Annals of the Institute of Statistical Mathematics (2020) 72:451–470
https://doi.org/10.1007/s10463-018-0694-5

A test for the presence of stochastic ordering under censoring: the k-sample case

Hammou El Barmi[1]

Received: 1 November 2017 / Revised: 30 September 2018 / Published online: 1 November 2018

Abstract
In this paper, we develop an empirical likelihood-based test for the presence of stochastic ordering under censoring in the k-sample case. The proposed test statistic is formed by taking the supremum of localized empirical likelihood ratio test statistics. Its asymptotic null distribution has a simple representation in terms of a standard Brownian motion process. Through simulations, we show that it outperforms in terms of power existing methods for the same problem at all the distributions that we consider. A real-life example is used to illustrate the applicability of this new test.

Keywords Censored data · Empirical likelihood · Order-restricted inference · Stochastic ordering

1 Introduction

Stochastic ordering between univariate distributions is a very important concept in statistics and applied probability. It arises naturally in numerous situations, and it has useful applications in many areas including economics, engineering, finance and public health. Many stochastic orders exist in the literature, and they include, in increasing order of strength, stochastic ordering, uniform stochastic ordering and likelihood ratio ordering. Shaked and Shanthikumar (2007) provide a thorough review of the literature on these and other stochastic orders. It is well documented in Silvapulle and Sen (2005) and Robertson et al. (1988) that incorporating ordering constraints when they hold can increase the efficiency of estimation procedures. For this reason, it is important to develop tests for their presence.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10463-018-0694-5) contains supplementary material, which is available to authorized users.

✉ Hammou El Barmi
hammou.elbarmi@baruch.cuny.edu

1 Paul H. Chook Department of Information Systems and Statistics, Baruch College, The City University of New York, One Baruch Way, New York, NY 10010, USA

Let X_1 and X_2 be two nonnegative random variables with cumulative distribution functions (cdf) F_1 and F_2 and survival functions (SFs) $\bar{F}_1$ and $\bar{F}_2$, respectively. X_1 is said to be *stochastically smaller* than X_2 or, equivalently, F_1 is *stochastically smaller* than F_2, denoted by $F_1 \preceq_{SO} F_2$, if $\bar{F}_1(t) \leq \bar{F}_2(t)$ for all t.

Stochastic ordering has been widely studied since it was introduced in Lehmann (1955). Many tests for its presence in the two-sample case without censoring exist in the literature. Robertson and Wright (1981) developed a likelihood ratio test in the multinomial case, and Lee and Wolf (1976) proposed a Mann–Whitney–Wilcoxon-type test based on the nonparametric maximum likelihood estimators (NPMLEs) of the cdfs. Other tests were discussed in Dykstra et al. (1983), Franck (1984) and Mau (1988). Chang and McKeague (2016) and El Barmi (2017) developed nonparametric likelihood ratio-based tests for the same problem under right censoring. For more than two populations, Wang (1996) discussed the likelihood ratio test in the multinomial case, and El Barmi and Johnson (2006) showed that the limiting distribution of his test statistic is of Chi-bar-square type and gave the expression of the weighting values. El Barmi and Mukerjee (2005) provided an asymptotically distribution-free test based on a sequential testing procedure originally introduced by Hogg (1962). Even though this test is applicable in both the multinomial and the continuous cases, with or without censoring, the value of their test statistic depends heavily on how the test is carried out. Recently, Davidov and Herman (2010) developed a new nonparametric test and El Barmi and McKeague (2013) developed an empirical likelihood-based test for this situation when there is no censoring. Finally, we note that Liu et al. (1993) provide a test based on the sum of two-sample weighted log-rank statistics. However, their test can fail to detect stochastic ordering since the one-sided weighted log-rank statistics are tests for the presence of uniform stochastic ordering which is more restrictive than stochastic ordering.

The purpose of the present paper is to extend the results in El Barmi and McKeague (2013) to the censored case and the results in Chang and McKeague (2016) and El Barmi (2017) to the k-sample case. Specifically, we develop an empirical likelihood-based test for $\mathcal{H}_0 : F_1 = F_2 = \cdots = F_k$ against $\mathcal{H}_1 - \mathcal{H}_0$ where $\mathcal{H}_1$ is the stochastic ordering alternative given by $\mathcal{H}_1 : F_1 \preceq_{SO} F_2 \preceq_{SO} \cdots \preceq_{SO} F_k$ and $F_1, F_2, \ldots, F_k$ are k continuous cdfs. This test is of interest, for example, in dose–response experiments when it is believed that increasing the dosage increases the response or leaves it unchanged. In such situations, we might be interested in testing the homogeneity of the distributions corresponding to the different dosages (i.e., $\mathcal{H}_0$) against the alternative that they are increasing with the dosage (i.e., $\mathcal{H}_1$). A plot of the Kaplan–Meier estimators of these distributions can be used to check whether this hypothesis is plausible.

The computation of the test statistic as well as the study of its asymptotic null distribution does not follow easily from the two-sample case. As a result, it took some effort and it required first developing a novel method for testing $\mathcal{H}_0$ against $\mathcal{H}_2 - \mathcal{H}_0$ where $\mathcal{H}_2$ imposes no constraints on F_j, $j = 1, 2, \ldots, k$, that may be of independent interest. Once the theory has been developed for this situation, we use it to develop the desired test. We note that there is an extensive literature for testing $\mathcal{H}_0$ against $\mathcal{H}_2 - \mathcal{H}_0$. Typically, Kolmogorov–Smirnov, Cramer–von Mises or their k-sample extensions (see Kiefer 1959) are used for this situation in the uncensored case or their extensions to the presence of censoring.

The rest of the paper is organized as follows: In Sect. 2 we give the main results, and in Sect. 3 we give an algorithm to compute the test statistic and present the results of a simulation study that show that the proposed test outperforms in terms of power and other tests used for the same situation at all the distributions that we consider. We also give an example to illustrate the theory developed here, and in Sect. 4, we give some concluding remarks. Throughout the paper, $||\mathbf{x}||_{\mathbf{w}}$ and $E_{\mathbf{w}}[\mathbf{x}|A]$ are used to indicate the L_2-norm and the usual least squares projection of the vector $\mathbf{x}$ onto the set A with weights $\mathbf{w}$, respectively, and $\stackrel{d}{\to}$ and $\stackrel{w}{\Rightarrow}$ are used to denote convergence in distribution and weak convergence, respectively.

Because of the length of the proofs, Appendix containing them is given in Supplementary Material.

