Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Multivariate Analysis 97 (2006) 1894–1912

Journal of Multivariate Analysis

www.elsevier.com/locate/jmva

A unified approach to testing for and against a set of linear inequality constraints in the product multinomial setting

Hammou El Barmi*, Matthew Johnson

Department of Statistics and Computer Information Systems, Baruch College, City University of New York, Box 11-220, One Boruch Way, NY 10010, USA

Received 20 December 2004
Available online 8 August 2005

Abstract

A problem that is frequently encountered in statistics concerns testing for equality of multiple probability vectors corresponding to independent multinomials against an alternative they are not equal. In applications where an assumption of some type of stochastic ordering is reasonable, it is desirable to test for equality against this more restrictive alternative. Similar problems have been considered heretofore using the likelihood ratio approach. This paper aims to generalize the existing results and provide a unified technique for testing for and against a set of linear inequality constraints placed upon on any r $(r \geqslant 1)$ probability vectors corresponding to r independent multinomials. The paper shows how to compute the maximum likelihood estimates under all hypotheses of interest and obtains the limiting distributions of the likelihood ratio test statistics. These limiting distributions are of chi bar square type and the expression of the weighting values is given. To illustrate our theoretical results, we use a real life data set to test against second-order stochastic ordering.

AMS 1991 subject classification: 62F30; 62H17; 60E15

Keywords: Chi bar square; Inequality constraints; Lagrange multipliers; Likelihood ratio; Orthant probabilities; Stochastic ordering

* Corresponding author. Fax.: +1 646 312 3351.
E-mail address: hammou_elbarmi@baruch.cuny.edu (H. El Barmi).

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2005.06.006

1. Introduction

A commonly occurring problem in statistics is that of testing for equality of two probability vectors corresponding to independent multinomials against an alternative that they are not equal. Sometimes it is reasonable to assume that these vectors satisfy some type of a stochastic ordering and it might be of interest to test for equality against such an assumption. For example, Robertson and Wright [16] consider testing for equality of two probability vectors against the alternative that they are stochastically ordered. They obtain the maximum likelihood estimates and show that the likelihood ratio test statistic has, asymptotically, a chi bar square distribution and give the expression of the weighting values. Wang [23] extends their work to more than two probability vectors.

In this paper, we present a unified approach to testing for or against a set of linear inequality constraints placed upon $r \geqslant 1$ probability vectors corresponding to r independent multinomial distributions (if $r \geqslant 2$). The results here extend those in [6] and the approach is based on the results in [22].

Specifically, let $\mathbf{p}_i = (p_{i1}, p_{i2}, \ldots, p_{ik})^T$ denote the probability vector corresponding to the ith distribution, $1 \leqslant i \leqslant r$, and consider testing $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_0$ and $\mathcal{H}_1$ against $\mathcal{H}_2 - \mathcal{H}_1$, where

$$\mathcal{H}_0 : \sum_{i=1}^{r} \sum_{j=1}^{k} x_{ij}^{(s)} p_{ij} = 0, \ s = 1, 2, \ldots, c, \tag{1.1}$$

$$\mathcal{H}_1 : \sum_{i=1}^{r} \sum_{j=1}^{k} x_{ij}^{(s)} p_{ij} \leqslant 0, \ s = 1, 2, \ldots, c. \tag{1.2}$$

$\mathcal{H}_2$ imposes no constraints on the probability vectors. Here $c \leqslant r(k-1)$ and the $x_{ij}^{(s)}$s are fixed and known constants.

Throughout the paper we also use $\mathcal{H}_i$ to denote the set of all the probability vectors that satisfy the constraints in $\mathcal{H}_i$ and assume that $\mathcal{H}_i \neq \emptyset$. It is well known [20] that the likelihood ratio test statistic for testing $\mathcal{H}_0$ against $\mathcal{H}_2 - \mathcal{H}_0$ has, asymptotically, a chi-square distribution with c degrees of freedom. This paper extends the existing results and obtains the test statistics for testing $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_0$ and $\mathcal{H}_1$ against $\mathcal{H}_2 - \mathcal{H}_1$ as well as their limiting distributions, which are shown to be of a chi bar square type. To illustrate our theoretical results, we consider the problem of the testing against second-order stochastic ordering. This type of ordering of distributions is weaker that the regular stochastic ordering. A random variable X with distribution function F is second-order stochastically smaller that a random variable Y with distribution function G if

$$\int_x^\infty (1 - F(u))\, du \leqslant \int_x^\infty (1 - G(u))\, du \quad \text{for all } x.$$

This ordering plays a prominent role in the general framework of analyzing choice under uncertainty by considering the maximization of the expected utilities [14]. More specifically, a risk averter prefers an investment portfolio $\mathcal{B}$ with random return Y over an investment $\mathcal{A}$ with random return X if and only if $E(U(Y)) \geqslant E(U(X))$ for all nondecreasing and concave utility functions U. It turns out that this condition is equivalent to Y being second-order stochastically larger than X [19]. Liu and Wang [15] consider testing for and against this type of ordering when $r = 2$ using grouped data and our results extend their work to $r > 2$.

The rest of the paper is organized as follows. In Section 2, we show how to compute the maximum likelihood estimators under all the hypotheses of interest. Section 3 derives the limiting distributions of the test statistics and Section 4 gives examples to illustrate our theoretical results. In Section 5, we give some concluding remarks and the proofs are given in the appendix in Section 6.

2. Estimation

Consider independent multinomial random vectors $(X_{i1}, X_{i2}, \ldots, X_{ik})^T, i = 1, 2, \ldots, r$. Assume that the ith vector summarizes the results of observing the outcome of n_i independent random experiments, each of which can result in any one of k mutually exclusive outcomes, $1, 2, \ldots, k$, with positive probabilities $p_{i1}, p_{i2}, \ldots, p_{ik}$, respectively. Let $\mathbf{p}_i = (p_{i1}, p_{i2}, \ldots, p_{ik})^T$ and $\mathbf{p} = (\mathbf{p}_1^T, \mathbf{p}_2^T, \ldots, \mathbf{p}_k^T)^T$.

The likelihood (except for a multiplicative constant) for any outcome $\{(n_{i1}, n_{i2}, \ldots, n_{ik})^T, i = 1, 2, \ldots, r\}$ is given by

$$\mathcal{L}(\mathbf{p}) = \prod_{i=1}^{r} \mathcal{L}_i(\mathbf{p}_i) \propto \prod_{i=1}^{r} \prod_{j=1}^{k} p_{ij}^{n_{ij}}. \tag{2.3}$$

The unrestricted maximum likelihood estimate of p_{ij} is given by $\hat{p}_{ij} = n_{ij}/n_i$ but in general, under $\mathcal{H}_0$ and $\mathcal{H}_1$, the maximum likelihood estimates do not exist in a closed form. El Barmi and Dykstra's algorithm [6,7] can be utilized to compute the maximum likelihood estimates under these two restricted hypotheses. Specifically, El Barmi and Dykstra [5–7] show that, if, for $s = c + 1, c + 2, \ldots, c + r - 1$ and $j = 1, 2, \ldots, k$,

$$x_{ij}^{(s)} = \begin{cases} 1, & s = c + i, \\ -1, & s = c + i + 1, \\ 0 & \text{otherwise} \end{cases}$$

and $\tilde{p}_{ij} = n_{ij}/n$, for all (i, j) where $n = \sum_{i=1}^{r} n_i$, then the maximum likelihood estimate $\hat{p}_{ij}^{(0)}$ of p_{ij} under $\mathcal{H}_0$ is given by

$$\hat{p}_{ij}^{(0)} = \frac{r\,\tilde{p}_{ij}}{1 + \sum_{s=1}^{c+r-1} \hat{\alpha}_s^{(0)} x_{ij}^{(s)}},$$

where the $\hat{\alpha}_s^{(0)}$s solve

$$\max \sum_{i=1}^{r} \sum_{j=1}^{k} \tilde{p}_{ij} \ln\left(1 + \sum_{s=1}^{r+c-1} \alpha_s x_{ij}^{(s)}\right). \tag{2.4}$$

They also show that the maximum likelihood estimate $\hat{p}_{ij}^{(1)}$ of p_{ij} under $\mathcal{H}_1$ is given by

$$\hat{p}_{ij}^{(1)} = \frac{r\tilde{p}_{ij}}{1 + \sum_{l=1}^{c+r-1} \hat{\alpha}_s^{(1)} x_{ij}^{(s)}},$$

where the $\hat{\alpha}_s^{(1)}$s solve

$$\max \sum_{i=1}^{r} \sum_{j=1}^{k} \tilde{p}_{ij} \ln \left(1 + \sum_{s=1}^{c+r-1} \alpha_s x_{ij}^{(s)}\right) \tag{2.5}$$

subject to $\alpha_s \geqslant 0$, $1 \leqslant s \leqslant c$ and $\alpha_s \in \mathcal{R}$ if $s \geqslant c+1$.

