)k}$, then $T_{12} \rightarrow^L \bar{\chi}^2(V^{-1}, S^0)$.*

PROOF. The result of the theorem follows immediately from Theorems 2,3 (replacing K in (3.19) by R^{mk}, T^i and S respectively) and the equality (2.3). □

4. Computing the critical value for the test

It is well known that H_1 that θ is in the polyhedral cone S is a composite hypothesis. In view of Theorem 4 we know that the asymptotic distribution of T_{12} depends on the location of θ_0 in S. To compute the critical value of the test, we need the least favorable null distribution, which is given by the following theorem. Robertson and Wegman (1978) has established a similar result for testing hypotheses that a collection of parameters satisfy some order restriction.

THEOREM 5. *Let $P_{\theta_0 \in D}(E)$ be the probability of the event E computed under the assumption that $\theta_0 \in D$. Then*

$$\lim_{n\to\infty} P_{\theta_0 \in S}(T_{12} \geq t) \leq \lim_{n\to\infty} P_{\theta_0 \in S^{(m-1)k}}(T_{12} \geq t)$$

holds for any real t, where $S^{(m-1)k} = \{\theta : A_j'\theta = 0, j = 1, \ldots, (m-1)k\}$.

PROOF. First, Theorem 4(1) states that all elements in S_0 can be removed from consideration as a least favorable value since T_{12} converges in probability to zero for these values of θ_0.

Next, for $0 < i < (m-1)k$, let $\hat{\beta}_{T^i}$ and $\hat{\beta}_S$ be optimal solutions in (3.19) when $K = T^i$ and $K = S$ respectively. By the comment given in Section 2 $\hat{\beta}_{T^i}$ and $\hat{\beta}_S$ are the projections of Z onto T^i and S in the V metric, and $G(\hat{\beta}_{T^i})$ and $G(\hat{\beta}_S)$ are the distances from Z to T^i and S. Because $S \subset T^i$, $G(\hat{\beta}_{T^i}) \leq G(\hat{\beta}_S)$. Thus for any real t we have

$P(G(\hat{\beta}_{T^i}) \geq t) \leq P(G(\hat{\beta}_S) \geq t)$. The proof is then completed by noticing the result of Theorem 3. □

From Theorem 5 it follows that $S^{(m-1)k}$ is the least favorable null hypothesis among hypotheses satisfying H_1 in the sense of yielding the largest type I error probability. Then by Theorem 4(3) we get that the least favorable null distribution of the test is $\bar{\chi}^2(V^{-1}, S^0)$. Therefore, for a chosen level α, the critical value c_α could be chosen to satisfy

$$P(\bar{\chi}^2(V^{-1}, S^0) \geq c_\alpha) = \alpha. \tag{4.1}$$

We now supply some methods to computing $\hat{\theta}$, the restricted MLE of θ under H_1. As stated in Subsection 3.1 we can transform increasing convex order constraints into a polyhedral cone constraint. Substituting $p_{i,k+1}$ with $1 - \sum_{j=1}^{k} p_{ij}$ for $i = 1, \ldots, m$, we express the log-likelihood function of $(p_1, \ldots, p_m)$ as

$$\begin{aligned} \log L(\theta) &= \sum_{i=1}^{m} \left[\sum_{j=1}^{k} (n_{ij} \log p_{ij}) + n_{i,k+1} \log \left(1 - \sum_{t=1}^{k} p_{it} \right) \right] + \text{const} \\ &= -N(\theta) + \text{const}. \end{aligned}$$

Then it can be easily seen that $\hat{\theta}$ is the optimal solution of the optimization problem

$$\begin{cases} \min \ N(\theta) \\ s.t. \ \ \theta \in S \cap E. \end{cases} \tag{4.2}$$

Since for $i = 1, \ldots, m$ and $j = 1, \ldots, k$, $\partial N(\theta)/\partial p_{ij} = -n_{ij} p_{ij}^{-1} + n_{i,k+1}(1 - \sum_{t=1}^{k} p_{it})^{-1}$, $\partial^2 N(\theta)/\partial p_{ij} \partial p_{il} = n_{i,k+1}(1 - \sum_{t=1}^{k} p_{it})^{-2} (j \neq l)$ and $\partial^2 N(\theta)/\partial p_{ij}^2 = n_{ij} p_{ij}^{-2} + n_{i,k+1}(1 - \sum_{t=1}^{k} p_{it})^{-2}$, it is easily verified that the Hessian matrix of $N(\theta)$ is positive definite. Thus $N(\theta)$ is strictly convex and has a unique minimum belonging to the convex set $S \cap E$.

Several algorithms in mathematical programming can be directly used to compute $\hat{\theta}$. For example, we can get the value of $\hat{\theta}$ by making use of penalty function methods or feasible direction methods (e.g., see Bazaraa and Shetty (1979)). However in practice the specific optimal solutions could be computed by a system of Matlab functions. We have computed the restricted MLE in the example of Section 5 by the function "fmincon" (the trust region method) and the function "constr" (the penalty function method) in which the starting points must be chosen to be feasible points. A computer program that implements all the computations in the example is available from the authors on request.