2 Development of the test statistic for the presence of stochastic ordering

Suppose that, for $i = 1, 2, \ldots, k$, we are given a random sample of lifetimes $X_{i1}, X_{i2}, \ldots, X_{in_i}$ from a continuous cdf F_i and a SF $\bar{F}_i$ and that these $X_{ij}s$ are, respectively, censored by $C_{i1}, C_{i2}, \ldots, C_{in_i}$, a random sample from a cdf G_i and a SF $\bar{G}_i$. As a result, we only observe $(Z_{ij}, \delta_{ij}) = (\min(X_{ij}, C_{ij}), I[X_{ij} \leq C_{ij}], i = 1, 2, \ldots, k, j = 1, 2, \ldots, n_i$. We assume that the resulting k samples are independent. Let $\mathbf{n} = (n_1, n_2, \ldots, n_k)^T$ and $\pi_i(t) = \bar{F}_i(t)\,\bar{G}_i(t)$ denote the probability of remaining under study at time t in the ith population. We assume that complete observations occur on a subset of times $T_1 < T_2 < \cdots < T_m$ and let $T_0 = 0$ and $T_{m+1} = \infty$ for convenience. Let $n = \sum_{i=1}^{k} n_i$ and assume that $\lim_{n\to\infty} n_i/n \to \gamma_i > 0$ for all $i = 1, 2, \ldots, k$. Define, for $j = 1, 2, \ldots, m$,

$$
\begin{aligned}
d_{ij} &= \text{number of complete observations from the ith sample at } T_j,\\
\ell_{ij} &= \text{number of observations censored from the ith sample at } [T_j, T_{j+1}),\\
n_{ij} &= \sum\nolimits_{\ell=j}^{n_i} (d_{i\ell} + \ell_{i\ell}) = \text{number of items from the } i\text{th population at risk at } T_j
\end{aligned}
$$

and $n_{i0} = n_i, d_{i0} = 0$. In addition, define

$$
\theta_{ij} = \frac{\bar{F}_i(T_j)}{\bar{F}_i(T_{j-1})}, \quad j = 1, 2, \ldots, m,
$$

and note that $\bar{F}_i(T_j) = \prod_{\ell=1}^{j} \theta_{i\ell}, i = 1, 2, \ldots, k$. Under no restrictions, the NPMLE of $\bar{F}_i$ is the usual Kaplan–Meier estimator (Kaplan and Meier 1958) and is given by $\hat{\bar{F}}_i(t) = \prod_{\{T_j \leq t\}} \hat{\theta}_{ij}$ where

$$
\hat{\theta}_{ij} = \frac{n_{ij} - d_{ij}}{n_{ij}}, \quad i = 1, 2, \ldots, m, \tag{1}
$$

($\prod_\emptyset = 1$) and is obtained by maximizing $L(\bar{F}_i) = \prod_{j=1}^m \theta_{ij}^{n_{ij}-d_{ij}}(1-\theta_{ij})^{d_{ij}}$. Next, we develop empirical likelihood-based tests for testing $\mathcal{H}_0$ against $\mathcal{H}_2 - \mathcal{H}_0$ and $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_0$, where $\mathcal{H}_0$, $\mathcal{H}_1$ and $\mathcal{H}_2$ are defined before. These tests will be carried out on a fixed interval $[\tau_1, \tau_2]$.

2.1 Test for $\mathcal{H}_0$ against $\mathcal{H}_2 - \mathcal{H}_0$

The approach we follow is based on translating the problem into testing a family of "local" null hypotheses of the form $\mathcal{H}_0^t : (\bar{F}_1(t), \bar{F}_2(t), \ldots, \bar{F}_k(t))^T \in \mathcal{I}_0$ against $\mathcal{H}_2^t : (\bar{F}_1(t), \bar{F}_2(t), \ldots, \bar{F}_k(t))^T \notin \mathcal{I}_0$ for a given t where $\mathcal{I}_0 = \{\mathbf{x} \in \mathbf{R}^k, x_1 = x_2 = \cdots = x_k\}$. The local empirical likelihood ratio test rejects $\mathcal{H}_0^t$ for small values of

$$\mathcal{R}_{02}(t) = \frac{\sup\left\{\prod_{i=1}^k L(\bar{F}_i) : (\bar{F}_1(t), \bar{F}_2(t), \ldots, \bar{F}_k(t))^T \in \mathcal{I}_0\right\}}{\sup\left\{\prod_{i=1}^k L(\bar{F}_i)\right\}}, \tag{2}$$

where $L(\bar{F}_i)$ is as given before, and we use the convention $\sup \emptyset = 0$ and $0/0 = 1$.

Clearly, under no constraints, $L(\bar{F}_i)$ achieves its maximum value at a vector whose jth component is given in (1). To compute the numerator in (2), write

$$L(\bar{F}_i) = \left\{\prod_{j=1}^{n(t)} \theta_{ij}^{n_{ij}-d_{ij}}(1-\theta_{ij})^{d_{ij}}\right\} \times \left\{\prod_{j=n(t)+1}^{m} \theta_{ij}^{n_{ij}-d_{ij}}(1-\theta_{ij})^{d_{ij}}\right\} \tag{3}$$

where for $u > 0, n(u) \equiv \sum_{i=1}^n I[T_i \leq u]$ is the number of distinct uncensored observations in the time interval $[0, u]$. Since $\bar{F}_i(t) = \prod_{j=1}^{n(t)} \theta_{ij}$ for each i, the terms in braces in (3) can be maximized separately under $\mathcal{H}_0$, and since the constraints on the θ_{ij}s in the second term are exactly the same in the numerator and the denominator of (2), this term has not effect and thus makes no contribution to $\mathcal{R}_{02}(t)$. The remaining term

$$\prod_{i=1}^{k} \prod_{j=1}^{n(t)} \theta_{ij}^{n_{ij}-d_{ij}}(1-\theta_{ij})^{d_{ij}} \tag{4}$$

is then maximized under $\mathcal{H}_0^t$. To do so, we can use the algorithm described in Sect. 3 but to derive the limiting distributions of $\mathcal{R}_{02}(t)$ and the test that we propose for $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_2$, we find it useful to proceed as follows: First, we consider $\bar{F}_i(t), i = 1, 2, \ldots, k$, fixed and maximize (4) subject to $\prod_{j=1}^{n(t)} \theta_{ij} = \bar{F}_i(t)$ or equivalently $\sum_{j=1}^{n(t)} \log(\theta_{ij}) = \mu_i$ where $\mu_i \equiv \mu_i(t) = \log(\bar{F}_i(t)), i = 1, 2, \ldots, k$. (We have suppressed the argument t in μ_i for simplicity.)