El Barmi and Dykstra [6] provide the following iterative algorithm which is guaranteed to converge to the true solution to find the $\hat{\alpha}_s^{(0)}$s and the $\hat{\alpha}_s^{(1)}$s, the maximizing values of (2.4) and (2.5), respectively.

Algorithm.

Step 1: initialize $\alpha^s = 0$, $s = 1, 2, \ldots, s = c + r - 1$, $v = 1$.

Step 2: Find the optimal value of α_v over $\mathcal{R}$ with all the other αs held fixed. This value of α_v replaces its previous value.

- If $v < c + r - 1$ set $v = v + 1$, if $v = c + r - 1$, set $v = 1$.
- Go to step 2.

These steps are repeated for $v = 1, 2, \ldots$ until sufficient accuracy is attained. We note that $\hat{\alpha}_1^{(0)}, \hat{\alpha}_2^{(0)}, \ldots, \alpha_{c+r-1}^{(0)}$ are the Lagrange multipliers corresponding to maximizing the likelihood function under $\mathcal{H}_0$ and subject to $(\mathbf{p}_1^T, \mathbf{p}_2^T, \ldots, \mathbf{p}_r^T)^T \in \mathcal{P}^r$ where $\mathcal{P}$ is the set of the probability vectors in $\mathcal{R}^k$.

To compute $\hat{\alpha}_1^{(1)}, \hat{\alpha}_2^{(1)}, \ldots, \hat{\alpha}_{c+r-1}^{(1)}$, Step 2 of the algorithm is replaced by Step 2′: Find the optimal value of α_v over $\mathcal{R}$ with all the other αs held fix. Whenever $v \leqslant c$ and this value of α_v is non-negative, it replaces its previous value, otherwise we use 0.

3. Hypotheses testing

In this section, we consider testing $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_0$ and $\mathcal{H}_1$ against $\mathcal{H}_2 - \mathcal{H}_1$. We obtain the likelihood ratio test statistics and show that the limiting distributions are of chi bar square type and provide the expression of the weighting values. Throughout the rest of the paper, we assume that $\gamma_i = \lim_{n\to\infty} n_i/n > 0$, for all i.

Let Λ_{ij} denote the likelihood ratio test statistic for testing $\mathcal{H}_i$ against $\mathcal{H}_j - \mathcal{H}_i$, $(i, j) =$ (0, 1) or (1,2). The likelihood ratio approach rejects $\mathcal{H}_i$ in favor of $\mathcal{H}_j$ for large values of $T_{ij} = -2 \ln \Lambda_{ij}$.

Let $\tilde{B} = diag[\frac{1}{\gamma_1}B_1, \frac{1}{\gamma_2}B_2, \ldots, \frac{1}{\gamma_r}B_r]$ where $B_i = (p_{is}(\delta_{st} - p_{it})_{s,t\neq k})$. For $s = 1, 2, \ldots, c$, $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, k-1$, let $y_{ij}^{(s)} = x_{ij}^{(s)} - x_{ik}^{(s)}$ and let H

be a matrix whose transpose is

$$H^T = \begin{bmatrix} y_{11}^{(1)} & \cdots & y_{1,k-1}^{(1)} & y_{21}^{(1)} & \cdots & y_{2,k-1}^{(1)} & \cdots & y_{r1}^{(1)} & \cdots & y_{r,k-1}^{(1)} \\ y_{11}^{(2)} & \cdots & y_{1,k-1}^{(2)} & y_{21}^{(2)} & \cdots & y_{2,k-1}^{(2)} & \cdots & y_{r1}^{(2)} & \cdots & y_{r,k-1}^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{11}^{(c)} & \cdots & y_{1,k-1}^{(c)} & y_{21}^{(c)} & \cdots & y_{2,k-1}^{(c)} & \cdots & y_{r1}^{(c)} & \cdots & y_{r,k-1}^{(c)} \end{bmatrix}$$
$$= [H_1^T \mid H_2^T \mid \ldots \mid H_k^T],$$

where for $1 \leqslant l \leqslant k$

$$H_l^T = \begin{bmatrix} y_{l1}^{(1)} & \cdots & y_{l,k-1}^{(1)} \\ y_{l1}^{(2)} & \cdots & y_{l,k-1}^{(2)} \\ \vdots & \vdots & \vdots \\ y_{l1}^{(c)} & \cdots & y_{l,k-1}^{(c)} \end{bmatrix}.$$

Assume that H has full rank and define the matrices P, Q and R by

$$\begin{bmatrix} \tilde{B}^{-1} & -H \\ -H^T & 0 \end{bmatrix} = \begin{bmatrix} \mathrm{P} & \mathrm{Q} \\ Q^T & \mathrm{R} \end{bmatrix}. \tag{3.6}$$

A direct computation of these matrices shows that

$$R = -[H^T \tilde{B} H]^{-1} = -\left[\sum_{i=1}^{r} \frac{1}{\gamma_i} H_i^T B_i H_i\right]^{-1},$$

$$Q = -\tilde{B} H [H^T \tilde{B} H]^{-1} = \begin{bmatrix} \frac{1}{\gamma_1} B_1 H_1 R \\ \frac{1}{\gamma_2} B_2 H_2 R \\ \cdots \\ \frac{1}{\gamma_r} B_r H_r R \end{bmatrix}$$

and $P = \tilde{B} - \tilde{B} H [H^T \tilde{B} H]^{-1} H^T \tilde{B} = (P_{ij})_{1 \leqslant i,j \leqslant r}$ where

$$P_{ii} = \frac{1}{\gamma_i} B_i + \frac{1}{\gamma_i^2} B_i H_i R H_i^T B_i$$

and

$$P_{ij} = P_{ji}^T = \frac{1}{\gamma_i \gamma_j} B_i H_i R H_j^T B_j.$$

Let $\overline{\mathbf{p}}_i = (\overline{p}_{i1}, \overline{p}_{i2}, \ldots, \overline{p}_{i,k-1})^T$ and $\overline{\mathbf{p}} = (\overline{\mathbf{p}}_1^T, \overline{\mathbf{p}}_2^T, \ldots, \overline{\mathbf{p}}_k^T)^T$. For technical reasons (namely, to avoid dealing with singular matrices), we express the likelihood in terms of $\overline{\mathbf{p}}_i$s, that is, our likelihood function is

$$\mathcal{L}(\overline{\mathbf{p}}) = \prod_{i=1}^{r} \mathcal{L}_i(\overline{\mathbf{p}}_i) \propto \prod_{i=1}^{r} \left[\prod_{j=1}^{k-1} p_{ij}^{n_{ij}} \left(1 - \sum_{j=1}^{k-1} p_{ij}\right)^{n_{ik}} \right]. \tag{3.7}$$