There are still two problems for applying the testing procedure above: how to compute the weights ω_i and how to handle the unknown parameter θ_0. From the formulas (3.4) and (5.5) of Shapiro (1988) we have

$$\begin{aligned} \omega_i(mk, V^{-1}, S^0) &= \omega_{mk-i}(mk, V^{-1}, S) \\ &= \omega_{(m-1)k-i}((m-1)k, AV^{-1}A'), \qquad i = 0, 1, \ldots, (m-1)k, \end{aligned} \tag{4.3}$$

while the remaining weights vanish, where A is defined in (3.4), and $\omega_{(m-1)k-i}((m-1)k, AV^{-1}A')$, $i = 0, 1, \ldots, (m-1)k$, can be obtained according to the method described in Section 2. However, the distribution of $\bar{\chi}^2$ depends on the unknown parameter $\theta_0(\in S^{(m-1)k})$ through the weights. For this problem one may use $\bar{\theta} = (n^{-1}\sum_{i=1}^m n_{i1}, \ldots, n^{-1}\sum_{i=1}^m n_{ik}, \ldots, n^{-1}\sum_{i=1}^m n_{i1}, \ldots, n^{-1}\sum_{i=1}^m n_{ik})'$ as an estimate of the unknown θ_0 and compute the weights based on this estimate, where $\bar{\theta}$ can be obtained by setting $p_{1j} = \cdots = p_{mj} (j = 1, \ldots, k+1)$ and solving the resulting optimization problem. The estimate $\bar{\theta}$ is the MLE of θ_0 under the least favorable null condition $\theta_0 \in S^{(m-1)k}$, and $\bar{\theta}$ converges to θ_0 in probability. Then $V(\bar{\theta})$ also converges to V in probability. Therefore it is very reasonable to use $\bar{\theta}$ for the unknown θ_0.

5. An example

In order to illustrate the theory developed in earlier sections, we consider some data given in Data Set II from Kalbfleisch and Prentice (1980). These data consist of survival times for patients with carcinoma of the oropharynx and several covariates. Patients diagnosed with squamous carcinoma of the oropharynx were classified by the degree to which the regional lymph nodes were affected by this disease. Since lymph node deterioration is an indication of the seriousness of the carcinoma, one would expect the populations with increased or decreased effects of the disease on the lymph nodes between them to be ordered in the increasing convex order. The data were grouped into seven classes in Table 1 of Dykstra *et al.* (1991). Just like in Y. Wang (1996), we delete all censored data and Group VII (in which most data are censored). We next merge every two next groups into a single group, that is, we group the data into three intervals (0,260], (260,540] and (540,900]. Table 1 lists n_{ij} for the grouped samples. We treat the grouped data as occurring at the interval midpoints. Thus we obtain four distributions with the common set of outcomes $b_1 = 130, b_2 = 400, b_3 = 720$. Let p_{i+1} be the probability distribution of Population $i, i = 0, 1, 2, 3$. Thus $m = 4, k = 2$, and $E = \{\theta : p_{i1} > 0, p_{i2} > 0, p_{i1} + p_{i2} < 1, i = 1, \ldots, 4\}$. Consider the hypothesis testing problem $H_1 : p_4 \le_{icx} p_3 \le_{icx} p_2 \le_{icx} p_1 \leftrightarrow H_2 - H_1$. Substituting $b_1 = 130, b_2 = 400$ and $b_3 = 720$ into (3.3) we get $S = \{\theta : A\theta \ge 0\}$, where

$$A = \begin{pmatrix} 1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 59 & 32 & -59 & -32 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 59 & 32 & -59 & -32 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 59 & 32 & -59 & -32 \end{pmatrix}.$$

Let $a = (33, 25, 11, 7, 9, 4, 4, 8, 5, 8, 14, 6)'$, $b(\theta) = (\log p_{41}, \log p_{42}, \log(1-p_{41}-p_{42}), \log p_{31}, \log p_{32}, \log(1-p_{31}-p_{32}), \log p_{21}, \log p_{22}, \log(1-p_{21}-p_{22}), \log p_{11}, \log p_{12}, \log(1-p_{11}-$

Table 1. Number of observations for grouped data.

Population	n_{i1}	n_{i2}	n_{i3}	n_i
Pop 0	8	14	6	28
Pop 1	4	8	5	17
Pop 2	7	9	4	20
Pop 3	33	25	11	69

$p_{12}))'$. Solving the optimization problem (4.2) with $N(\theta) = -a'b(\theta)$ we obtain

$$\hat{\theta} = (0.4783, 0.3623, 0.3500, 0.4500, 0.2667, 0.4889, 0.2667, 0.4889)'.$$

Then $T_{12} = 2[a'b(\tilde{\theta}) - a'b(\hat{\theta})] = 0.3912$, where $\tilde{\theta} = (33/69, 25/69, 7/20, 9/20, 4/17, 8/17, 8/28, 14/28)'$ is the unconstrained MLE of θ.