Write the Lagrangian corresponding to this optimization problem as

$$\sum_{i=1}^{k}\left[\sum_{j=1}^{n(t)}\left[(n_{ij}-d_{ij})\log(\theta_{ij})+d_{ij}\log(1-\theta_{ij})\right]+n_i\beta_i\left(\sum_{j=1}^{n(t)}\log(\theta_{ij})-\mu_i\right)\right].$$

Its solution is

$$\tilde{\theta}_{ij}^{(0)}\equiv\tilde{\theta}_{ij}^{(0)}(\mu_i)=1-\frac{d_{ij}}{n_{ij}+n_i\beta_i}=1-\frac{d_{ij}/n_i}{n_{ij}/n_i+\beta_i} \tag{5}$$

where $\beta_i\equiv\beta_i(t)(\mu_i)$ satisfies

$$\sum_{j=1}^{n(t)}\log\left(1-\frac{d_{ij}/n_i}{n_{ij}/n_i+\beta_i}\right)-\mu_i=0,\quad i=1,2,\ldots,k. \tag{6}$$

The values in (5) are then plugged into (4) to obtain the profile likelihood

$$L(\boldsymbol{\mu})=\prod_{i=1}^{k}\prod_{j=1}^{n(t)}\left(1-\frac{d_{ij}/n_i}{n_{ij}/n_i+\beta_i}\right)^{n_{ij}-d_{ij}}\left(\frac{d_{ij}/n_i}{n_{ij}/n_i+\beta_i}\right)^{d_{ij}} \tag{7}$$

where $\boldsymbol{\mu}=(\mu_1,\mu_2,\ldots,\mu_k)^T$. The profile likelihood is then maximized subject to $\mu_1=\mu_2=\cdots=\mu_k$ by considering

$$M=\frac{1}{n}\log(L(\boldsymbol{\mu}))+\boldsymbol{\lambda}^T\mathbf{h}(\boldsymbol{\mu})$$

where $\mathbf{h}=(h_1,h_2,\ldots,h_{k-1})^T$ with $h_j(\boldsymbol{\mu})=\mu_{j+1}-\mu_j$, $j=1,2,\ldots,k-1$, and $\boldsymbol{\lambda}$, a vector of Lagrange multipliers, is used instead of $\boldsymbol{\lambda}(t)$ for brevity. Differentiating M with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$, we get

$$\frac{1}{n}\nabla\log L(\boldsymbol{\mu})+H\boldsymbol{\lambda}=\mathbf{0}\qquad\text{and}\qquad h_j(\boldsymbol{\mu})=0,\quad j=1,2,\ldots,k-1,$$

where $H=[\nabla h_1(\boldsymbol{\mu}),\nabla h_2(\boldsymbol{\mu}),\ldots,\nabla h_{k-1}(\boldsymbol{\mu})]$. Consequently, to compute the numerator of $\mathcal{R}_{02}(t)$, we seek the solution of

$$Q_{i\mathbf{n}}(\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\lambda})=\mathbf{0},\quad i=1,2,3, \tag{8}$$

where

$$Q_{1\mathbf{n}}(\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\lambda}) = \begin{pmatrix} \sum_{j=1}^{n(t)} \log\left(1-\frac{d_{1j}/n_1}{n_{1j}/n_1+\beta_1}\right)-\mu_1 \\ \sum_{j=1}^{n(t)} \log\left(1-\frac{d_{2j}/n_2}{n_{2j}/n_2+\beta_2}\right)-\mu_2 \\ \vdots \\ \sum_{j=1}^{n(t)} \log\left(1-\frac{d_{kj}/n_k}{n_{kj}/n_k+\beta_k}\right)-\mu_k \end{pmatrix}$$

$$Q_{2\mathbf{n}}(\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\lambda}) = \frac{1}{n}\nabla \log L(\boldsymbol{\mu}) + H\boldsymbol{\lambda} \quad \text{and}$$
$$Q_{3\mathbf{n}}(\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\lambda}) = \mathbf{h}(\boldsymbol{\mu}).$$

We denote this solution by $(\hat{\boldsymbol{\beta}}^{(0)}, \hat{\boldsymbol{\mu}}^{(0)}, \hat{\boldsymbol{\lambda}}^{(0)})$ and let $\boldsymbol{\mu}^{(0)}$ be the true value of $\boldsymbol{\mu}$. Assume that $\mathcal{H}_0$ holds in which case we write $\boldsymbol{\mu}^{(0)} = (\log(\bar{F}(t)), \log(\bar{F}(t)), \ldots, \log(\bar{F}(t)))^T$ where F is the common distribution under $\mathcal{H}_0$. Next we derive the asymptotic distribution of the estimators $(\hat{\boldsymbol{\beta}}^{(0)}, \hat{\boldsymbol{\mu}}^{(0)}, \hat{\boldsymbol{\lambda}}^{(0)})$. The method we use here is similar to a method developed in Qin and Lawless (1995). First we give a lemma to show that from $Q_{1n}(\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\lambda}) = \mathbf{0}$, we can determine uniquely $\boldsymbol{\beta}(t) = \boldsymbol{\beta}(t)(\boldsymbol{\mu})$ in a neighborhood of $\boldsymbol{\mu}^{(0)}$.

Lemma 1 *Let $a_n = n^{-1/3-\delta}$ where $0 < \delta < 1/6$, then the equation $Q_{1\mathbf{n}}(\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\lambda}) = \mathbf{0}$ has almost surely roots $\boldsymbol{\beta} = \boldsymbol{\beta}(t)(\boldsymbol{\mu}) = O(a_n)$ in the sphere $\{\boldsymbol{\mu} : ||\boldsymbol{\mu} - \boldsymbol{\mu}^{(0)}|| \le a_n\}$. In addition, $\boldsymbol{\beta}$ is continuous and differentiable with respect to $\boldsymbol{\mu}$ where $\boldsymbol{\mu}$ is in this sphere.*

The following lemma shows that the equations in (8) have a solution.

Lemma 2 *Let $a_n = n^{-1/3-\delta}$. The equations in (8) almost surely have solutions in $\mathcal{U}_{a_n} = \{(\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\lambda}) : ||\boldsymbol{\mu}-\boldsymbol{\mu}^{(0)}|| + ||\boldsymbol{\beta}|| + ||\boldsymbol{\lambda}|| \le a_n\}$ as $n \to \infty$ and any solution of (8) in $\mathcal{U}_{a_n}$ maximizes $L(\boldsymbol{\mu})$ subject to $\mathbf{h}(\boldsymbol{\mu}) = \mathbf{0}$.*

For $i = 1, 2, \ldots, k$, let

$$c_i(t) = \int_0^t \frac{\mathrm{d}\Lambda_i(u)}{\pi_i(u)}, \quad \hat{c}_i(t) = n_i \sum_{T_j \le t} \frac{d_{ij}}{n_{ij}(n_{ij}-d_{ij})}$$

where $\Lambda_i(t)$ is the cumulative hazard function corresponding to F_i. Let $C \equiv C(t) = \mathrm{diag}(c_1(t), c_2(t), \ldots, c_k(t))$, $\hat{C} \equiv \hat{C}(t) = \mathrm{diag}(\hat{c}_1(t), \hat{c}_2(t), \ldots, \hat{c}_k(t))$, $w_i \equiv w_i(t) = \frac{c_i(t)}{\gamma_i}$ and $\tilde{w}_i \equiv \tilde{w}_i(t) = 1/w_i$, $i = 1, 2, \ldots, k$. The following key theorem gives the asymptotic distribution of $\hat{\boldsymbol{\mu}}^{(0)}$ and $\hat{\boldsymbol{\lambda}}^{(0)}$ and is key in proving that locally the local EL statistic for testing $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_0$ has a Chi-bar-square distribution under $\mathcal{H}_0$.

