A Likelihood Ratio Test Against Stochastic Ordering in Several Populations

Yazhen Wang [*]

JASA 91 (1996), pp. 1676-1683

Abstract

The likelihood ratio test is often used to test hypotheses involving a stochastic ordering. Distribution theory for the likelihood ratio test has been developed only for two stochastically ordered distributions. For testing equality of distributions against a stochastic ordering in several populations, this paper derives the null asymptotic distribution of the likelihood ratio test statistic, which is characterized by minimization problems and has no closed form. A Monte Carlo simulation is conducted to study the limiting distribution. Since the limiting distribution depends on the specific values of the unknown distributions under the null hypothesis, asymptotic and bootstrap approaches are proposed to overcome practical difficulties and implement tests based on the likelihood principle. Asymptotic validities for these tests are established and simulations are carried out to check their performances for finite

[*]Yazhen Wang is Assistant Professor of Statistics, University of Missouri, Columbia, MO 65211.
Research is supported in part by the Research Board Grant from University of Missouri and the Grant DMS-9404142 from National Science Foundation. The author is grateful to Professor Tim Wright for helpful discussion and comments and Professor Frederick Williams for correction of the language. Special thanks are due to two anonymous referees, an Associate Editor and the Editor for their stimulating comments and regarding the simulation of the limiting distribution in particular.

sample sizes. The tests are applied to an example involving data for survival time for carcinoma of the oropharynx.

KEY WORDS: Asymptotic test; Bootstrap test; Limiting distribution.

1. INTRODUCTION

Stochastic ordering between probability distributions has been widely studied since it was introduced by Lehmann in 1955. It arises in many situations in which it is believed that distributions must be stochastically ordered. Since it is often easy to make value judgements when such orderings exist, and incorporating these orderings into an inference increases its statistical efficiency, it is desirable to recognize their occurrence and to select models with such orderings. The likelihood ratio test is often used to test hypotheses involving stochastic orderings in several populations. In this article we derive the null asymptotic distribution of the likelihood ratio test statistic for testing equality of distributions against a stochastic ordering.

Closed form expressions for maximum likelihood estimates for more than two stochastically ordered distributions have not been found. Distribution theory for tests based on the likelihood principle has not been fully developed to the best of our knowledge. Brunk, Franck, Hanson and Hogg (1966) obtain a closed form for nonparametric maximum likelihood estimates of two stochastically ordered distributions and study their properties. Dykstra (1982) considers a similar problem with censored data and gives the maximum likelihood estimates in the form of Kaplan-Meier product limit estimators with modified data. For more than two stochastically ordered distributions, even though the solution has a nice characterization, obtaining the actual estimates is difficult. An iterative algorithm depending only on the solution to the pairwise problem is proposed by Dykstra and Feltz (1989) and Feltz and Dykstra (1985). Lee (1987), Robertson and Wright (1974), and Sampson and Whitaker (1989) study estimation of multivariate distributions under a stochastic ordering.

Testing procedures which are based on maximum likelihood estimates of two stochastically ordered distributions are discussed in Robertson and Wright (1981), Lucas and Wright (1991), Lee and Wolfe (1976), Franck (1989) and Dykstra, Madsen and Fairbanks (1983). Robertson and Wright (1981) and Lucas and Wright (1991) derive asymptotic distributions for likelihood ratio test statistics for stochastic ordering in discrete univariate and bivariate case. Their asymptotic distributions are chi-bar-squared distributions, a mixture of chi-squared distributions (Robertson, Wright and Dykstra 1988, chap. 2). Lee and Wolfe (1976) propose a Mann - Whitney - Wilcoxon test for stochastic ordering based on the restricted maximum likelihood estimates. Dykstra et al. (1983) consider a nonparametric likelihood ratio test for stochastic ordering based on the restricted maximum likelihood estimates. Franck (1984) discusses a similar test based on the restricted maximum likelihood estimates. None of the last three references derive the asymptotic distributions of their test statistics. Mau (1988) presents a test based on isotonic regression of sample means for censored data. Dykstra, Kochar, and Robertson (1991) study estimation under uniform stochastic ordering and testing against a uniform stochastic ordering.

Our approach to deriving the asymptotic distribution of the likelihood ratio test statistic is as follows. The likelihood ratio test statistic is a function of the restricted maximum likelihood estimates, so the null asymptotic distribution of the likelihood ratio test statistic is easily obtained from those of the restricted maximum likelihood estimates. The restricted maximum likelihood estimates are characterized by a maximization problem with a concave objective function and linear constraints. By empirical process arguments (Kim and Pollard 1990), we first show that the restricted maximum likelihood estimates are $n^{1/2}$-consistent and then restrict the objective function in a $n^{-1/2}$-shrinking neighborhood of the common distribution function under the null hypothesis and think of it as an empirical process indexed by distribution functions. The process converges weakly to a Gaussian process and hence the limiting distribution is derived by the continuous mapping theorem.

The rest of this paper is organized as follows. The asymptotic distribution for the likelihood ratio test statistic is given in Section 2. The limiting distribution depends on the specific values of the distributions under the null hypothesis. Unlike the case with two stochastically ordered distributions, for more than two stochastically ordered distributions no closed form is available for the asymptotic distribution. This section also conducts a Monte Carlo simulation to study the limiting distribution. Section 3 proposes an asymptotic test and a bootstrap test to overcome practical difficulties and implement tests based on the likelihood principle. These approaches are parameter free and use computer-based simulations to find critical values of the tests. Asymptotic validities for these tests are established and simulations are carried out to check their performances for finite sample sizes. Section 4 illustrates an application of the methods to an example involving data for survival time for carcinoma of the oropharynx. Section 5 features some concluding remarks. All proofs are included in the Appendix.

2. DISTRIBUTION THEORY FOR THE LIKELIHOOD RATIO TEST

2.1 The Likelihood Ratio Test

Suppose k populations have distribution functions $F_1, \ldots, F_k$. We say $F_1, \ldots, F_k$ are stochastically ordered, denoted by $F_1 \leq \cdots \leq F_k$, if $F_1(x) \leq \cdots \leq F_k(x)$ for all x. Let $H_0 : F_1 = \cdots = F_k$ and $H_1 : F_1 \leq \cdots \leq F_k$. We consider testing H_0 against $H_1 - H_0$.