Define

$$\mathcal{D} \ln \mathcal{L}(\overline{\mathbf{p}}) = \left(\mathcal{D} \ln \mathcal{L}_1(\overline{\mathbf{p}}_1)^T, \ldots, \mathcal{D} \ln \mathcal{L}_r(\overline{\mathbf{p}}_r)^T\right)^T,$$

where

$$\mathcal{D} \ln \mathcal{L}_i(\overline{\mathbf{p}}_i) = \left(\frac{\partial}{\partial p_{ij}} \ln \mathcal{L}_i(\mathbf{p}_i)\right)_{j \neq k}.$$

If $\overline{\mathbf{p}}^0 = (\overline{\mathbf{p}}_1^{0T}, \ldots, \overline{\mathbf{p}}_r^{0T})^T \in \mathcal{H}_0$ is the true value of $\overline{\mathbf{p}} = (\mathbf{p}_1^T, \mathbf{p}_2^T, \ldots, \mathbf{p}_r^T)^T$, similar arguments used in [22] give as $n \to \infty$,

$$\sqrt{n}[\hat{\overline{\mathbf{p}}}^{0T} - \overline{\mathbf{p}}^{0T}] = \frac{1}{\sqrt{n}} P^0 \mathcal{D} \ln \mathcal{L}(\overline{\mathbf{p}}^0) + o_p(1) \tag{3.8}$$

and

$$\frac{1}{\sqrt{n}}[\hat{\lambda}_1^{(0)}, \ldots, \hat{\lambda}_c^{(0)}]^T = \frac{1}{\sqrt{n}} Q^{0T} \mathcal{D} \ln \mathcal{L}(\overline{\mathbf{p}}^0) + o_p(1), \tag{3.9}$$

where $\hat{\boldsymbol{\lambda}}^0 = (\hat{\lambda}_1^{(0)}, \ldots, \hat{\lambda}_c^{(0)})^T$ is the vector of Lagrange multipliers corresponding to maximizing (3.7) under $\mathcal{H}_0$ and P^0 and Q^0 denote the matrices P and Q under $\mathcal{H}_0$.

Combining (3.8) and (3.9) gives the following theorem which is a generalization of a theorem provided by Silvey [22] for the one sample case.

Theorem 3.1. *Under* $\mathcal{H}_0$,

$$\sqrt{n}[\hat{\overline{\mathbf{p}}}^{(0)T} - \overline{\mathbf{p}}^{(0)T}, \frac{1}{n}\hat{\boldsymbol{\lambda}}^{(0)T}]^T \xrightarrow{d} N(\mathbf{0}, V),$$

where

$$V = \begin{bmatrix} P^0 & 0 \\ 0 & -R^0 \end{bmatrix}.$$

Proof. The proof follows from combining Eqs. (3.8) and (3.9), the fact that $\frac{1}{\sqrt{n}} \mathcal{D} \ln \mathcal{L}(\overline{\mathbf{p}}^0)$ converges in distribution to $N(\mathbf{0}, [\tilde{B}^0]^{-1})$ under $\mathcal{H}_0$ and the definition of P^0, Q^0 and R^0. Here $\tilde{B}^0$ is the value of $\tilde{B}$ under $\mathcal{H}_0$.

As a consequence of this theorem, $-\frac{1}{n}\hat{\boldsymbol{\lambda}}^{0T}[R^0]^{-1}\hat{\boldsymbol{\lambda}}^0$ has asymptotically a χ_c^2 and can be used to test $\mathcal{H}_0$ against $\mathcal{H}_2 - \mathcal{H}_0$ as proposed in [2] for the one sample case.

In order to establish the distributions of the likelihood ratio test statistics for testing $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_0$ and $\mathcal{H}_1$ against $\mathcal{H}_2 - \mathcal{H}_1$, we consider first testing $\mathcal{H}_0$ against $\mathcal{H}_{1:\pi} - \mathcal{H}_0$ where

$$\mathcal{H}_{1:\pi} : \sum_{i=1}^{r} \sum_{j=1}^{k} x_{ij}^{(s)} p_{ij} = 0, \quad s \in \pi \tag{3.10}$$

and $\pi \subset \{1, 2, \ldots, c\}$.

Let $\hat{\overline{\mathbf{p}}}^{(0)}(\pi) = (\hat{\overline{\mathbf{p}}}_1^{0T}(\pi), \ldots, \hat{\overline{\mathbf{p}}}_r^{0T}(\pi))^T$ denote the maximum likelihood estimator of $\overline{\mathbf{p}}$ under $\mathcal{H}_{1:\pi}$ and $\hat{\boldsymbol{\lambda}}^{(0)}(\pi)$ denote the vector of the Lagrange multipliers corresponding to maximizing (3.7) under $\mathcal{H}_{1:\pi}$. Then,

$$\sqrt{n}[\hat{\overline{\mathbf{p}}}^{(0)}(\pi) - \overline{\mathbf{p}}^{(0)}] = \frac{1}{\sqrt{n}} P^0(\pi)\mathcal{D}\ln \mathcal{L}(\overline{\mathbf{p}}^0) + o_p(1) \tag{3.11}$$

and

$$\frac{1}{\sqrt{n}}\hat{\boldsymbol{\lambda}}(\pi) = \frac{1}{\sqrt{n}} Q^{0T}(\pi)\mathcal{D}\ln \mathcal{L}_n(\overline{\mathbf{p}}^0) + o_p(1), \tag{3.12}$$

where $P^0(\pi)$ and $Q^0(\pi)$ are the values of P^0 and Q^0 in (3.6) when H is replaced by $H(\pi)$. □

The following result is well known [20] but we include it for the development of the main theorem (Theorem 3.3).

Theorem 3.2. *Under $\mathcal{H}_0$ and for any t, we have*

$$\lim_{n\to\infty} P(T_{01:\pi} \geqslant t) = P(\chi^2_{c-card(\pi)} \geqslant t).$$

Proof. See Appendix. □

Next we consider testing $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_0$ and $\mathcal{H}_1$ against $\mathcal{H}_2 - \mathcal{H}_1$. Let $\mathcal{F}$ denote the class of all subsets of $\{1, 2, \ldots, c\}$ and let T_{01} (T_{12}) denote the log-likelihood ratio test statistic for testing $\mathcal{H}_0$ $(\mathcal{H}_1)$ against $\mathcal{H}_1 - \mathcal{H}_0$ $(\mathcal{H}_2 - \mathcal{H}_1)$. For a proper subset π of $\{1, 2, \ldots, c\}$, with complement $\tilde{\pi}$, define

$$a_j(\mathbf{p}) = \sum_{\pi, card(\pi)=j} P(N(0, \Sigma_1(\pi)) \geqslant 0) P(N(0, \Sigma) > 0),$$

where

$$\Sigma_1(\pi) = [H^T(\pi)\tilde{B}^0 H(\pi)]^{-1}$$

and

$$\Sigma(\pi) = H^T(\tilde{\pi})\tilde{B}^0 H(\tilde{\pi}) - H^T(\tilde{\pi})\tilde{B}^0 H(\tilde{\pi})\Sigma_1 H^T(\tilde{\pi})\tilde{B}^0 H(\tilde{\pi}).$$

Let

$$a_0(\mathbf{p}) = P(N(0, \Sigma_0) \geqslant 0)$$

and

$$a_c(\mathbf{p}) = P(N(0, \Sigma_c) \geqslant 0),$$

where $\Sigma_0 = \Sigma_c^{-1} = H^T\tilde{B}^0 H$. These weights, which are sums of products of normal orthant probabilities, do not exist in general in a closed form (see [13,18] for more discussion on this as well as related references). In the analysis in Section 4.2 we use the algorithm suggested by Genz [10,11] to approximate the orthant probabilities $a_j(\mathbf{p})$ above, which play an important role in the following theorem.