Theorem 1

$$\sqrt{n}\hat{\boldsymbol{\xi}}^{(0)} \equiv \begin{pmatrix} \sqrt{n}(\hat{\boldsymbol{\mu}}^{(0)} - \boldsymbol{\mu}^{(0)}) \\ \sqrt{n}\hat{\boldsymbol{\lambda}}^{(0)} \end{pmatrix} \xrightarrow{d} N\left(\mathbf{0}, \begin{pmatrix} P & \mathbf{0} \\ \mathbf{0} & R \end{pmatrix}\right)$$

where $P \equiv P(t) = \tilde{C}(I - H(H^T\tilde{C}H)^{-1}H^T\tilde{C})$ and $R \equiv R(t) = (H^T\tilde{C}H)^{-1}$ *and* $\tilde{C} \equiv \tilde{C}(t) = \mathrm{diag}(w_1, w_2, \ldots, w_k)$.

The following result shows that the local EL statistic is asymptotically equivalent to a Lagrange multiplier statistic and has asymptotically a χ^2 distribution with $k-1$ degrees of freedom.

Lemma 3 *Under* $\mathcal{H}_0$

$$-2\log(\mathcal{R}_{02}(t)) = n[\hat{\boldsymbol{\lambda}}^{(0)}(t)]^T H^T \hat{\hat{C}} H \hat{\boldsymbol{\lambda}}^{(0)}(t) + O_p(n^{-1/2}) \xrightarrow{d} \chi^2_{k-1}.$$

The proof of this result follows using a Taylor expansion of $-2\log(\mathcal{R}_{02}(t))$ to the fifth order and Theorem 1. It is omitted as it follows from the proof of a more general result given in Theorem 2.

Remark 1 Equation (S.1) in Supplementary Material implies that $n_1\hat{\beta}_1^{(0)} = n\hat{\lambda}_1^{(0)}$, $n_i\hat{\beta}_i^{(0)} = n(\hat{\lambda}_i^{(0)} - \hat{\lambda}_{i-1}^{(0)})$, $i = 2, 3, \ldots, k-1$, and $n_k\hat{\beta}_k^{(0)} = -n\hat{\lambda}_{k-1}^{(0)}$. Consequently, (4) is maximized under $\mathcal{H}_0^t$ by

$$\hat{\theta}_{ij}^{(0)} = 1 - \frac{d_{ij}}{n_{ij} + n_i\hat{\beta}_i^{(0)}} = 1 - \frac{d_{ij}}{n_{ij} + n\left(\hat{\lambda}_i^{(0)} - \hat{\lambda}_{i-1}^{(0)}\right)}, \quad 1 \le i \le k, \quad 1 \le j \le n(t),$$

where $\hat{\lambda}_0^{(0)} = \hat{\lambda}_k^{(0)} = 0$. Also a careful inspection of the proof of Theorem 1 shows that when $k = 2$, $H = (1, -1)^T$, $\hat{\beta}_1^{(0)} = n\hat{\lambda}^{(0)}/n_1$ and $\hat{\beta}_2^{(0)} = -n\hat{\lambda}^{(0)}/n_2$. Therefore

$$\hat{\theta}_{1j}^{(0)} = 1 - \frac{d_{ij}}{n_{ij} + n\hat{\lambda}^{(0)}} \quad \text{and} \quad \hat{\theta}_{2j}^{(0)} = 1 - \frac{d_{ij}}{n_{ij} - n\hat{\lambda}^{(0)}}, \quad j = 1, 2, \ldots, n(t).$$

In addition,

$$\frac{n_1\hat{\beta}_1^{(0)}}{\sqrt{n}} = -\frac{n_2\hat{\beta}_2^{(0)}}{\sqrt{n}} = \sqrt{n}\hat{\lambda}^{(0)} = \frac{1}{c(t)}\sqrt{n}(\log(\hat{\hat{F}}_1(t)) - \log(\hat{\hat{F}}_2(t))) + o_p(1)$$
$$\xrightarrow{d} \frac{1}{c(t)} N(0, c(t)) = N(0, 1/c(t))$$

under $\mathcal{H}_0$, where $c(t) = \frac{1}{\gamma_1}c_1(t) + \frac{1}{\gamma_2}c_2(t)$. This result was proved in Præstgaard and Huang (1996).

Let $\log(\hat{\mathbf{F}}(t)) = (\log(\hat{\hat{F}}_1(t)), \log(\hat{\hat{F}}_2(t)), \ldots, \log(\hat{\hat{F}}_k(t)))^T$. The following two lemmas are key in the derivation of the limiting distribution of the test statistics that we propose for testing $\mathcal{H}_0$ against $\mathcal{H}_2 - \mathcal{H}_0$ and $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_0$.

Lemma 4 *Under $\mathcal{H}_0$, if τ_1 and τ_2 are such that $\bar{F}(\tau_1) < 1$ and $\min_{1\le i\le k}\pi_i(\tau_2) > 0$, then*

$$\hat{\boldsymbol{\beta}}^{(0)}(t) = O_p(n^{-1/2}), \tag{9}$$

$$\hat{\boldsymbol{\lambda}}^{(0)}(t) = R(t)H^T \log(\hat{\bar{\mathbf{F}}}(t)) + O_p(n^{-1}) \tag{10}$$

and

$$\hat{\boldsymbol{\mu}}^{(0)}(t) = P(t)\tilde{C}^{-1}(t)\log(\hat{\bar{\mathbf{F}}}(t)) + O_p(n^{-1}) \tag{11}$$

uniformly in t over $[\tau_1, \tau_2]$ where $R(t)$ and $P(t)$ are as defined in Theorem 1.

Let $\mathbf{V} = (V_1, V_2, \ldots, V_k)^T$ with $V_i(t) = \gamma_i^{-1/2} W_i(c_i(t)), i = 1, 2, \ldots, k,$ and $W_1, W_2, \ldots, W_k$ are independent Brownian motions. If

$$\mathbf{U}_{1n}(t) = \sqrt{n}\hat{\boldsymbol{\lambda}}^{(0)}(t) \quad \text{and} \quad \mathbf{U}_{2n}(t) = \sqrt{n}(\hat{\boldsymbol{\mu}}^{(0)}(t) - \boldsymbol{\mu}^{(0)}(t)),$$

then the following result holds.

Lemma 5 *Under $\mathcal{H}_0$, if τ_1 and τ_2 satisfy $\bar{F}(\tau_1) < 1$ and $\min_{1\le i\le k}\pi_i(\tau_2) > 0$, then*

$$\left\{\begin{pmatrix}\mathbf{U}_{1n}(t)\\ \mathbf{U}_{2n}(t)\end{pmatrix}, \tau_1 \le t \le \tau_2\right\} \overset{w}{\Rightarrow} \left\{\begin{pmatrix}\mathbf{U}_{1}(t)\\ \mathbf{U}_{2}(t)\end{pmatrix}, \tau_1 \le t \le \tau_2\right\} \tag{12}$$

where

$$\mathbf{U}_1(t) = R(t)H^T\mathbf{V}(t), \quad \text{and} \quad \mathbf{U}_2(t) = P(t)\tilde{C}^{-1}(t)\mathbf{V}(t)). \tag{13}$$

In addition, $\mathbf{U}_1$ and $\mathbf{U}_2$ are independent Gaussian processes.