Suppose we have interval censored data, which are common situations in medical research. That is, with $-\infty = s_0 < s_1 < \cdots < s_m \leq \infty$ checking times, an experiment is checked at the times $s_1, \ldots, s_m$ and only the number of failures (or deaths) between the checking times are observed. Let d_{ij} be the number of observations from the ith population in $(s_{j-1}, s_j]$, $n_i = \sum_{j=1}^{m} d_{ij}$ the total number of observations from the ith population, and $n = \sum_{i=1}^{k} n_i$ the total number of observations from all k populations (Dykstra and Feltz 1989; Feltz and Dykstra 1985).

Let $\mathbf{F} = (F_1, \ldots, F_k)$. The likelihood function for $\mathbf{F}$ is

$$L(\mathbf{F}) = \prod_{i=1}^{k} \prod_{j=1}^{m} \{F_i(s_j) - F_i(s_{j-1})\}^{d_{ij}}. \tag{1}$$

The maximum likelihood estimate, $\tilde{\mathbf{F}} = (\tilde{F}_1, \ldots, \tilde{F}_k)$, of $\mathbf{F}$ subject to the stochastic ordering is defined to be

$$\begin{aligned} \tilde{\mathbf{F}} = Argmax\{&L(\mathbf{F}) : \mathbf{F} = (F_1, \ldots, F_k), \\ &F_1 \leq \cdots \leq F_k\}. \end{aligned} \tag{2}$$

For $k = 2$, $(\tilde{F}_1, \tilde{F}_2)$ has a closed form (Brunk et al. 1966), and for $k \geq 3$, no closed form is available for $(\tilde{F}_1, \ldots, \tilde{F}_k)$ (Dykstra and Feltz 1989; Feltz and Dykstra 1985). An iterative algorithm is proposed to compute the estimate by Dykstra and Feltz (1989) and Feltz and Dykstra (1985).

Under H_0, $F_1, \ldots, F_k$ are equal, we denote by F_0 the common distribution function, and let $p_j = F_0(s_j) - F_0(s_{j-1})$, $\mathbf{p} = (p_1, \ldots, p_m)$, and $\mathbf{F}_0 = (F_0, \ldots, F_0)$. Denote by

$$\tilde{F}_0(s_j) = \frac{1}{n} \sum_{i=1}^{k} \sum_{r=1}^{j} d_{ir}, \quad j = 1, \ldots, m, \tag{3}$$

the empirical distribution for the pooled sample, and set $\tilde{\mathbf{F}}_0 = (\tilde{F}_0, \ldots, \tilde{F}_0)$. Then $\tilde{\mathbf{F}}_0$ is the maximum likelihood estimate of $\mathbf{F}$ under H_0.

The likelihood ratio test statistic is

$$S_n = 2\{logL(\tilde{\mathbf{F}}) - logL(\tilde{\mathbf{F}}_0)\}, \tag{4}$$

and the likelihood ratio test will reject H_0 if $S_n \geq C$, where the critical value C satisfies

$$P_{H_0}(S_n \geq C) = \alpha. \tag{5}$$

2.2 The Limiting Distribution

In order to find the critical value C defined in (5), we need to know either the exact null distribution or the null asymptotic distribution of S_n. The exact distribution of

S_n is very difficult to characterize. We give its limiting distribution in the following theorem.

Theorem 1 *Assume* $p_j > 0$, $j = 1, \ldots, m$, *and* $n_i/n \to a_i$, *as* $n \to \infty$, $0 < a_i < 1$, $i = 1, \ldots, k$, $\sum_{i=1}^{k} a_i = 1$. *Then under* H_0, S_n *converges weakly to the random variable*

$$\min\{V(\mathbf{t}) : \mathbf{t} \in \mathbf{T}_0\} - min\{V(\mathbf{t}) : \mathbf{t} \in \mathbf{T}\},$$

where

$$\begin{aligned} \mathbf{T} = \{\mathbf{t} = \{t_{ij}\} \in \mathbf{R}^{mk-k} : \sum_{r=1}^{j} t_{\ell r} \geq \sum_{r=1}^{j} t_{\ell+1,r}, \\ j = 1, \ldots, m-1, \ell = 1, \ldots, k-1\}, \end{aligned} \tag{6}$$

$$\begin{aligned} \mathbf{T}_0 = \{\mathbf{t} = (t_{ij}) \in \mathbf{T} : t_{1j} = \cdots = t_{kj}, \\ j = 1, \ldots, m-1\}, \end{aligned} \tag{7}$$

$$V(\mathbf{t}) = \sum_{i=1}^{k} \sum_{j=1}^{m-1} A_{ij}\{t_{ij} - B_{ij}\}^2, \tag{8}$$

for $1 \leq i \leq k$ *and* $1 \leq j \leq m-1$,

$$A_{ij} = \frac{a_i \sum_{r=j}^{m} p_r \sum_{r=j+1}^{m} p_r}{p_j},$$

$$B_{ij} = \frac{a_i^{1/2}}{2A_{ij}} \left(\sum_{r=j+1}^{m} U_{ir} - \frac{\sum_{r=j+1}^{m} p_r}{p_j} U_{ij} \right),$$

$(U_{i1}, \ldots, U_{im})$, $i = 1, \ldots, k$, *are independent and identically distributed, and* $(U_{i1}, \ldots, U_{im})$ *follows a multivariate normal distribution with mean zero and covariance* $Var(U_{ij}) = p_j\,(1 - p_j)$, $Cov(U_{ij}, U_{ir}) = -p_j\,p_r$ *for* $j \neq r$.