Theorem 3.3. *Under $\mathcal{H}_0$ and for any t_1 and t_2, we have*

$$\lim_{n\to\infty} P(T_{01}\geqslant t_1, T_{12} > t_2) = \sum_{j=0}^{c} a_j(\mathbf{p}^{(0)}) P(\chi^2_{c-j}\geqslant t_1) P(\chi^2_j > t_2),$$

where $\chi^2_0 \equiv 0$.

As a consequence of this theorem, we have under $\mathcal{H}_0$

$$\lim_{n\to\infty} P(T_{01}\geqslant t) = \sum_{j=0}^{c} a_j(\mathbf{p}^{(0)}) P(\chi^2_{c-j}\geqslant t)$$

and

$$\lim_{n\to\infty} P(T_{12}\geqslant t) = \sum_{j=0}^{c} a_j(\mathbf{p}^{(0)}) P(\chi^2_{j}\geqslant t).$$

It is the case that

$$\lim_{n\to\infty} P(T_{01}\geqslant t_1) = \lim_{n\to\infty} \sum_{j=0}^{c} a_j(\hat{\mathbf{p}}^{0}) P(\chi^2_{c-j}\geqslant t) \tag{3.13}$$

almost surely if $\hat{\mathbf{p}}^0$ is a consistent estimator of $\mathbf{p}$ under $\mathcal{H}_0$. A natural estimator for $\mathbf{p}$ is its maximum likelihood estimator under $\mathcal{H}_0$. Empirical evidence suggests that critical points and p-values obtained from

$$\sum_{j=0}^{c} a_j(\hat{\mathbf{p}}^{(0)}) P(\chi^2_{c-j}\geqslant t) \tag{3.14}$$

work well due to lack of sensitivity of the level probabilities to changes in the weights [17].

4. Examples

In this section we discuss two examples to illustrate our theoretical results. The hypotheses in (1.1) and (1.2) can be utilized to test for and against different types of stochastic ordering.

4.1. Example 1

Here we assume that $r = 2$ and wish to test $\mathcal{H}_0 : p_{1i} = p_{2i},\ i = 1, 2, \ldots, k-1$, against $\mathcal{H}_1 - \mathcal{H}_0$ where $\mathcal{H}_1 : \sum_{i=1}^{j} p_{1i} \leqslant \sum_{i=1}^{j} p_{2i},\ j = 1, 2, \ldots, k-1$. That is, $\mathbf{p}_1$ is stochastically larger than $\mathbf{p}_2$. In this case the matrix $H^T = [H_1^T | H_2^T]$ where

$$H_1^T = -H_2^T = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix}.$$

It is easy to see that under $\mathcal{H}_0$,

$$H^T \tilde{B}^0 H = \left[\frac{1}{\gamma_1} + \frac{1}{\gamma_2}\right] H^T B_1^0 H.$$

A direct computation of the (i, j)th element of this symmetric matrix shows that

$$(H^T \tilde{B}^0 H)_{ij} \propto \left(\sum_{l=1}^{i} p_{1l}^0\right)\left(1 - \sum_{l=1}^{j} p_{1l}^0\right), \quad 1 \leqslant i \leqslant j \leqslant k-1$$

and its inverse is symmetric with its $(i, j)th$ element given by

$$[(H^T \tilde{B}^0 H)^{-1}]_{ij} \propto \begin{cases} \frac{1}{p_{1i}^0} + \frac{1}{p_{1,i+1}^0} & \text{if } i = j, \\ -\frac{1}{p_{1,i+1}^0} & \text{if } j = i+1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $Y_1, Y_2, \dots, Y_k$ be independent random variables with mean zero and variances $1/p_{11}^0$, $1/p_{12}^0, \dots, 1/p_{1k}^0$, respectively. Let also $U_i = Y_{i+1} - Y_i, i = 1, 2, \dots, k-1$. Then $(U_1, U_2, \dots, U_{k-1})^T$ has a multivariate normal distribution with zero mean vector and covariance matrix proportional to $[H^T \tilde{B}^0 H]^{-1}$. Therefore

$$\begin{aligned} a_{k-1}(\mathbf{p}^0) &\overset{d}{=} P(N(\mathbf{0}, [H^T \tilde{B}^0 H]^{-1}) \geqslant \mathbf{0}) \\ &= P(U_1 \geqslant 0, U_2 \geqslant 0, \dots, U_{k-1} \geqslant 0) \\ &= P(k, k, \mathbf{p}^0), \end{aligned}$$

where $P(k, k, \mathbf{p}^0)$ is the probability that the least-squares projection of $(Y_1, Y_2, \dots, Y_k)^T$ onto $\mathcal{I} = \{\mathbf{x} \in R^k,\ x_1 \leqslant x_2 \leqslant \dots \leqslant x_k\}$ with weights $p_{1i}^0, i = 1, 2, \dots, k$, has exactly k distinct levels [18]. It can also be shown that $a_j(\mathbf{p}^0) = P(j+1, k, \mathbf{p}^0)$, $j = 0, 1, \dots, k-2$. Therefore, under $\mathcal{H}_0$,

$$\lim_{n\to\infty} P(T_{01} \geqslant t_1, T_{12} > t_2) = \sum_{l=1}^{k} P(l, k, \mathbf{p}^{(0)}) P(\chi_{k-j}^2 \geqslant t_1) P(\chi_j^2 > t_2)$$

which gives the result in [16].

4.2. Example 2: Testing against second-order stochastic ordering

Data from [12] consists of the survival times and several covariates for 195 patients suffering from carcinoma of the oropharynx; approximately 26% of the survival times are

Table 1
Observed frequencies for the oropharynx carcinoma data set

Population	Survival times					
	0–160 d_{i1}	161–260 d_{i2}	261–360 d_{i3}	361–540 d_{i4}	541–900 d_{i5}	n_i
Population 0	3	5	5	9	6	28
Population 1	2	2	6	2	5	17
Population 2	2	5	4	5	4	20
Population 3	17	16	13	12	11	69

censored. One of the covariates is an ordinal categorical variable with four levels, which indicates increasing levels of deterioration of lymph nodes in each patient, measured at time of entry in the study. Because lymph node deterioration is an indication of the seriousness of the carcinoma, it is reasonable to expect that the four survival time distributions would be stochastically ordered by the severity of the lymph node deterioration.

Feltz and Dykstra [9], Dykstra and Feltz [3], and Dykstra et al. [4] examine the data under the assumption of uniform stochastic ordering. Dykstra et al. [4] collapse the survival times into seven classes as indicated in Table 1 of their paper, and find that the data provide evidence against the null hypothesis of equal survival functions in favor of uniform stochastic ordering.

Wang [23], in an effort to examine the hypothesis of stochastically ordered survival functions, removes censored data and patients with the longest survival times (Group VII in [4]) and merges Groups V and VI. The resulting data is presented in Table 1. The four populations (0, 1, 2, 3) correspond to the four levels of lymph node deterioration and the survival times correspond to the ranges defining each of the five groups of data. Wang [23] goes on to show that there is no statistical evidence to reject the hypothesis that the first three populations are the same. The paper then finds that there is some statistical evidence (p-$value$ $= 0.091$) to support the claim that Population 3 stochastically dominates the pooled Population found by combining Populations 0,1 and 2.

Liu and Wang [15] test the hypothesis of equality of the survival functions versus an alternative of second stochastic ordering for the same collapsed data set by collapsing cells 3, 4 and 5 and combining the populations 0, 1 and 2.