To test $\mathcal{H}_0$ against $\mathcal{H}_2 - \mathcal{H}_0$, we propose to use

$$\mathcal{T}_{02} = \sup_{\tau_1\le t\le\tau_2} [-2\log(\mathcal{R}_{02}(t))].$$

Theorem 2 *Under $\mathcal{H}_0$, if τ_1 and τ_2 are such that $\bar{F}(\tau_1) < 1$ and $\min_{1\le i\le k}\pi_i(\tau_2) > 0$, then*

$$\mathcal{T}_{02} \overset{d}{\to} \sup_{\tau_1\le t\le\tau_2} ||\mathbf{V}(t) - E_{\tilde{\mathbf{w}}(t)}[\mathbf{V}(t)|\mathcal{I}_0]||^2_{\tilde{\mathbf{w}}(t)}. \tag{14}$$

If in addition all the samples have a common censoring distribution G, in which case $c_1 = c_2 = \cdots = c_k \equiv c$, then

$$\mathcal{T}_{02} \overset{d}{\to} \sup_{u_1\le u\le u_2} \sum_{i=1}^{k} \gamma_i \frac{\left(\gamma_i^{-1/2}B_i(u) - \sum_{j=1}^{k}\gamma_j^{1/2}B_j(u)\right)^2}{u(1-u)} \tag{15}$$

where $B_1, B_2, \ldots, B_k$ *are independent Brownian bridges and* $u_i = c(\tau_i)/(1 + c(\tau_i)), i = 1, 2.$

Remark 2 It is easy to check that when $k = 2$,

$$||\mathbf{V}(t) - E_{\tilde{\mathbf{w}}(t)}[\mathbf{V}(t)|\mathcal{I}_0]||^2_{\tilde{\mathbf{w}}(t)} = \frac{(\gamma_1^{-1/2} W_1(c_1(t)) - \gamma_2^{-1/2} W_2(c_2(t)))^2}{c(t)}$$

where again $c(t) = \gamma_1^{-1} c_1(t) + \gamma_2^{-1} c_2(t)$. If we let $W(t) = \gamma_1^{-1/2} W_1(t) - \gamma_2^{-1/2} W_2(t)$, then $W(t)$ is a standard Brownian motion. Using the fact that $B(t) = (1-t)W(t/(1-t))$ is standard Brownian Bridge, (14) can be expressed as

$$\mathcal{T}_{02} \xrightarrow{d} \sup_{\tau_1 \le t \le \tau_2} \frac{W^2(c(t))}{c(t)} \stackrel{d}{=} \sup_{u_1 \le u \le u_2} \frac{B^2(u)}{u(1-u)}$$

where $u_i = c^{-1}(\tau_i/(1-\tau_i)), i = 1, 2$, and $c^{-1}(u) = \inf\{t, c(t) \ge u\}$.

Clearly the limiting distributions in (14) and (15) are not distribution-free. To implement the test when the censoring distributions are not the same, we need to pre-specify the interval $[\tau_1, \tau_2]$. The choice can be based on the smallest and largest observed values or some other biological consideration as suggested in Chang and McKeague (2016). The critical values in this case cannot be tabulated, and to approximate the p values and the powers of the tests we propose, we adapt a technique developed in Parzen et al. (1997) for constructing confidence intervals for the difference of two survival functions. A similar technique was also used in El Barmi et al. (2008) to compare the cumulative incidence functions of a competing risk over several populations and in Chang et al. (2016) to test for stochastic ordering under biased sampling. Specifically, let

$$\hat{\mathbf{V}}(t) = \sqrt{n}\left(\log(\hat{\bar{F}}_1(t)/\bar{F}_1(t)), \log(\hat{\bar{F}}_2(t)/\bar{F}_2(t)), \ldots, \log(\hat{\bar{F}}_k(t)/\bar{F}_k(t))\right)$$

and note that, under $\mathcal{H}_0$, $H^T\hat{\mathbf{V}}(t) = \sqrt{n}H^T \log(\hat{\bar{\mathbf{F}}})$. Therefore, if $\tilde{\boldsymbol{\lambda}}(t) = [H^T\hat{\bar{C}}(t)H]^{-1}H^T\hat{\mathbf{V}}(t)/\sqrt{n}$. Then by (10), under $\mathcal{H}_0$, $\tilde{\boldsymbol{\lambda}}(t) = \hat{\boldsymbol{\lambda}}^{(0)}(t) + O_p(n^{-1})$ uniformly on $[\tau_1, \tau_2]$. Using the functional delta method

$$\hat{\mathbf{V}} \stackrel{w}{\Longrightarrow} \mathbf{V}$$

on $[\tau_1, \tau_2]$. As a result, if we let $\hat{\mathcal{T}}_{02} = n \sup_{\tau_1 \le t \le \tau_2} [\tilde{\boldsymbol{\lambda}}(t)]^T [H^T\hat{\bar{C}}H]\tilde{\boldsymbol{\lambda}}(t)$, then

$$\hat{\mathcal{T}}_{02} \xrightarrow{d} \sup_{\tau_1 \le t \le \tau_2} \mathbf{V}(t)^T HR(t)H^T\mathbf{V}(t) = \sup_{\tau_1 \le t \le \tau_2} ||\mathbf{V}(t) - E_{\tilde{\mathbf{w}}(t)}[\mathbf{V}(t)|\mathcal{I}_0]||^2_{\tilde{\mathbf{w}}(t)}$$

where the last equality follows from the proof of Theorem 2 in Supplementary Material. Moreover, under $\mathcal{H}_0$, $\mathcal{T}_{02}$ and $\hat{\mathcal{T}}_{02}$ have exactly the same asymptotic distribution. Now the results in Parzen et al. (1997) imply that the distribution of this limiting distribution

Table 1 Selected estimated asymptotic critical points of $\mathcal{T}_{02}$

k	Significance level α		
	0.10	0.05	0.01
2	3.683	4.479	6.191
3	5.083	6.312	9.168
4	5.499	6.586	8.396
5	6.064	7.116	9.321

can be approximated by holding the data fixed and simulating the independent standard normal covariates Z_{ij}s in

$$\tilde{V}_i(t) = -\sqrt{n} \sum_{T_j \le t} \frac{d_{ij} I[X_{ij} \le t] Z_{ij}}{n_{ij}} = -\sqrt{n_i} \sum_{j=1}^{n_i} \frac{d_{ij} I[X_{ij} \le t] Z_{ij}}{\sum_{\ell=1}^{n_i} I[X_{i\ell} \ge t]}, \quad 1 \le i \le k,$$

and computing $\tilde{\mathcal{T}}_{02} = n \sup_{\tau_1 \le t \le \tau_2} [\tilde{\mathbf{V}}(t)]^T H [H^T \hat{\hat{C}} H] H \tilde{\mathbf{V}}(t)$ where $\tilde{\mathbf{V}}(t) = (\tilde{V}_1(t), \tilde{V}_2(t), \ldots, \tilde{V}_k(t))^T$. To approximate the p value, we may calculate the percentage of the simulated values of $\tilde{\mathcal{T}}_{02}$ that are greater than the observed value of $\mathcal{T}_{02}$. In addition, to approximate the power of $\mathcal{T}_{02}$ at a given alternative, for each simulated data, we compute the value of $\mathcal{T}_{02}$ and simulate the values of $\tilde{\mathcal{T}}_{02}$ by repeatedly generating the independent normal variates Z_{ij}s while holding the observed data fixed. We then compare appropriate quantiles of these values of $\tilde{\mathcal{T}}_{02}$ with the value of $\mathcal{T}_{02}$.