Denote by $G(x; k, m, \mathbf{p})$ the limiting distribution function of S_n under H_0. For $k = 2$, $G(x; 2, m, \mathbf{p})$ is a chi-bar-squared distribution with level probability $P(j, m, \mathbf{p})$ (Robertson and Wright 1981; Robertson et al. 1988, sec. 5.4). The chi-bar-squared distribution can be heuristically derived by Theorem 1 as follows.

For $k = 2$, the cone $\mathbf{T}$ consists of all $\mathbf{t} = (t_{ij}) \in \mathbf{R}^{2m-2}$ satisfying the following simple order on cumulative sums,

$$\begin{aligned} t_{11} &\geq t_{21}, \\ t_{11} + t_{12} &\geq t_{21} + t_{22}, \\ t_{11} + t_{12} + t_{13} &\geq t_{21} + t_{22} + t_{23}, \\ \cdots & \quad \cdots \\ t_{11} + \cdots + t_{1,m-1} &\geq t_{21} + \cdots + t_{2,m-1}. \end{aligned}$$

Because $\mathbf{T}$ is defined by the simple order, and $V(\mathbf{t})$ is the sum of squared differences between t_{ij} and normal random variables, minimizing $\mathbf{V}(\mathbf{t})$ over $\mathbf{T}$ can be converted into an isotonic regression problem. Therefore, the distribution of the minimum of $V(\mathbf{t})$ on $\mathbf{T}$ can be expected to be a chi-bar-squared distribution (Robertson et al. 1988, chap. 2).

2.3 Monte Carlo Simulations

For $k \geq 3$, no closed form for $G(x; k, m, \mathbf{p})$ is available. Simulations are carried out to study $G(x; k, m, \mathbf{p})$. For a given $\mathbf{p}$, generating k independent normal random vectors $(U_{i1}, \ldots, U_{im})$, $i = 1, \ldots, k$, and evaluating $V(\mathbf{t})$ according to the description in Theorem 1, we solve the minimization problems of finding the minima of $V(\mathbf{t})$ over $\mathbf{T}_0$ and over $\mathbf{T}$. A simulated observation from $G(x; k, m, \mathbf{p})$ is obtained by taking the difference of the two minima. Repeating the procedure we get a sample of simulated data from $G(x; k, m, \mathbf{p})$. $G(x; k, m, \mathbf{p})$ is estimated by the empirical distribution function of the simulated sample. In the simulation the empirical distribution functions are computed by 5000 simulated data. We plot the $G(x; k, m, \mathbf{p})$ for various values of k and $\mathbf{p}$ in Figures 1 and 2. In particular, Figure 1 illustrates that $G(x; k, m, \mathbf{p})$ varies substantially as the number, k, of populations increases.

Put Figures 1 and 2 about here

Simulations are conducted to check the approximation of the null distribution function of S_n for finite n by $G(x; k, m, \mathbf{p})$. For a given $\mathbf{p}$ we generate k independent multinomial random vectors $(d_{i1}, \ldots, d_{im})$, $i = 1, \ldots, k$, and then compute the value of S_n by (4). The distribution function of S_n is estimated by the empirical distribution function of 5000 simulated values of S_n.

The simulation results indicate that, when the sample sizes n_i are approximately equal, the distribution function of S_n converges to $G(x; k, m, \mathbf{p})$ very fast and is well approximated by $G(x; k, m, \mathbf{p})$ even for small and moderate sample sizes. Figure 3 compares $G(x; k, m, \mathbf{p})$ with the null distribution function of S_n with $n_1 = n_2 = 30$ for $k = 2$, $m = 3$, and $\mathbf{p} = (1/3, 1/3, 1/3)$. For this example the difference between the 95% quantiles of $G(x; k, m, \mathbf{p})$ and the null distribution of S_n is about 0.003.

Put Figure 3 about here

However, when the ratio of the maximum and the minimum of n_i is far from 1, the simulations reveal that the convergence becomes very slow. For example, Figure 4 displays $G(x; k, m, \mathbf{p})$ and the null distribution function of S_n with $(n_1, n_2) = (30, 10), (90, 30)$, and $(900, 300)$ for $k = 2$, $m = 3$ and $\mathbf{p} = (1/3, 1/3, 1/3)$. In the example we keep the ratio $n_1/n_2 = 3$. Note that the difference between $G(x; k, m, \mathbf{p})$ and the null distribution function of S_n is clearly visible even for $(n_1, n_2) = (900, 300)$. Comparing with the case of $n_1 = n_2 = 30$ in Figure 3, we see that the convergence is much slower.

Put Figure 4 about here

It is of interest to observe that, although the limiting distribution $G(x; 2, m, \mathbf{p})$ doesn't depend on the limit ratio of n_1 and n_2 (Robertson and Wright 1981; Robertson et al. 1988, pp. 254), the simulations suggest that the ratio does have a strong effect on the speed of the convergence. Simulation results also show that the null distribution function of S_n lies below $G(x; 2, m, \mathbf{p})$ if $n_1/n_2 > 1$ and above $G(x; 2, m, \mathbf{p})$ if $n_1/n_2 < 1$.

3. AN ASYMPTOTIC TEST AND A BOOTSTRAP TEST

Because the limiting distribution $G(x; k, m, \mathbf{p})$ depends on the unknown common distribution F_0 (or $\mathbf{p}$), to find the critical value defined by (5) and carry out the likelihood ratio test, the usual approach is to maximize $1 - G(x; k, m, \mathbf{p})$ over all $\mathbf{p}$ and use the critical value based on the maximum. For $k = 2$, the maximum is a mixture of two chi-squared distributions (Robertson and Wright 1981; Robertson et al. 1988, sec. 5.4). However, the maximum distributions for $k \geq 3$ are not available. Moreover, this approach yields a very conservative test for many practical situations (Robertson and Wright 1981; Lucas and Wright 1991).

3.1 An Asymptotic Test

One approach is to replace $\mathbf{p}$ by its estimate $\tilde{\mathbf{p}}$ and then find the critical value C_A by solving

$$G(C_A; k, m, \tilde{\mathbf{p}}) = 1 - \alpha.$$

The asymptotic test will reject H_0 if $S_n \geq C_A$.