Here we use the full table to test for second-order stochastic ordering. In order to apply our approach we assume that all the observations in a given interval are equal to the mid-interval point, which are located at $t_1 = 80$, $t_2 = 210$, $t_3 = 310$, $t_4 = 450$, and $t_5 = 720$. Further, assume that

$$P(X_i = t_j) = p_{ij}, \quad i = 0, 1, 2, 3, \ j = 1, 2, \ldots, 5.$$

We wish to test $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_0$ where

$$\mathcal{H}_0 \ : \ p_{0j} = p_{1j} = p_{2j} = p_{3j}, \quad j = 1, 2, \ldots, 5$$

Table 2
Fitted values for the number of patients within each survival time class for the full data set under the assumption of second stochastic ordering

Population	Group				
	I	II	III	IV	V
Pop. 0	2.88	4.81	4.81	8.65	6.84
Pop. 1	2.14	2.14	6.42	2.14	4.16
Pop. 2	2.00	5.00	4.00	5.00	4.00
Pop. 3	17.00	16.00	13.00	12.00	11.00

$$\begin{aligned}\mathcal{H}_1 \ : \ & \int_x^\infty (1 - F_0(u))\, du \geqslant \int_x^\infty (1 - F_1(u))\, du \geqslant \int_x^\infty (1 - F_2(u))\, du \\ & \geqslant \int_x^\infty (1 - F_3(u))\, \mathrm{d}u\end{aligned}$$

for all $x \geqslant 0$. Under our assumption, $\mathcal{H}_1$ reduces to

$$\begin{aligned} p_{i4} &\leqslant p_{i+1,4}, \\ \sum_{j=1}^{3} (t_5 - t_3) p_{ij} + p_{i4}(t_5 - t_4) &\leqslant \sum_{j=1}^{3} (t_5 - t_3) p_{i+1,j} + p_{i+1,4}(t_5 - t_4), \\ \sum_{j=1}^{2} (t_5 - t_2) p_{ij} + p_{i3}(t_5 - t_3) + p_{i4}(t_5 - t_4) &\leqslant \sum_{j=1}^{2} (t_5 - t_3) p_{i+1,j} + p_{i+1,3}(t_5 - t_3) \\ &\quad + p_{i+1,4}(t_5 - t_4), \\ \sum_{j=1}^{4} (t_5 - t_j) p_{ij} &\leqslant \sum_{j=1}^{4} (t_5 - t_j) p_{i+1,j}, \end{aligned}$$

$i = 0, 1, 2.$

The advantage of our general method over the works of Wang [23] and Liu and Wang [15] is that it can handle the case of testing equality of several populations classified into multiple groups versus an alternative of second stochastic ordering. To demonstrate we fit the entire data set in Table 1, under the assumption of second stochastic ordering. The fitted cell counts are reported in Table 2. The value of the test statistics for testing for equality of the probability vectors versus second stochastic ordering is 10.43, which yields a p-value of 0.316. This p-value is computed according to (3.13) with approximated weights.

Wang [23] finds no statistical evidence against the assumption of equality of the first three populations (Pops 0–2) in Table 1, So, like Wang [23] we test for equality of the first three populations. However, we test the hypothesis against an alternative of second stochastic ordering. The fitted cell counts appear in Table 3. The test statistic for testing the hypothesis of equality of the probability vectors versus an alternative of second stochastic ordering is 4.35, which yields a p-value of 0.648 using (3.13) with approximated weights. So, like Wang [23], we find no statistical evidence of differences between the probability vectors.

Table 3
Fitted values for the number of patients within each survival time class for the first three classes of patients under the assumption of second stochastic ordering

Population	Group				
	I	II	III	IV	V
Pop. 0	2.88	4.81	4.81	8.65	6.84
Pop. 1	2.14	2.14	6.42	2.14	4.16
Pop. 2	2.00	5.00	4.00	5.00	4.00

Table 4
Fitted values for the number of patients within each survival time class for the collapsed data set under the assumption of second stochastic ordering

Population	Group				
	I	II	III	IV	V
Pops 0–2	7	12	15	16	15
Pop. 3	17	16	13	12	11

Table 4 contains the fitted frequencies for the data set obtained after collapsing the first three populations. The observed data satisfies the assumption of second stochastic ordering, and therefore the observed frequencies are fit exactly. The test statistic for comparing the null hypothesis of equal survivals against the alternative of second stochastic ordering is 6.08, and the *p-value* $= 0.173$ using (3.13) with approximated weights; there is not significant evidence against the assumption of equality in favor of second stochastic ordering.

4.3. Example 3

The data is this example is the result of a clinical trial regarding the outcome for patients who experienced trauma due to subarachnoid hemorrhage are given in Table [21,1]. In this case there are four treatments (Placebo=1, Low Dose=2, Medium Dose=3 and High Dose=4) and five possible outcomes (Death=1, Vegetative State=2, Major Disability=3, Minor Disbaility=4 and Good Recovery =5). For $i = 1, 2, \ldots, 4$ and $j = 1, \ldots, 5$, let

$$p_{ij} = P(\text{Outcome} = j \,|\, \text{Treatment} = i).$$

Consider testing the null hypothesis of no treatment effect against the alternative of higher dose being more effective. Specifically, we want to test $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_0$ where

$$\mathcal{H}_0 : p_{1j} = p_{2j} = p_{3j} = p_{4j}, \quad j = 1, \ldots, 5,$$

Table 5
The results of a clinical trial comparing the effectiveness of varying levels of a treatment on patients who lead suffered from a subarachnoid hemorrhage

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

Table 6
The fitted values of the clinical trial data under the null hypothesis, which states that there is no difference between dosages

Treatment	Death	Vegitative state	Major disability	Minor disability	Good recovery
Placebo	50.20	16.82	50.73	57.03	35.22
Low dose	45.42	15.22	45.89	51.60	31.86
Medium dose	48.77	16.34	49.28	55.40	34.21
High dose	46.61	15.62	47.10	52.96	32.70

Table 7
The fitted values of the clinical trial data under the alternative hypothesis, which states that the outcome distribution is stochastically ordered by the level of the dosage

Treatment	Death	Vegitative state	Major disability	Minor disability	Good recovery
Placebo	59.00	25.00	46.00	48.00	32.00
Low dose	48.23	21.10	44.21	47.23	29.22
Medium dose	42.85	13.76	53.09	62.92	31.38
High dose	40.96	4.05	49.66	58.78	41.55

and

$$\mathcal{H}_1 : \sum_{l=1}^{j} p_{4l} \leqslant \sum_{l=1}^{j} p_{3l} \leqslant \sum_{l=1}^{j} p_{2l} \leqslant \sum_{j=1}^{j} p_{1l}, \quad j = 1, \ldots, 4.$$

The fitted values under $\mathcal{H}_0$ and $\mathcal{H}_1$ are given, respectively, in Tables 6 and 7 The values of the test statistic for testing $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_0$ is 28.43 and the p-value based on (3.14) is 0.00028.

5. Concluding remarks

In this paper we have shown how to test for or against a set of linear inequality constraints placed upon the probability vectors of independent multinomials using the likelihood approach. Examples of this include testing for or against second-order stochastic ordering. Our result extend, in particular, those of Robertson and Wright [16], El Barmi and Dykstra

[7], Wang [23] and Liu and Wang [15]. We have also provided examples to illustrate our theoretical results.

Acknowledgements

The authors would like to thank the editor and a referee for helpful comments and suggestions that led to a much improved paper. The authors are especially grateful to the referee for his painstaking scrutiny that found several errors and many typos.

Appendix

In this section we give the proof of the main results. The following identities which follow from the definition of $P^0(P^0(\pi))$, $Q^0(Q^0(\pi))$ and $R^0(R^0(\pi))$ will be used in the proofs that follow.

(a)	$[\tilde{B}^0]^{-1}P^0 - HQ^0 = I,$	(b)	$-H^TP = 0,$
(d)	$[\tilde{B}^0]^{-1}Q^0 - HR^0 = 0,$	(e)	$-H^TQ^0 = I,$
(f)	$[\tilde{B}^0]^{-1}P^0(\pi) - HQ^0(\pi) = I,$	(g)	$-H^T(\pi)P(\pi) = 0,$
(h)	$[\tilde{B}^0]^{-1}Q^0(\pi) - H(\pi)R^0(\pi) = 0,$	(e)	$-H^T(\pi)Q^0(\pi) = I.$

The following lemmas will be used in the proof of Theorem 3.2.