To implement our test under a common censoring distribution, one can fix u_1 and u_2 (for example, take $u_1 = 0.1$ and $u_2 = 0.90$ and use $\tau_i = \hat{c}^{-1}(u_i/(1-u_i)), i = 1, 2$, where $\hat{c}^{-1}(x) = \inf\{t, \hat{c}(t) \ge x\}$ and $\hat{c} = \sum_{i=1}^{k} \frac{n_i}{n} \hat{c}_i$. is the pooled estimate of c. The validity of this methods follows from the results in Chang and McKeague (2016). The estimated cutoff points corresponding to $u_1 = 1 - u_2 = 0.10$ are given in Table 1. For more discussion on the choice of $[\tau_1, \tau_2]$, see Davidov and Herman (2010).

Remark 3 When there is no censoring, $c(t)/(1+c(t)) = F(t)$ where F is the common distribution under $\mathcal{H}_0$ and $u_i = F(t_i), i = 1, 2$. In this case, we can use $\tau_i = \hat{F}^{-1}(u_i)$ where $\hat{F}$ is the pooled estimate of the common cdf F.

2.2 Test for $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_0$

Next, we consider testing $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_0$. Adapting the same approach as before, we first test $\mathcal{H}_0^t$ against $\mathcal{H}_1^t - \mathcal{H}_0^t$ where $\mathcal{H}_0^t$ is defined above and $\mathcal{H}_1^t$ is $\mathcal{H}_1^t : (\bar{F}_1(t), \bar{F}_2(t), \ldots, \bar{F}_k(t)) \in \mathcal{I}_1$ with $\mathcal{I}_1 = \{\mathbf{x} \in \mathbf{R}^k, x_1 \le x_2 \le \cdots \le x_k\}$. The localized empirical likelihood ratio test statistic in this case is

$$\mathcal{R}_{01}(t) = \frac{\sup\left\{\prod_{i=1}^{k} L(\bar{F}_i) : (\bar{F}_1(t), \bar{F}_2(t), \ldots, \bar{F}_k(t))^T \in \mathcal{I}_0\right\}}{\sup\left\{\prod_{i=1}^{k} L(\bar{F}_i) : (\bar{F}_1(t), \bar{F}_2(t), \ldots, \bar{F}_k(t))^T \in \mathcal{I}_1\right\}}.$$

To compute $\mathcal{R}_{01}(t)$ we require also maximizing (4) under $\mathcal{H}_1^t$ or equivalently maximizing (7) subject to $h_j(\boldsymbol{\mu}) \leq 0$, $j = 1, 2, \ldots, k-1$. We denote by $(\hat{\boldsymbol{\beta}}^{(1)}, \hat{\boldsymbol{\mu}}^{(1)}, \hat{\boldsymbol{\lambda}}^{(1)})$ the optimal values of $(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\lambda})$ in this case. The solution of this maximization is given by

$$\hat{\theta}_{ij}^{(1)} = 1 - \frac{d_{ij}}{n_{ij} + n_i \hat{\beta}_i^{(1)}} = 1 - \frac{d_{ij}}{n_{ij} + n(\hat{\lambda}_i^{(1)} - \hat{\lambda}_{i-1}^{(1)})}, \quad 1 \leq i \leq k, 1 \leq j \leq n(t),$$

where again $\hat{\lambda}_1^{(1)} = \hat{\lambda}_k^{(1)} = 0$. An algorithm to compute this solution is also described in Sect. 3, but for the asymptotics, we proceed as in the previous section. The following lemma holds.

Lemma 6 *Under* $\mathcal{H}_0$

$$-2\log(\mathcal{R}_{01}(t)) \xrightarrow{d} ||E_{\tilde{\mathbf{w}}(t)}[\mathbf{V}(t)|\mathcal{I}_1] - E_{\tilde{\mathbf{w}}(t)}[\mathbf{V}(t)|\mathcal{I}_0]||^2_{\tilde{\mathbf{w}}(t)}. \tag{16}$$

The limiting distribution of $-2\log(\mathcal{R}_{01}(t))$ is known as a Chi-bar-square distribution. Specifically, under $\mathcal{H}_0$,

$$\lim_{n\to\infty} P(-2\log(\mathcal{R}_{01}(t)) \geq a) = \sum_{\ell=1}^{k} p_{\tilde{\mathbf{w}}}(\ell, k) P\left(\chi^2_{\ell-1} \geq a\right)$$

where χ^2_ℓ denotes the central Chi-square distribution with ℓ degrees of freedom and $\chi^2_0 \equiv 0$. The weight $p_{\tilde{\mathbf{w}}}(\ell, k)$, also known as a level probability, is the probability that the least squares projection of $(V_1(t), V_2(t), \ldots, V_k(t))$ with weights $\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_k$ onto $\mathcal{I}_1$ has exactly ℓ levels. These level probabilities are sums of products of normal orthant probabilities, and in general, they do not exist in a closed form. However, when $\tilde{w}_1 = \tilde{w}_2 = \cdots = \tilde{w}_k$, they do not depend on $\tilde{\mathbf{w}}$, which is then omitted, and they satisfy the following recurrence relation

$$p(\ell, k) = \frac{1}{k} p(\ell - 1, k - 1) + \frac{k-1}{k} p(\ell, k - 1)$$

where $p(0, k-1) = p(k, k-1) = 0$. For more on this, see Robertson et al. (1988) and Silvapulle and Sen (2005). To test H_0 against H_1, we introduce the test statistic

$$\mathcal{T}_{01} = \sup_{\tau_1 \leq t \leq \tau_2} [\mathcal{R}_{01}(t)].$$

The following result gives the asymptotic null distribution of $\mathcal{T}_{01}$.