Although no closed form is available for $G(x; k, m, \tilde{\mathbf{p}})$ if $k \geq 3$, following the description in Section 2.3 we simulate the distribution $G(x; k, m, \tilde{\mathbf{p}})$ and then easily find its quantile C_A.

3.2 A Bootstrap Test

Another practical approach is to use the bootstrap method. Having observed $\{d_{ij}\}$, we compute S_n by (4) and then bootstrap the likelihood ratio test statistic S_n^*. The critical value C_B for the bootstrap test is the $1 - \alpha$ quantile of the conditional distribution of S_n^* given $\{d_{ij}\}$, and the bootstrap test will reject H_0 if $S_n \geq C_B$.

In practice, C_B is approximated by the $1-\alpha$ quantile of the empirical distribution of J bootstrap replications of S_n for a large integer J (Beran and Ducharme 1991, pp.11; Efron and Tibshirani 1993, chap. 16).

3.3 Asymptotic Validities

An important question concerns how well these methods perform in constructing tests with the desired type I error. The following theorem shows both critical values, C_A and C_B, are asymptotically consistent, meaning that tests based on these critical values will have the desired type I error for large samples.

Theorem 2 *Under the assumption of Theorem 1, if $\alpha < 0.5$, then*

$$\lim_{n\to\infty} P_{H_0}(S_n \geq C_A) = \alpha,$$

and

$$\lim_{n\to\infty} P_{H_0}(S_n \geq C_B) = \alpha.$$

Since the limiting distribution of S_n depends on the unknown common distribution under H_0, S_n is not an asymptotically pivotal statistic and hence the two tests are asymptotically equivalent (Beran and Ducharme 1991; Hall 1992, chap. 3).

3.4 Simulation Results

A simulation study is carried out to check the performance of the asymptotic test and the bootstrap test for finite sample sizes. It shows that, if sample sizes n_i are approximately equal, both the asymptotic test and the bootstrap test perform very well even for small to moderate sample sizes. For example, for $k = 2, 4$ and 6, $m = 3$, $n_i = 10, 30$ and 90, and various $\mathbf{p}$, the asymptotic tests and bootstrap tests with $\alpha = 0.05$ have actual type I errors ranging from 0.046 to 0.054. In the simulations the critical values C_B are computed by the method described in Section 3.2 with $J = 5000$, and Type I errors are evaluated by 500 repetitions.

4. AN EXAMPLE

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 four distributions to be stochastically ordered. Restricted maximum likelihood estimates and testing against uniform stochastic ordering for the data have been studied by Dykstra and Feltz (1989), Feltz and Dykstra (1985), and Dykstra et al. (1991). The data were grouped into seven classes as indicated in Table 1 of Dykstra et al. (1991). Since the techniques developed apply to data without censoring, we delete all censored data and Group VII (in which most data are censored), and merge Groups V and VI into a single group. d_{ij} for the grouped samples are given in Table 1.

Table 1. No. of observations for the grouped data

Population	d_{i1}	d_{i2}	d_{i3}	d_{i4}	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

It may not be appropriate to use the method to test all four populations. For this moderate data set, the number of observations corresponding to Population 3 is three times larger than that corresponding to each of the other three populations, and hence the null distribution of the likelihood ratio test statistic may not be well approximated by the asymptotic distribution.

However, the numbers of observations corresponding to Populations 0, 1, and 2 are pretty close. For testing the three populations, the p value = 0.497. Such a big p value indicates that there is no statistical evidence that the distributions for the first three populations differ.

Treating the distributions of the first three populations the same, we pool their observations together. The number of observations corresponding to the three populations is approximately equal to the number of observations corresponding to Population 3. For testing the pooled population and Population 3, the p value = 0.091. The p value provides evidence that the distribution of Population 3 dominates the others.

5. CONCLUSIONS AND DISCUSSION

For testing equality of distributions against a stochastic ordering in several populations, the distribution theory for the likelihood ratio test has been developed. The limiting distribution of the likelihood ratio test statistic is characterized by minimization problems and has no closed form. A Monte Carlo simulation is conducted to study the limiting distribution. Because the limiting distribution depends on the specific values of the unknown distributions under the null hypothesis, asymptotic and bootstrap approaches are proposed to overcome practical difficulties and implement tests based on the likelihood principle. Asymptotic validities for these tests are established and simulations are carried out to check their performances for finite sample sizes. Finally the tests are applied to a real example.

The ideas for deriving the limiting distribution work for other ordering configurations and multivariate cases. For example, for a given ordering configuration

$$\Gamma = \{(a,b) : F_a \leq F_b, 1 \leq a, b \leq k\},$$

the limiting distribution of S_n will be the same as that described in Theorem 1 except for that the cone $\mathbf{T}$ is replaced by

$$\mathbf{T} = \{\mathbf{t} = \{t_{ij}\} \in \mathbf{R}^{mk-k} : \sum_{r=1}^{j} t_{ar} \geq \sum_{r=1}^{j} t_{br},$$
$$j = 1, \ldots, m-1, (a,b) \in \Gamma\}.$$

For data with censoring in the sense of Dykstra and Feltz (1989) and Feltz and Dykstra (1985), Praestgaard (1994) derives the limiting distributions of restricted maximum likelihood estimates. The results can be used to derive an asymptotic distribution of the likelihood ratio test statistic for censored data.

For testing for a stochastic ordering, that is, H_1 is a null hypothesis, it is easy to show that the asymptotic least favorable distribution is the limiting distribution of the likelihood ratio test statistic when the k distributions are equal (Robertson and Wright 1981; Robertson et al. 1988, sec. 5.4). By the same method, we can derive that the asymptotic least favorable distribution is the distribution of $min\{V(\mathbf{t}) : \mathbf{t} \in \mathbf{T}\}$, where $V(\mathbf{t})$ and $\mathbf{T}$ are defined in Theorem 1. A test based on the asymptotic least favorable distribution can be built. However, this test may be very conservative. The asymptotic and bootstrap approaches can be used to construct reasonable tests which are not conservative.