Lemma 6.1. *The following identities hold*

$$[\tilde{B}^0]^{-1/2}(P^0(\pi) - P^0)[\tilde{B}^0]^{-1}Q^0(\pi)R^0(\pi)Q^{0T}(\pi)[\tilde{B}^0]^{-1/2} = 0,$$

$$[\tilde{B}^0]^{-1/2}(P^0(\pi) - P^0)[\tilde{B}^0]^{-1}Q^0(\pi) = 0,$$

$$[\tilde{B}^0]^{-1/2}(P^0(\pi) - P^0)[\tilde{B}^0]^{-1}P^0(\pi)H(\tilde{\pi}) = [\tilde{B}^0]^{-1/2}P^0(\pi)H(\tilde{\pi}),$$

$$[\tilde{B}^0]^{-1/2}Q^0(\pi)R^0(\pi)Q^{0T}(\pi)[\tilde{B}^0]^{-1}Q^0(\pi) = [\tilde{B}^0]^{-1/2}Q^0(\pi),$$

$$[\tilde{B}^0]^{-1/2}Q^0(\pi)R^0(\pi)Q^{0T}(\pi)[\tilde{B}^0]^{-1}P^0(\pi)H(\tilde{\pi}) = 0.$$

Proof. Follows immediately from the identities above. □

The proof of the following lemma can be found in [8].

Lemma 6.2. *Suppose* $\mathbf{X}$ *has a multivariate normal distribution with zero mean vector and covariance matrix I, P is an idempotent symmetric matrix of rank r and* $\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_k$ *are k vectors satisfying either* $P\mathbf{d}_i = \mathbf{0}$ *or* $P\mathbf{d}_i = \mathbf{d}_i$ *for all i. Then the conditional distribution of* $\mathbf{X}^TP\mathbf{X}$ *given* $\mathbf{d}^T\mathbf{X} \leqslant 0$, $i = 1, 2, \ldots, k$, *is that of a chi-squared random variable with r degrees of freedom.*

Proof of Theorem 3.2. We show that the likelihood ratio test statistic in this case has, asymptotically, a chi-square distribution with *c-card*(π) degrees of freedom. Without loss of generality, assume that $card(\pi) = c_1$ and that $H(\pi)$, the sub-matrix of H that corresponds to the constraints in $\mathcal{H}_{1:\pi}$, is made of the first c_1 columns of H. Let $P(\pi)$ and $Q(\pi)$ be the values P and Q in (3.8) for matrix $H(\pi)$ (instead of H).

Let $T_{01:\pi}$ denote the likelihood ratio test statistic for testing $\mathcal{H}_0$ against $\mathcal{H}_0 - \mathcal{H}_1(\pi)$. Next, we show that its limiting distribution is a chi-square with $c - c_1$ degrees of freedom. Let $\hat{\overline{\mathbf{p}}}$ denote the unrestricted maximum likelihood estimator of $\overline{\mathbf{p}}$. Since

$$\sqrt{n}[\hat{\overline{\mathbf{p}}} - \overline{\mathbf{p}}^{(0)}] = \frac{1}{\sqrt{n}} \tilde{B}^0 \mathcal{D} \ln \mathcal{L}(\overline{\mathbf{p}}^0) + o_p(1), \tag{6.15}$$

(3.8) and (3.11) imply that

$$\sqrt{n}[\hat{\overline{\overline{\mathbf{p}}}}^{(0)} - \hat{\overline{\overline{\mathbf{p}}}}] = \frac{1}{\sqrt{n}}[P^0 - \tilde{B}^0]\mathcal{D} \ln \mathcal{L}(\overline{\mathbf{p}}^0) + o_p(1) \tag{6.16}$$

and

$$\sqrt{n}[\hat{\mathbf{p}}(\pi) - \hat{\mathbf{p}}] = \frac{1}{\sqrt{n}}[P(\pi) - \tilde{B}^0]\mathcal{D} \ln \mathcal{L}(\overline{\mathbf{p}}^0) + o_p(1). \tag{6.17}$$

Applying a Taylor's expansion of $\ln \mathcal{L}(\hat{\overline{\mathbf{p}}}^{(0)})$ and $\ln \mathcal{L}(\hat{\overline{\mathbf{p}}}^{(0)}(\pi))$ around $\hat{\overline{\mathbf{p}}}$ under $\mathcal{H}_0$ we find that

$$\begin{aligned} \ln \mathcal{L}(\hat{\overline{\mathbf{p}}}^{(0)}) &= \ln \mathcal{L}(\hat{\overline{\mathbf{p}}}) - \frac{1}{2} n(\hat{\overline{\mathbf{p}}}^{(0)} - \hat{\overline{\mathbf{p}}})^T [\tilde{B}^0]^{-1} (\hat{\overline{\mathbf{p}}}^{(0)} - \hat{\overline{\mathbf{p}}}) + o_p(1), \\ \ln \mathcal{L}(\hat{\overline{\mathbf{p}}}(\pi)) &= \ln \mathcal{L}(\hat{\overline{\mathbf{p}}}) - \frac{1}{2} n(\hat{\overline{\mathbf{p}}}(\pi) - \hat{\overline{\mathbf{p}}})^T [\tilde{B}^0]^{-1} (\hat{\overline{\mathbf{p}}}(\pi) - \hat{\overline{\mathbf{p}}}) + o_p(1). \end{aligned}$$

Therefore, we have

$$\begin{aligned} T_{01:\pi} &= -2\left(\mathcal{D} \ln \mathcal{L}(\hat{\overline{\mathbf{p}}}^{(0)}) - \mathcal{D} \ln \mathcal{L}(\hat{\overline{\mathbf{p}}}(\pi))\right) \\ &= n(\hat{\mathbf{p}}(\pi) - \hat{\mathbf{p}})^T [\tilde{B}^0]^{-1} (\hat{\mathbf{p}}(\pi) - \hat{\mathbf{p}}) - n(\hat{\mathbf{p}}(\pi) - \hat{\mathbf{p}})^T [\tilde{B}^0]^{-1} (\hat{\mathbf{p}}(\pi) - \hat{\mathbf{p}}) + o_p(1) \\ &= [n^{-1/2} \mathcal{D} \ln \mathcal{L}(\overline{\mathbf{p}}^0)]^T (P^0 - \tilde{B}^0)[\tilde{B}^0]^{-1} (P^0 - \tilde{B}^0)[n^{-1/2} \mathcal{D} \ln \mathcal{L}(\overline{\mathbf{p}}^0)] \\ &\quad -[n^{-1/2} \mathcal{D} \ln \mathcal{L}(\overline{\mathbf{p}}^0)]^T (P^0(\pi) - \tilde{B}^0)[\tilde{B}^0]^{-1} (P^0(\pi) - \tilde{B}^0) \\ &\quad \times [n^{-1/2} \mathcal{D} \ln \mathcal{L}(\overline{\mathbf{p}}^0)] + o_p(1) \\ &= [n^{-1/2} \mathcal{D} \ln \mathcal{L}(\overline{\mathbf{p}}^0)]^T (P^0(\pi) - P)[n^{-1/2} \mathcal{D} \ln \mathcal{L}(\overline{\mathbf{p}}^0)] + o_p(1), \end{aligned} \tag{6.18}$$

where the last equality follows from the identities above.