Theorem 3 *Under* $\mathcal{H}_0$, *if* τ_1 *and* τ_2 *are such that* $\bar{F}(\tau_1) < 1$ *and* $\min_{1\leq i\leq k}\pi_i(\tau_2) > 0$, *then*

$$\mathcal{T}_{01} \xrightarrow{d} \sup_{\tau_1 \leq t \leq \tau_2} ||E_{\tilde{\mathbf{w}}(t)}[\mathbf{V}(t)|\mathcal{I}_1] - E_{\tilde{\mathbf{w}}(t)}[\mathbf{V}(t)|\mathcal{I}_0]||^2_{\tilde{\mathbf{w}}(t)}. \tag{17}$$

If in addition all the samples are censored by the same distribution G, *in which case* $c_1 = c_2 = \cdots = c_k \equiv c$,

$$\mathcal{T}_{01} \xrightarrow{d} \sup_{u_1 \le u \le u_2} \sum_{i=1}^{k} \gamma_i \frac{\left(E_{\boldsymbol{\gamma}}[\tilde{\mathbf{B}}(u)|\mathcal{I}_1]_i - \sum_{j=1}^{k} \gamma_j^{1/2} B_j(u)\right)^2}{u(1-u)} \tag{18}$$

where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_k)^T$, $\tilde{\mathbf{B}} = (\gamma_1^{-1/2} B_1, \gamma_2^{-1/2} B_2, \ldots, \gamma_k^{-1/2} B_k)^T$ *with* $B_1, B_2, \ldots, B_k$ *independent Brownian bridges and* $u_i = c(\tau_i)/(1 + c(\tau_i))$, $i = 1, 2$.

Remark 4 When $k = 2$, it is easy to show that

$$E_{\tilde{\mathbf{w}}(t)}[\mathbf{V}(t)|\mathcal{I}_1]_i = V_i(t) I[V_1(t) \le V_2(t)] + E_{\tilde{\mathbf{w}}(t)}[\mathbf{V}(t)|\mathcal{I}_0]_i I[V_1(t) > V_2(t)], i = 1, 2,$$

and

$$E_{\tilde{\mathbf{w}}(t)}[\mathbf{V}(t)|\mathcal{I}_0]_1 = E_{\tilde{\mathbf{w}}(t)}[\mathbf{V}(t)|\mathcal{I}_0]_2 = \bar{V}(t) = \frac{\tilde{w}_1(t)V_1(t) + \tilde{w}_2(t)V_2(t)}{\tilde{w}_1(t) + \tilde{w}_2(t)}.$$

Consequently,

$$||E_{\tilde{\mathbf{w}}(t)}[\mathbf{V}(t)|\mathcal{I}_1] - E_{\tilde{\mathbf{w}}(t)}[\mathbf{V}(t)|\mathcal{I}_0]||^2_{\tilde{\mathbf{w}}(t)} = \frac{[(V_2(t) - V_1(t))_+]^2}{c(t)}$$

where for any real number a, $a_+ = \max(a, 0)$. Using this and the same steps as in Remark 2, we get

$$\mathcal{T}_{01} \xrightarrow{d} \sup_{u_1 \le u \le u_2} \frac{B_+^2(u)}{u(1-u)}$$

where $B(u)$, u_1, u_2 and $c^{-1}(u)$ are as defined in Remark 2. This is the result obtained in Chang and McKeague (2016) who showed via simulations that this test outperforms a weighted Kaplan–Meier test proposed in Pepe and Fleming (1989) and the one-sided log-rank test under alternatives with crossing hazard rates.

Again the limiting distributions in (17) and (18) are not tractable, but the same the technique described above can be used to approximate the p values and the power corresponding to $\mathcal{T}_{01}$. Specifically, let

$$\hat{\mathcal{T}}_{01} = \sup_{\tau_1 \le t \le \tau} |E_{\hat{\tilde{\mathbf{w}}}(t)}[\hat{\mathbf{V}}(t)|\mathcal{I}_1] - E_{\hat{\tilde{\mathbf{w}}}(t)}[\hat{\mathbf{V}}(t)|\mathcal{I}_0]||^2_{\hat{\tilde{\mathbf{w}}}(t)}$$

where $\hat{\mathbf{V}}(t)$ is as defined before and $\hat{\tilde{\mathbf{w}}}(t) = \left(\frac{n_1}{n\hat{c}_1(t)}, \frac{n_2}{n\hat{c}_2(t)}, \ldots, \frac{n_k}{n\hat{c}_k(t)}\right)^T$. Under $\mathcal{H}_0$, $\mathcal{T}_{01}$ and $\hat{\mathcal{T}}_{01}$ have exactly the same asymptotic distribution which is given in (17). Moreover, the true asymptotic distribution of $\hat{\mathcal{T}}_{01}$ can again be approximated using

Table 2 Powers of the tests $\mathcal{T}_{01}$ and $\mathcal{S}_{01}$ when $k = 3$ under Lehman alternatives ($\bar{F}_i(x) = [\bar{F}_0(x)]^{a_i}, i = 1, 2, 3$), $u_1 = 0.1, u_2 = 0.9, \alpha = 0.05$

a_1	a_2	a_3	$\mathcal{T}_{01}$	$\mathcal{S}_{01}$	$\mathcal{T}_{01}$	$\mathcal{S}_{01}$	$\mathcal{T}_{01}$	$\mathcal{S}_{01}$	$\mathcal{T}_{01}$	$\mathcal{S}_{01}$
10% Censoring										
1.00	1.00	1.00	0.039	0.025	0.049	0.034	0.051	0.036	0.044	0.037
1.50	1.25	1.00	0.253	0.123	0.450	0.163	0.312	0.118	0.345	0.160
1.75	1.50	1.00	0.446	0.152	0.706	0.264	0.600	0.159	0.541	0.231
2.00	1.75	1.00	0.630	0.221	0.881	0.379	0.768	0.243	0.720	0.311
2.25	2.00	1.00	0.870	0.371	0.950	0.496	0.890	0.301	0.854	0.387
25% Censoring										
1.00	1.00	1.00	0.040	0.025	0.045	0.036	0.037	0.031	0.046	0.038
1.50	1.25	1.00	0.223	0.096	0.387	0.143	0.303	0.090	0.311	0.132
1.75	1.50	1.00	0.401	0.151	0.611	0.203	0.501	0.146	0.505	0.201
2.00	1.75	1.00	0.564	0.193	0.822	0.302	0.730	0.196	0.666	0.274
2.25	2.00	1.00	0.710	0.234	0.925	0.388	0.859	0.264	0.799	0.359

For columns 4 and 5, $n_1 = n_2 = n_3 = 30$, for columns 6 and 7, $n_1 = n_2 = n_3 = 50$, for columns 8 and 9, $n_1 = 50, n_2 = 40, n_3 = 30$ and for columns 10 and 11, $n_1 = 30, n_2 = 40, n_3 = 50$