APPENDIX: PROOFS

Proof of Theorem 1

First we reparametrize $\mathbf{F}$ and $\mathbf{F}_0$. Set $\boldsymbol{\theta} = \{\theta_{ij} : 1 \leq i \leq k, 1 \leq j \leq m-1\}$, $\theta_{ij} = log\{1 - F_i(s_j)\} - log\{1 - F_i(s_{j-1})\}$; $\tilde{\boldsymbol{\theta}} = \{\tilde{\theta}_{ij} : 1 \leq i \leq k, 1 \leq j \leq m-1\}$, $\tilde{\theta}_{ij} = log\{1-\tilde{F}_i(s_j)\} - log\{1-\tilde{F}_i(s_{j-1})\}$; $\boldsymbol{\theta}_0 = \{\theta_{ij} = \theta_{0j} : 1 \leq i \leq k, 1 \leq j \leq m-1\}$, $\theta_{0j} = log\{1-F_0(s_j)\} - log\{1-F_0(s_{j-1})\}$; $\tilde{\boldsymbol{\theta}}_0 = \{\tilde{\theta}_{ij} = \tilde{\theta}_{0j} : 1 \leq i \leq k, 1 \leq j \leq m-1\}$, $\tilde{\theta}_{0j} = log\{1-\tilde{F}_0(s_j)\} - log\{1-\tilde{F}_0(s_{j-1})\}$. Then $\tilde{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}_0$ are the maximum likelihood estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$, respectively.

Using the relations between $\mathbf{F}$ and $\boldsymbol{\theta}$, between $\tilde{\mathbf{F}}$ and $\tilde{\boldsymbol{\theta}}$, between $\mathbf{F}_0$ and $\boldsymbol{\theta}_0$ and between $\tilde{\mathbf{F}}_0$ and $\tilde{\boldsymbol{\theta}}_0$, and (1), (2), (3) and (4), we can derive that

$$\tilde{\boldsymbol{\theta}} = Argmax\{f_n(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}, \tag{A.1}$$

$$\tilde{\boldsymbol{\theta}}_0 = Argmax\{f_n(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta_0\}, \tag{A.2}$$

and

$$S_n = 2n\,\{f_n(\tilde{\boldsymbol{\theta}}) - f_n(\tilde{\boldsymbol{\theta}}_0)\}, \tag{A.3}$$

where

$$f_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{m-1}[d_{ij}\,log\{1 - exp(\theta_{ij})\}$$

$$+\theta_{ij} \sum_{r=j+1}^{m} d_{ir}], \tag{A.4}$$

$$\Theta = \{\boldsymbol{\theta} : \sum_{r=1}^{j} \theta_{ir} \geq \sum_{r=1}^{j} \theta_{i+1,r},\ -\infty \leq \theta_{ij} \leq 0,$$
$$j = 1, \ldots, m-1, i = 1, \ldots, k-1\}, \tag{A.5}$$

and

$$\Theta_0 = \{\boldsymbol{\theta} \in \Theta : \theta_{1j} = \cdots = \theta_{kj},$$
$$j = 1, \ldots, m-1\} \tag{A.6}$$

(Feltz and Dykstra 1985, sec. 3).

Remember the relation among $\boldsymbol{\theta}_0$, $\mathbf{F}_0$ and $\mathbf{p} = (p_1, \ldots, p_m)$, and keep in mind that all arguments are under H_0, that is, $F_1 = \cdots = F_k = F_0$, and hence for each $1 \leq i \leq k$, $(d_{i1}, \ldots, d_{im})$ follows a multinomial distribution with parameters n_i and $\mathbf{p} = (p_1, \ldots, p_m)$.

Denote by $|\cdot|$ the Euclid distance in R^{mk} and let $h(x) = log(1 - e^x)$. Then

$$f_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{m-1} \{d_{ij}\, h(\theta_{ij})$$
$$+\theta_{ij} \sum_{r=j+1}^{m} d_{ir}\}, \tag{A.7}$$

and let

$$f(\boldsymbol{\theta}) = E f_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{m-1} n_i \{p_j\, h(\theta_{ij})$$
$$+\theta_{ij} \sum_{r=j+1}^{m} p_r\}. \tag{A.8}$$

Lemma A.1 *For j = 1, ..., m-1,*

$$h'(\theta_{0j}) = -\sum_{r=j+1}^{m} p_r / p_j, \tag{A.9}$$

$$h''(\theta_{0j}) = -\sum_{r=j}^{m} p_r \sum_{r=j+1}^{m} p_r/p_j^2, \tag{A.10}$$

$f(\boldsymbol{\theta})$ is a strictly concave function and $\boldsymbol{\theta}_0$ maximizes $f(\boldsymbol{\theta})$ over Θ, and there exist $M_1, M_2 > 0$ and $\rho > 0$ such that for $|\boldsymbol{\theta} - \boldsymbol{\theta}_0| \leq \rho$,

$$M_1 \, |\boldsymbol{\theta} - \boldsymbol{\theta}_0|^2 \leq f(\boldsymbol{\theta}_0) - f(\boldsymbol{\theta}) \leq M_2 \, |\boldsymbol{\theta} - \boldsymbol{\theta}_0|^2. \tag{A.11}$$

Proof. Direct computations and the fact that $\boldsymbol{\theta}_0 \in \Theta$. □

Lemma A.2

$$\lim_{c\to\infty} \lim_{n\to\infty} P(|\tilde{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0| > c\, n^{-1/2} \text{ and } |\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| > c\, n^{-1/2}) = 0.$$

Proof. Since $\tilde{\mathbf{F}}_0$ is a $n^{1/2}$-consistent estimate of $\mathbf{F}_0$, using the relations between $\tilde{\mathbf{F}}_0$ and $\tilde{\boldsymbol{\theta}}_0$ and between $\mathbf{F}_0$ and $\boldsymbol{\theta}_0$, and the delta method, we can easily show that $\tilde{\boldsymbol{\theta}}_0$ is a $n^{1/2}$-consistent estimate of $\boldsymbol{\theta}_0$ and hence the first result is proved.