Since $n^{-1/2}\mathcal{D} \ln \mathcal{L}(\overline{\mathbf{p}}^0)$ converges in distribution as n goes to infinity to $N(\mathbf{0}, [\tilde{B}^0]^{-1})$ and since

$$(P^0(\pi) - P^0)[\tilde{B}^0]^{-1}(P^0(\pi) - P^0)[\tilde{B}^0]^{-1} = (P^0(\pi) - P^0)[\tilde{B}^0]^{-1},$$

$T_{01:\pi}$ converges in distribution to a chi-square random variable with $rank(P^0(\pi) - P^0)$ degrees of freedom. But

$$\begin{aligned} rank(P^0(\pi) - P^0) &= rank\left((P^0(\pi) - P^0)[\tilde{B}^0]^{-1}\right) \\ &= trace\left((P^0(\pi) - P^0)[\tilde{B}^0]^{-1}\right) \end{aligned}$$

since $(P^0(\pi) - P^0)[\tilde{B}^0]^{-1}$ is idempotent. Since

$$trace\left((P^0(\pi) - P^0)[\tilde{B}^0]^{-1}\right) = trace(P^0(\pi)[\tilde{B}^0]^{-1}) - trace(P^0[\tilde{B}^0]^{-1}) = c - c_1,$$

we have the desired conclusion. □

Proof of Theorem 3.3. For any observed data set, $\hat{\mathbf{p}}^{(1)}$ equals $\hat{\mathbf{p}}(\pi)$ for exactly one π. Moreover $\hat{\mathbf{p}}^{(1)} = \hat{\mathbf{p}}(\pi)$ if and only if

$$\sum_{i=1}^{r}\sum_{j=1}^{k} x_{ij}^{(s)} \hat{p}_{ij}(\pi) < 0 \quad \forall s \in \tilde{\pi}, \tag{6.19}$$

$$\hat{\lambda}_s(\pi) > 0 \quad \forall s \in \pi. \tag{6.20}$$

Let $T_{12:\pi}$ be the test likelihood ratio test statistic for testing $\mathcal{H}_{1:\pi}$ against $\mathcal{H}_2 - \mathcal{H}_1$. It follows from [22] that

$$T_{12:\pi} = -\frac{1}{n}[\mathcal{D}\ln\mathcal{L}(\mathbf{p}^0)]^T Q^0(\pi)R^0(\pi)Q^0(\pi)\mathcal{D}\ln\mathcal{L}(\mathbf{p}^0) + o_p(1). \tag{6.21}$$

Using (3.11) and (3.12), (6.21) we have

$$\begin{aligned}
&P(T_{01} \geqslant t_1, T_{12} \geqslant t_2)\\
&\quad = \sum_{\pi\in\mathcal{F}} P(T_{01} \geqslant t_1, T_{12} \geqslant t_2, \hat{\mathbf{p}}^{(1)} = \hat{\mathbf{p}}(\pi))\\
&\quad = \sum_{\pi\in\mathcal{F}} P\left(T_{01} \geqslant t_1, T_{12} \geqslant t_2, \sum_{i=1}^{r}\sum_{j=1}^{k} x_{ij}^{(s)} \hat{p}_{ij}(\pi) < 0\,\forall s \in \tilde{\pi}, \hat{\lambda}_s(\pi) > 0\,\forall s \in \pi\right)\\
&\quad = \sum_{\pi\in\mathcal{F}} P\left(T_{01:\pi} \geqslant t_1, T_{12:\pi} \geqslant t_2, \sum_{i=1}^{r}\sum_{j=1}^{k} x_{ij}^{(s)} \hat{p}_{ij}(\pi) < 0\,\forall s \in \tilde{\pi}, \hat{\lambda}_s(\pi) > 0\,\forall s \in \pi\right)\\
&\quad = \sum_{\pi\in\mathcal{F}} P([n^{-1/2}\mathcal{D}\ln\mathcal{L}(\overline{\mathbf{p}}^0)]^T(P^0(\pi) - P^0)[n^{-1/2}\mathcal{D}\ln\mathcal{L}(\overline{\mathbf{p}}^0)] + o_p(1) \geqslant t_1,\\
&\qquad -\frac{1}{n}[\mathcal{D}\ln\mathcal{L}(\overline{\mathbf{p}}^0)]^T Q^0(\pi)R^0(\pi)Q^{0T}(\pi)\mathcal{D}\ln\mathcal{L}(\overline{\mathbf{p}}^0) + o_p(1) \geqslant t_2\\
&\qquad \times \sum_{i=1}^{r}\sum_{j=1}^{k} x_{ij}^{(s)} \hat{p}_{ij}(\pi) < 0\,\forall s \in \tilde{\pi}, \hat{\lambda}_s(\pi) > 0\,\forall s \in \pi)\\
&= \sum_{\pi\in\mathcal{F}} P([n^{-1/2}\mathcal{D}\ln\mathcal{L}(\overline{\mathbf{p}}^0)]^T(P^0(\pi) - P^0)[n^{-1/2}\mathcal{D}\ln\mathcal{L}(\overline{\mathbf{p}}^0)] + o_p(1) \geqslant t_1,\\
&\qquad -\frac{1}{n}[\mathcal{D}\ln\mathcal{L}(\overline{\mathbf{p}}^0)]^T Q^0(\pi)R^0(\pi)Q^{0T}(\pi)\mathcal{D}\ln\mathcal{L}(\overline{\mathbf{p}}^0) + o_p(1) \geqslant t_2,\\
&\qquad \times n^{-1/2}H^T(\tilde{\pi})P^0(\pi)\mathcal{D}\ln\mathcal{L}(\overline{\mathbf{p}}^0) + o_p(1) < 0,\\
&\qquad \times n^{-1/2}Q^{0T}(\pi)\mathcal{D}\ln\mathcal{L}(\overline{\mathbf{p}}^0) + o_p(1) > 0).
\end{aligned}$$

Since $n^{-1/2}\mathcal{D}\ln\mathcal{L}(\overline{\mathbf{p}}^0)$ converges in distribution as n goes to infinity to $N(\mathbf{0},[\tilde{B}^0]^{-1})$,

$$\begin{aligned}
&\lim_{n\to\infty} P(T_{01}\geqslant t_1, T_{12}\geqslant t_2)\\
&\quad=\sum_{\pi\in\mathcal{F}} P(\mathbf{U}^T(P^0(\pi)-P^0)\mathbf{U}\geqslant t_1, \mathbf{U}^T Q^0(\pi)R^0(\pi)Q^{0T}(\pi)\mathbf{U}\geqslant t_2,\\
&\qquad H^T(\tilde{\pi})P^0(\pi)\mathbf{U}<\mathbf{0}, Q^{0T}(\pi)\mathbf{U}\geqslant\mathbf{0}),
\end{aligned}$$

where $\mathbf{U}$ has a multivariate normal distribution with zero mean vector and variance $[\tilde{B}^0]^{-1}$. Let $\mathbf{Z}$ have a multivariate normal distribution with zero mean vector and variance equal to the identity matrix I, then

$$\begin{aligned}
&\lim_{n\to\infty} P(T_{01}\geqslant t_1, T_{12}\geqslant t_2)\\
&\quad=\sum_{\pi\in\mathcal{F}} P(\mathbf{U}^T(P^0(\pi)-P^0)\mathbf{U}\geqslant t_1, \mathbf{U}^T Q^0(\pi)R^0(\pi)Q^{0T}(\pi)\mathbf{U}\geqslant t_2,\\
&\qquad H^T(\tilde{\pi})P^0(\pi)\mathbf{U}<\mathbf{0}, Q^{0T}(\pi)\mathbf{U}\geqslant\mathbf{0})\\
&\quad=\sum_{\pi\in\mathcal{F}} P(\mathbf{Z}^T[\tilde{B}^0]^{-1/2}(P^0(\pi)-P^0)[\tilde{B}^0]^{-1/2}\mathbf{Z}\geqslant t_1,\\
&\qquad \mathbf{Z}^T[\tilde{B}^0]^{-1/2}Q^0(\pi)R^0(\pi)Q^{0T}(\pi)[\tilde{B}^0]^{-1/2}\mathbf{Z}\geqslant t_2,\\
&\qquad H^T(\tilde{\pi})P^0(\pi)[\tilde{B}^0]^{-1/2}\mathbf{Z}<\mathbf{0}, Q^{0T}(\pi)[\tilde{B}^0]^{-1/2}\mathbf{Z}\geqslant\mathbf{0}).
\end{aligned}$$