We now compute the p value for the test. First, we use Monte Carlo techniques to obtain the weights. Here $\bar{p}_i = (52/134, 56/134, 26/134)'$. From the expression of V^{-1} (see the proof of Theorem 1), a direct computation can give the matrix $\Delta = AV^{-1}A'$. By the equality (4.3) we must compute the weights $\omega_i(6,\Delta), i = 0,\ldots,6$.

We take 10,000 draws from a multivariate normal distribution with mean zero and covariance matrix Δ. For each sample point R we minimize $(x-R)'\Delta^{-1}(x-R)$ subject to $x = (x_1,\ldots,x_6)' \geq 0$. Denote the optimal solution by $\hat{x}$. Then we count the number of elements of the vector $\hat{x}$ greater than zero. In this case $\omega_i(6,\Delta)$ is computed as the proportion of the 10,000 draws in which $\hat{x}$ has exactly i elements greater than zero. Implementing this technique in Matlab program gives that $\omega_0(6,\Delta) = 0.0090$, $\omega_1(6,\Delta) = 0.0881$, $\omega_2(6,\Delta) = 0.2659$, $\omega_3(6,\Delta) = 0.3532$, $\omega_4(6,\Delta) = 0.2187$, $\omega_5(6,\Delta) = 0.0600$, $\omega_6(6,\Delta) = 0.0051$. Therefore, the p value for the test is

$$\begin{aligned}
&P(\bar{\chi}^2(V^{-1}, S^0) \geq 0.3912)\\
&\quad = \sum_{i=0}^{8} \omega_i(8, V^{-1}, S^0) P(\chi_i^2 \geq 0.3912)\\
&\quad = \sum_{i=0}^{6} \omega_{6-i}(6,\Delta) P(\chi_i^2 \geq 0.3912)\\
&\quad = 0.5317\omega_5(6,\Delta) + 0.8223\omega_4(6,\Delta) + 0.9421\omega_3(6,\Delta)\\
&\qquad +0.9832\omega_2(6,\Delta) + 0.9956\omega_1(6,\Delta) + 0.9989\omega_0(6,\Delta) = 0.9026.
\end{aligned}$$

This large p value indicates that the null hypothesis is true. This result gives a positive support to our proposed testing scheme.

Acknowledgements

The authors are grateful to the Editor, Associate Editor and the referees for their thoughtful comments and suggestions. This work has been supported by NSFC and NSFJS.

References

Attouch, H. (1985). *Variational Convergence for Function and Operators*, Pitman, London.

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

Bazaraa, M. S. and Shetty, C. M. (1979). *Nonlinear Programming: Theory and Algorithms*, Wiley, New York.

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

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

Dykstra, R. L. (1982). Maximum likelihood estimation of the survival functions of stochastically ordered random variables, *J. Amer. Statist. Assoc.*, **77**, 621–628.

Dykstra, R. L. and Feltz, C. J. (1989). Nonparametric maximum likelihood estimation of the 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, *Ann. Statist.*, **19**, 870–888.

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

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

Franck, W. E. (1984). A likelihood ratio test for stochastic ordering, *J. Amer. Statist. Assoc.*, **79**, 686–691.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*, Wiley, New York.

Kudô, A. (1963). Multivariate analogue of the one-sided test, *Biometrika*, **50**, 403–418.

Lee, Y. J. and Wolfe, D. A. (1976). A distribution-free test for stochastic ordering, *J. Amer. Statist. Assoc.*, **71**, 722–727.

Mann, H. B. and Wald, A. (1943). On stochastic limit and order relationships, *Ann. Math. Statist.*, **14**, 217–226.

Nüesch, P. (1966). On the problem of testing location in multivariate populations for restricted alternatives, *Ann. Math. Statist.*, **37**, 113–119.

Perlman, M. D. (1969). One-sided testing problems in multivariate analysis, *Ann. Math. Statist.*, **40**, 549–567.

Prakasa Rao, B. L. S. (1975). Tightness of probability measures generated by stochastic processes on metric spaces, *Bull. Inst. Math. Acad. Sinica*, **3**, 353–367.

Prakasa Rao, B. L. S. (1987). *Asymptotic Theory of Statistical Inference*, Wiley, New York.

Robertson, T. and Wegman, E. J. (1978). Likelihood ratio tests for order restrictions in exponential families, *Ann. Statist.*, **6**, 485–505.

Robertson, T. and Wright, F. T. (1981). Likelihood ratio tests for and against a stochastic ordering between multinomial populations, *Ann. Statist.*, **9**, 1248–1257.

Ross, S. M. (1983). *Stochastic Processes*, Wiley, New York.

Shaked, M. and Shanthikumar, J. G. (1994). *Stochastic Orders and Their Applications*, Academic Press, New York.

Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis, *International Statistical Review*, **56**, 49–62.

Stoyan, D. (1983). *Comparison Methods for Queues and Other Stochastic Models*, Wiley, New York.

Wang, J. (1996). The asymptotics of least-squares estimators for constrained nonlinear regression, *Ann. Statist.*, **24**, 1316–1326.

Wang, Y. (1996). A likelihood ratio test against stochastic ordering in several population, *J. Amer. Statist. Assoc.*, **91**, 1676–1683.