We now prove the second result. By the central limit theorem we obtain

$$\begin{aligned} f_n(\boldsymbol{\theta}) = f(\boldsymbol{\theta}) + n^{-1/2} \sum_{i=1}^{k} \sum_{j=1}^{m-1} a_i^{1/2} \, [U_{ij} \, h(\theta_{ij}) \\ + \theta_{ij} \sum_{r=j+1}^{m} U_{ir}] \, \{1 + o_p(1)\}, \end{aligned} \tag{A.12}$$

where $\{U_{ij} : 1 \leq i \leq k, 1 \leq j \leq m\}$ is defined in Theorem 1. $f_n(\boldsymbol{\theta})$ and $f(\boldsymbol{\theta})$ achieve their maxima on Θ at $\tilde{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$, respectively, so (A.12) implies that $\tilde{\boldsymbol{\theta}}$ is a consistent estimate of $\boldsymbol{\theta}_0$. Applying (A.12) to $f_n(\tilde{\boldsymbol{\theta}})$ and $f_n(\boldsymbol{\theta}_0)$, and using (A.11) we obtain

$$\begin{aligned} 0 \leq f_n(\tilde{\boldsymbol{\theta}}) - f_n(\boldsymbol{\theta}_0) &= f(\tilde{\boldsymbol{\theta}}) - f(\boldsymbol{\theta}_0) + \\ & n^{-1/2} \sum_{i=1}^{k} \sum_{j=1}^{m-1} a_i^{1/2} \, [U_{ij} \, \{h(\tilde{\theta}_{ij}) - h(\theta_{0j})\} \\ & + (\tilde{\theta}_{ij} - \theta_{0j}) \sum_{r=j+1}^{m} U_{ir}] \, \{1 + o_p(1)\} \\ & \leq -M_1 \, |\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|^2 + M \, \max_{i,j}\{U_{ij}\} \, n^{-1/2} \\ & |\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| \, \{1 + o_p(1)\}, \end{aligned}$$

where M is a positive constant. The above inequality implies that $|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|$ is of order $n^{-1/2}$. □

Proof of Theorem 1: Lemma A. 2 with (A.1) and (A.2) implies that the maxima $f_n(\tilde{\boldsymbol{\theta}}) = max\{f_n(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ and $f_n(\tilde{\boldsymbol{\theta}}_0) = max\{f_n(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_0\}$ are achieved in a $n^{-1/2}$-shrinking neighborhood of $\boldsymbol{\theta}_0$, so we restrict $\boldsymbol{\theta}$ in the $n^{-1/2}$-shrinking neighborhood of $\boldsymbol{\theta}_0$ and rescale the parameter $\boldsymbol{\theta}$ as follows: Let $\boldsymbol{\theta}= \boldsymbol{\theta}_0 + \mathbf{t}\, n^{-1/2}$ or $\theta_{ij} = \theta_{0j} + t_{ij}\, n^{-1/2}$, $i = 1, \ldots, k$, $j = 1, \ldots, m-1$. Then $\boldsymbol{\theta}\in \Theta$ is equivalent to $\mathbf{t} \in \mathbf{T}$ and $\boldsymbol{\theta}_0 \in \Theta_0$ is equivalent to $\mathbf{t} \in \mathbf{T}_0$. Set

$$Z_n(\mathbf{t}) = n\{f_n(\boldsymbol{\theta}_0 + \mathbf{t}\, n^{-1/2}) - f_n(\boldsymbol{\theta}_0)\}, \tag{A.13}$$

and

$$\begin{aligned} V_1(\mathbf{t}) &= \sum_{i=1}^{k} \sum_{j=1}^{m-1} (A_{ij}\, B_{ij}\, t_{ij} - A_{ij}\, t_{ij}^2/2) \\ &= -V(\mathbf{t})/2 + \sum_{i=1}^{k} \sum_{j=1}^{m-1} A_{ij}\, B_{ij}^2/2. \end{aligned} \tag{A.14}$$

$V(\mathbf{t})$ is a positive convex quadratic function with normal random variables as coefficients of linear terms, then, with probability one, $\{V(\mathbf{t}) : \mathbf{t} \in \mathbf{T}\}$ has a unique minimum and hence $\{V_1(\mathbf{t}) : \mathbf{t} \in \mathbf{T}\}$ has a unique maximum.

Using (A.7), (A.9), (A.10) and (A.13), and doing Taylor expansions at θ_{0j}, we have

$$\begin{aligned} Z_n(\mathbf{t}) = \sum_{i=1}^{k} \sum_{j=1}^{m-1} [\, & d_{ij} \{h'(\theta_{0j})\, t_{ij}\, n^{-1/2} + \\ & h''(\theta_{0j})\, t_{ij}^2/(2n)\} + t_{ij}\, n^{-1/2} \sum_{r=j+1}^{m} d_{ij}] \\ & + O_p(n^{-1/2}) \\ = \sum_{i=1}^{k} \sum_{j=1}^{m-1} & n_i\, [\, (d_{ij}/n_i - p_j)\, \{h'(\theta_{0j})\, t_{ij}\, n^{-1/2} + \\ & h''(\theta_{0j})\, t_{ij}^2/(2n)\} + \sum_{r=j+1}^{m} (d_{ir}/n_i - p_r)\, t_{ij} \end{aligned}$$

$$
\begin{aligned}
& n^{-1/2} + \{p_j\, h'(\theta_{0j}) + \sum_{r=j+1}^{m} p_r\}\, t_{ij}\, n^{-1/2} \\
& + p_j\, h''(\theta_{0j})\, t_{ij}^2/(2n)\,] + O_p(n^{-1/2}) \\
= & \sum_{i=1}^{k} \sum_{j=1}^{m-1} [\, t_{ij}\, (n_i/n)^{1/2}\, \{h'(\theta_{0j})\, n_i^{1/2} \\
& (d_{ij}/n_i - p_j) + \sum_{r=j+1}^{m} n_i^{1/2}\, (d_{ir}/n_i - p_r)\} + \\
& p_j\, h''(\theta_{0j})\, n_i\, t_{ij}^2/(2n)\,] + O_p(n^{-1/2}),
\end{aligned}
$$