Using Lemmas 6.1, 6.2 and Lemma D (Robertson et al. [18], page 71), we get

$$\begin{aligned}
&\sum_{\pi\in\mathcal{F}} P(\mathbf{Z}^T[\tilde{B}^0]^{-1/2}(P^0(\pi)-P^0)[\tilde{B}^0]^{-1/2}\mathbf{Z}\geqslant t,\\
&\quad Q^{0T}(\pi)[\tilde{B}^0]^{-1/2}\mathbf{Z}\geqslant\mathbf{0}, H^T(\pi)P^0(\pi)[\tilde{B}^0]^{-1/2}\mathbf{Z}<\mathbf{0})\\
&\quad=\sum_{\pi\in\mathcal{F}} P(\mathbf{Z}^T[\tilde{B}^0]^{-1/2}(P^0(\pi)-P^0)[\tilde{B}^0]^{-1/2}\mathbf{Z}\geqslant t_1)\\
&\qquad\times P(\mathbf{Z}^T[\tilde{B}^0]^{-1/2}Q^0(\pi)R^0(\pi)Q^{0T}(\pi)[\tilde{B}^0]^{-1/2}\mathbf{Z}\geqslant t_2)\\
&\qquad\times P(Q^{0T}(\pi)[\tilde{B}^0]^{-1/2}\mathbf{Z}\geqslant\mathbf{0})P(H^T(\pi)P^0(\pi)[\tilde{B}^0]^{-1/2}\mathbf{Z}<\mathbf{0}).
\end{aligned}$$

Note that by Theorem 3.2,

$$P(\mathbf{Z}^T[\tilde{B}^0]^{-1/2}(P^0(\pi)-P^0)[\tilde{B}^0]^{-1/2}\mathbf{Z}\geqslant t_1)=P(\chi^2_{c-card(\pi)}\geqslant t_1).$$

Since $[\tilde{B}^0]^{-1/2}Q^0(\pi)R^0(\pi)Q^{0T}(\pi)[\tilde{B}^0]^{-1/2}$ is idempotent with rank equal to cardinal of π, we have

$$P(\mathbf{Z}^T[\tilde{B}^0]^{-1/2}Q^0(\pi)R^0(\pi)Q^{0T}(\pi)[\tilde{B}^0]^{-1/2}\mathbf{Z}\geqslant t_2)=P(\chi^2_{card(\pi)}\geqslant t_2).$$

Therefore

$$\begin{aligned}
\lim_{n\to\infty} P(T_{01}\geqslant t_1, T_{12}\geqslant t_2) &= \sum_{j=0}^{c} P(\chi^2_{c-j}\geqslant t_1)P(\chi^2_j\geqslant t_2)\\
&\quad\times\sum_{\pi\in\mathcal{F},card(\pi)=j} P(Q^{0T}(\pi)[\tilde{B}^0]^{-1/2}\mathbf{Z}\geqslant\mathbf{0})\\
&\quad\times P(H^T(\pi)P^0(\pi)[\tilde{B}^0]^{-1/2}\mathbf{Z}<\mathbf{0}).
\end{aligned}$$

Since

$$Q^{0T}(\pi)[\tilde{B}^0]^{-1}Q^0(\pi) = -R^0(\pi) = [H^T(\pi)\tilde{B}^0 H(\pi)]^{-1}$$

$$\begin{aligned} P(Q^{0T}(\pi)[\tilde{B}^0]^{-1/2}Z > 0) &= P(N(\mathbf{0}, Q^{0T}(\pi)[\tilde{B}^0]^{-1}Q^0(\pi)\geqslant 0) \\ &= P(N(\mathbf{0}, [H^T(\pi)\tilde{B}^0 H(\pi)]^{-1})\geqslant 0). \end{aligned}$$

Also

$$P(H^T(\tilde{\pi})P^0(\pi)[\tilde{B}^0]^{-1/2}Z < \mathbf{0}) = P(N(\mathbf{0}, H^T(\tilde{\pi})P^0(\pi)H(\tilde{\pi}) < \mathbf{0})$$

so that

$$\sum_{\pi\in\mathcal{F}, card(\pi)=j} P(Q^{0T}(\pi)[\tilde{B}^0]^{-1/2}\mathbf{Z}\geqslant\mathbf{0})P(H^T(\pi)P^0(\pi)[\tilde{B}^0]^{-1/2}\mathbf{Z} < \mathbf{0}) = a_j(\mathbf{p}^0).$$

Putting all this together gives the desired conclusion. □

References

[1] A. Agresti, B. Coull, The analysis of contingency tables under inequality constraints, J. Statist. Plann. Inference 107 (1–2) (2002) 45–73.
[2] J. Aitchison, S.D. Silvey, Maximum likelihood estimation of parameters subject to restraints, Ann. Math. Statist. (1957) 130–140.
[3] R.L. Dykstra, C.J. Feltz, Nonparametric maximum likelihood estimation of the survival functions with a general stochastic ordering and its dual, Biometrika 76 (1989) 331–341.
[4] R.L. Dykstra, S.C. Kochar, T. Robertson, Statistical inference for uniform stochastic ordering in several populations, Ann. Statist. 19 (1991) 870–888.
[5] H. El Barmi, R. Dykstra, Restricted multinomial MLE based upon Fenchel duality, Statist. Probab. Lett. 21 (1994) 121–130.
[6] H. El Barmi, R. Dykstra, Testing for or against a set of linear inequality constraints in a multinomial setting, Canad. J. Statist. 23 (1995) 131–143.
[7] H. El Barmi, R. Dykstra, Restricted product multinomial and product Poisson estimation based upon Fenchel duality, Statist. Probab. Lett. 29 (1996) 117–123.
[8] H. El Barmi, R. Dykstra, Likelihood ratio test against a set of inequality constraints, J. Nonparametric Statist. 11 (1999) 233–250.
[9] C. Feltz, R. Dykstra, Maximum likelihood estimation of the survival functions of N stochastically ordered random variables, J. Amer. Statist. Assoc. 80 (1985) 1012–1019.
[10] A. Genz, Numerical computation of multivariate normal probabilities, J. Comput. Graph. Statist. 1 (1992) 141–149.
[11] A. Genz, Comparison of methods for computation of multivariate normal probabilities, Comput. Sci. Statist. 25 (1993) 400–405.
[12] J.D. Kalbfleisch, R.L. Prentice, The Statistical Analysis of Failure Time Data, Wiley, New York, 1980.
[13] A. Kudô, A multivariate analogue of the one sided test, Biometrika 50 (1963) 403–418.
[14] H. Levy, Stochastic dominance and expected utility: survey and analysis, Manage. Sci. 38 (1992) 555–593.
[15] X. Liu, J. Wang, Testing the equality of multinomial populations ordered by increasing convexity under the alternative, Canad. J. Statist. 32 (2004) 159–168.
[16] T. Robertson, F.T. Wright, Likelihood ratio test for and against stochastic ordering between multinomial populations, Ann. Statist. 6 (1981) 1248–1257.
[17] T. Robertson, F.T. Wright, On approximation of the level probabilities and associated distributions in order restricted inference, Biometrika 70 (1983) 597–606.
[18] T. Robertson, F.T. Wright, R.L. Dykstra, Order Restricted Statistical Inference, Wiley, New York, 1988.

[19] S. Ross, Stochastic Processes, Wiley, New York, 1996.
[20] R. Serfling, Approximation Theorems of Mathematical Statistics, Wiley, New York, 1980.
[21] M.J. Silvapulle, P.K. Sen, Constrained Statistical Inference, Wiley, New York, 2005.
[22] S.D. Silvey, The Lagrangian multiplier test, Ann. Statist. 30 (1959) 389–407.
[23] Y. Wang, A likelihood ratio test against stochastic ordering in several populations, J. Amer. Statist. Assoc. 91 (1996) 1676–1683.