and thus simple direct argument can show that the process $\{Z_n(\mathbf{t}) : \mathbf{t} \in \mathbf{T}\}$ converges weakly to $\{V_1(\mathbf{t}) : \mathbf{t} \in \mathbf{T}\}$. By (A.1), (A.2), (A.13), and Lemma A. 2, we have that, with probability tending to one, $n\{f_n(\tilde{\boldsymbol{\theta}}_0) - f_n(\boldsymbol{\theta}_0)\} = max\{Z_n(\mathbf{t}) : \mathbf{t} \in \mathbf{T}_0\}$, and $n\{f_n(\tilde{\boldsymbol{\theta}}) - f_n(\boldsymbol{\theta}_0)\} = max\{Z_n(\mathbf{t}) : \mathbf{t} \in \mathbf{T}\}$. Hence, $n\{f_n(\tilde{\boldsymbol{\theta}}) - f_n(\boldsymbol{\theta}_0)\}$ and $n\{f_n(\tilde{\boldsymbol{\theta}}_0) - f_n(\boldsymbol{\theta}_0)\}$ converge weakly to $max\{V_1(\mathbf{t}) : \mathbf{t} \in \mathbf{T}\}$ and $max\{V_1(\mathbf{t}) : \mathbf{t} \in \mathbf{T}_0\}$, respectively.

Finally, from (A.3) and (A.14), we have

$$
\begin{aligned}
& S_n = 2n\{f_n(\tilde{\boldsymbol{\theta}}) - f_n(\tilde{\boldsymbol{\theta}}_0)\} \\
& = 2n\{f_n(\tilde{\boldsymbol{\theta}}) - f_n(\boldsymbol{\theta}_0)\} - 2n\{f_n(\tilde{\boldsymbol{\theta}}_0) - f_n(\boldsymbol{\theta}_0)\} \\
& \Longrightarrow 2\, max\{V_1(\mathbf{t}) : \mathbf{t} \in \mathbf{T}\} - \\
& \qquad 2\, max\{V_1(\mathbf{t}) : \mathbf{t} \in \mathbf{T}_0\} \\
& = min\{V(\mathbf{t}) : \mathbf{t} \in \mathbf{T}_0\} - min\{V(\mathbf{t}) : \mathbf{t} \in \mathbf{T}\}. \ \square
\end{aligned}
$$

Proof of Theorem 2

Lemma A.3 *$G(x; k, m, \mathbf{p})$ depends continuously on $\mathbf{p}$ if all $p_j > 0$ for $j = 1, \ldots, m$.*

Proof. $V(\mathbf{t})$ depends continuously on $\mathbf{p}$, so do the minima of $V(\mathbf{t})$ over $\mathbf{T}$ and over $\mathbf{T}_0$. $\square$.

Lemma A.4 *For each $\mathbf{p}$, $G(x; k, m, \mathbf{p})$ has a jump at $x = 0$ with jump size no more than a half and is continuous after zero.*

Proof. $V(\mathbf{t})$ is a quadratic function and the coefficients of its linear terms are normal random variables, so the distribution functions of the minima of $V(\mathbf{t})$ on $\mathbf{T}$ and on $\mathbf{T}_0$ are continuous. Hence, $G(x; k, m, \mathbf{p})$ has a jump at $x = 0$ and is continuous after zero. We now show that the jump size is less than a half. Denote by $\boldsymbol{\tau}$ the unique minimer of $V(\mathbf{t})$ on $\mathbf{T}$. Let

$$\tau_{0j} = \sum_{i=1}^{k} A_{ij}\, B_{ij} / \sum_{i=1}^{k} A_{ij}, \quad j = 1 \ldots, m-1,$$

and $\boldsymbol{\tau}_0 = \{\tau_{ij} = \tau_{0j}, 1 \leq i \leq k, 1 \leq j \leq m-1\}$, where A_{ij} and B_{ij} are defined in Theorem 1. It is easy to check that $\boldsymbol{\tau}_0 \in \mathbf{T}_0$ and $V(\mathbf{t})$ achieves its unique minimum on $\mathbf{T}_0$ at $\boldsymbol{\tau}_0$. Hence, the jump size is equal to $P\{V(\boldsymbol{\tau}_0) = V(\boldsymbol{\tau})\}$.

Given a z, define $\mathbf{t}(z) = \{t_{ij}(z) : t_{ij}(z) = \tau_{0j}$ except for $t_{11}(z) = z\}$. Note that $\boldsymbol{\tau}_0 \in \mathbf{T}_0$. If $z \geq \tau_{01}$, $\mathbf{t}(z) \in \mathbf{T}$ and hence $V(\boldsymbol{\tau}) \leq V(\mathbf{t}(z))$. On the other hand,

$$\begin{aligned} V(\mathbf{t}(z)) - V(\boldsymbol{\tau}_0) &= A_{11}\, \{(z - B_{11})^2 \\ &\qquad - (\tau_{01} - B_{11})^2\} \\ &= A_{11}\, (z - \tau_{01})\, (z - 2\, B_{11} + \tau_{01}), \end{aligned}$$

then $V(\mathbf{t}(z)) - V(\boldsymbol{\tau}_0) \leq 0$ if z is between τ_{01} and $2\, B_{11} - \tau_{01}$ and ≥ 0 otherwise. Therefore,

$$\begin{aligned} P\{V(\boldsymbol{\tau}_0) = V(\boldsymbol{\tau})\} &\leq P\{V(\boldsymbol{\tau}_0) = V(\boldsymbol{\tau}) \\ &\qquad \leq V(\mathbf{t}(z)) \text{ for } z \geq \tau_{01}\} \\ &\leq P\{V(\boldsymbol{\tau}_0) \leq V(\mathbf{t}(z)) \text{ for } z \geq \tau_{01}\} \\ &\leq P\{\tau_{01} \geq 2B_{11} - \tau_{01}\} \\ &= P\{\sum_{i=2}^{k} A_{i1}(B_{i1} - B_{11}) \geq 0\} = 1/2. \end{aligned}$$

The reason for the last equality is that $\sum_{i=2}^{k} A_{i1}(B_{i1} - B_{11})$ follows a normal distribution with mean zero. □

Proof of Theorem 2: By Lemma A. 4, $G(x; k, m, \mathbf{p})$ is continuous on $[\delta, \infty)$ for every $\delta > 0$. Restricting G on $[\delta, \infty)$ and using Theorem 1, Lemma A. 3 and Lemma

A. 4, we can easily prove the theorem by the arguments in proving Propositions 1.2 and 1.3 in Beran and Ducharme (1991). □

REFERENCES

Bagai, I., Deshpandé, J. V., and Kochar, S. C. (1989), "Distribution Free Tests for Stochastic Ordering in the Competing Risks Model," *Biometrika*, 76, 775-781.

Beran, R., and Ducharme, G. R. (1991), *Asymptotic Theory for Bootstrap Methods in Statistics*, Montréal: Les Publications CRM.

Bickel, P. J., and Freedman, D. A. (1991), "Some Asymptotic Theory for the Bootstrap," *The Annals of Statistics*, 9, 1196-1217.

Singh, K. (1991), "On the Asymptotic Accuracy of Efron's Bootstrap," *The Annals of Statistics*, 9, 1187-1195.

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," *Journal of the American Statistical Association*, 61, 1067-1080.

Dykstra, R. L. (1982), "Maximum Likelihood Estimation of the Survival Functions of Stochastically Ordered Random Variables," *Journal of the American Statistical Association*, 77, 621-628.

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

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

Dykstra, R. L., Madsen, R. W., and Fairbanks, K. (1983), "A Nonparametric Likelihood Ratio Test," *Journal of Statistical Computing and Simulation*, 18, 247-264.

Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, New York: Chapman & Hall.

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

Franck, W. E. (1984), "A Likelihood Ratio Test for Stochastic Ordering," *Journal of the American Statistical Association*, 79, 686-691.

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.

Kim, J., and Pollard, D. (1990), "Cube Root Asymptotics," *The Annals of Statistics*, 18, 191-219.

Lee, C. C. (1987), "Maximum Likelihood Estimates for Stochastically Ordered Multinomial Populations With Fixed and Random Zeros," in *Foundations of Statistical Inference*, eds. I. B. MacNeill and G. J. Umphrey, Boston: Reidel, pp. 189-197.

Lee, Y. J., and Wolfe, D. A. (1976), "A Distribution-Free Test for Stochastic Ordering," *Journal of the American Statistical Association*, 71, 722-727.

Lucas, L. A., and Wright, F. T. (1991), "Testing for and Against a Stochastic Ordering Between Multivariate Multinomial Populations," *Journal of Multivariate Analysis*, 38, 167-186.

Mau, J. (1988), "A Generalization of a Nonparametric Test for Stochastically Ordered Distributions to Censored Survival Data," *Journal of the Royal Statistical Society*, Ser. B, 50, 403-412.

Pollard, D. (1984), *Convergence of Stochastic Processes*, New York: Springer.

Praestgaard, J. T. (1994), "Asymptotic Theory for Nonparametric Estimation of Survival Curves Under Order Restrictions," Technical Report 228, University of Iowa, Dept. of Statistics.

Robertson, T., and Wright, F. T. (1974), "On the Maximum Likelihood Estimation of Stochastically Ordered Random Variables," *The Annals of Statistics*, 2, 528-534.

Robertson, T., and Wright, F. T. (1981), "Likelihood Ratio Tests for and Against a Stochastic Ordering Between Multinomial Populations," *The Annals of Statis-*

tics, 9, 1248-1257.

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

Sampson. A. R., and Whitaker, L. R. (1989), "Estimation of Multivariate Distributions Under Stochastic Ordering," *Journal of the American Statistical Association*, 84, 541-548.

Shorack, G., and Wellner, J. (1986), *Empirical Processes With Applications to Statistics*, New York: Wiley.

Wang, Y. (1994), "A Bartlett Adjustment for the Likelihood Ratio Statistic With an Ordered Alternative," *Statistics and Probability Letters*, 20, 347-352.

Wang, Y. (1995), "Asymptotic Expansions of the Likelihood Ratio Test Statistic With Ordered Hypotheses," *Sankhya A*, to be published.

Legend for figures:

Figure 1. The limiting distribution function $G(x; k, m, \mathbf{p})$ for $m = 3$, $\mathbf{p} = (1/3, 1/3, 1/3)$ and various values of k. Curves 1, 2, 3 and 4 correspond to $k = 2, 3, 4$ and 6, respectively.

Figure 2. The limiting distribution function $G(x; 4, 3, \mathbf{p})$ for $k = 4$, $m = 3$, and various values of $\mathbf{p}$. Curves 1, 2, 3 and 4 correspond to $\mathbf{p} = (1/5, 1/5, 3/5)$, $(1/5, 3/5, 1/5)$, $(1/3, 1/3,\ 1/3)$ and $(3/5, 1/5, 1/5)$, respectively.

Figure 3. The limiting distribution function $G(x; k, m, \mathbf{p})$ (dot curve) and the null distribution function of S_n with $n_1 = n_2 = 30$ (solid curve) for $k = 2$, $m = 3$, and $\mathbf{p} = (1/3, 1/3, 1/3)$.

Figure 4. The limiting distribution function $G(x; k, m, \mathbf{p})$ and the null distribution function of S_n with various values of n for $k = 2$, $m = 3$, and $\mathbf{p} = (1/3, 1/3, 1/3)$. Curve 0 corresponds to $G(x; k, m, \mathbf{p})$, and Curves 1, 2 and 3 correspond to the null distribution function of S_n with $(n_1, n_2) = (30, 10), (90, 30)$ and $(900, 300)$, respectively.