Table 3 Powers of the tests $\mathcal{T}_{01}$ and $\mathcal{S}_{01}$ when $k = 4$ under Lehman alternatives ($\bar{F}_i(x) = [\bar{F}_0(x)]^{a_i}, i = 1, 2, 3, 4$), $u_1 = 0.1, u_2 = 0.9, \alpha = 0.05$

a_1	a_2	a_3	a_4	$\mathcal{T}_{01}$	$\mathcal{S}_{01}$	$\mathcal{T}_{01}$	$\mathcal{S}_{01}$
10% Censoring							
1.0	1.00	1.00	1.0	0.043	0.035	0.053	0.038
1.5	1.3	1.2	1.0	0.358	0.116	0.588	0.513
1.6	1.4	1.2	1.0	0.355	0.110	0.599	0.169
1.8	1.6	1.4	1.0	0.492	0.122	0.755	0.191
25% Censoring							
1.0	1.0	1.0	1.0	0.043	0.035	0.053	0.038
1.5	1.3	1.2	1.0	0.310	0.089	0.511	0.143
1.6	1.4	1.2	1.0	0.311	0.091	0.521	0.134
1.8	1.6	1.4	1.0	0.422	0.130	0.672	0.168

For columns 5 and 6, $n_1 = n_2 = n_3 = n_4 = 30$, for columns 7 and 8, $n_1 = n_2 = n_3 = n_4 = 50$

the approach in Parzen et al. (1997) by holding the data fixed and simulating the independent standard normal covariates Z_{ij}s in $\tilde{V}(t) = (\tilde{V}_1(t), \tilde{V}_2(t), \ldots, \tilde{V}_k(t))^T$ and computing $\tilde{\mathcal{T}}_{01} = \sup_{\tau_1 \le t \le \tau} ||E_{\hat{\mathbf{w}}(t)}[\tilde{\mathbf{V}}(t)|\mathcal{I}_1] - E_{\hat{\mathbf{w}}(t)}[\tilde{\mathbf{V}}(t)|\mathcal{I}_0]||^2_{\hat{\hat{\mathbf{w}}}(t)}$. To approximate the p value, we may calculate the percentage of the simulated values of $\tilde{\mathcal{T}}_{01}$ that are greater than the observed value of $\mathcal{T}_{01}$. In addition, to approximate the power of $\mathcal{T}_{01}$ at a given alternative, for each simulated data under this alternative, we compute the value of $\mathcal{T}_{01}$, then simulate the values of $\tilde{\mathcal{T}}_{01}$ by repeatedly generating the independent normal variates Z_{ij}s while holding the observed data fixed. We then compare appropriate quantiles of these values of $\tilde{\mathcal{T}}_{01}$ with the value of $\mathcal{T}_{01}$ for each one of the simulated data sets.

Table 4 Powers of the tests $\mathcal{T}_{01}$ and $\mathcal{S}_{01}$ when $k = 3$, $u_1 = 0.1$, $u_2 = 0.9$ and $\alpha = 0.05$

F_1	F_2	F_3	$\mathcal{T}_{01}$	$\mathcal{S}_{01}$	$\mathcal{T}_{01}$	$\mathcal{S}_{01}$
10% Censoring						
Exp(1)	Exp(1)	Exp(1)	0.039	0.025	0.044	0.031
Exp(1)	0.05 + Exp(1)	0.10 + Exp(1)	0.125	0.077	0.189	0.113
Exp(1)	0.10 + Exp(1)	0.20 + Exp(1)	0.314	0.250	0.524	0.423
Uni(0, 2)	Uni(0, 2)	Uni(0, 2)	0.041	0.035	0.048	0.043
Uni(0, 2)	Uni(0, 1.75)	Uni(0, 1.5)	0.117	0.058	0.139	0.077
Uni(0, 2)	Uni(0, 1.5)	Uni(0, 1)	0.505	0.193	0.749	0.284
25% Censoring						
Exp(1)	Exp(1)	Exp(1)	0.039	0.025	0.044	0.031
Exp(1)	0.05 + Exp(1)	0.10 + Exp(1)	0.101	0.079	0.171	0.155
Exp(1)	0.10 + Exp(1)	0.20 + Exp(1)	0.246	0.198	0.523	0.457
Uni(0, 2)	Uni(0, 2)	Uni(0, 2)	0.041	0.035	0.043	0.039
Uni(0, 2)	Uni(0, 1.75)	Uni(0, 1.5)	0.112	0.057	0.211	0.088
Uni(0, 2)	Uni(0, 1.5)	Uni(0, 1)	0.444	0.178	0.711	0.227

For columns 4 and 5, $n_1 = n_2 = n_3 = 30$, for columns 6 and 7, $n_1 = n_2 = n_3 = 50$

To implement the new test when the censoring distributions are the same, one can again fix u_1 and u_2 and use $\tau_i = \hat{c}^{-1}(u_i/(1-u_i)), i = 1, 2$, where $\hat{c}^{-1}(x) = \inf\{t, \hat{c}(t) \geq x\}$ and $\hat{c}(t) = \sum_{i=1}^{k} \frac{n_i}{n} \hat{c}_i(t)$ is again the pooled estimate of c. The estimated cutoff points corresponding to $u_1 = 0.1$ and $u_2 = 0.9$ can be found in Davidov and Herman (2010). Finally, when there is no censoring, $c_i(t) = F(t)/\bar{F}(t)$ where F is the common distribution under $\mathcal{H}_0$, and it can be easily verified that

$$\sup_{\tau_1 \leq t \leq \tau_2} [-2\log(\mathcal{R}_{01}(t))] \xrightarrow{d} \sup_{F(\tau_1) \leq u \leq F(\tau_2)} \sum_{i=1}^{k} \gamma_i \frac{\left(E_{\boldsymbol{\gamma}}[\tilde{\mathbf{B}}(u)|\mathcal{I}_1]_i - \sum_{j=1}^{k} \gamma_j^{1/2} B_j(u)\right)^2}{u(1-u)}$$

where $\tilde{\mathbf{B}}$ is defined in Theorem 3. This is the limiting distribution given in Davidov and Herman (2010) for their test. These authors showed that their test outperforms existing tests at all the distributions that they considered.

3 Algorithm, simulations and a numerical example

3.1 Algorithm

To compute $\mathcal{R}_{02}(t)$ and $\mathcal{R}_{01}(t)$, we need to compute $\hat{\boldsymbol{\lambda}}^{(0)}(t)$ and $\hat{\boldsymbol{\lambda}}^{(1)}(t)$. A careful inspection of the duality results in Dykstra and Feltz (1989) shows that $\hat{\boldsymbol{\lambda}}^{(0)}(t)$ is the solution to

$$\min \sum_{i=1}^{k} \sum_{j=1}^{n(t)} (n_{ij} + n(\lambda_i - \lambda_{i-1}))\Phi\left(\frac{d_{ij}}{n_{ij} + n(\lambda_i - \lambda_{i-1})}\right)$$

subject to $(\lambda_0, \lambda_1, \ldots, \lambda_k)^T \in \{0\} \times R^{k-1} \times \{0\}$ where $\Phi(x) = x\log(x) + (1 - x)\log(1 - x)$. Its value can be computed using their algorithm as follows.

1. Initially set $\lambda_i^0 = 0$ for all j; set $\nu = 1$.
2. Find λ_i^ν, the optimal value of λ_i with all the other λs held fixed. This value of λ_i replaces the previous value of λ_i.
3. If $i < k - 1$, set $i = i + 1$; if $i = k - 1$, set $i = 1$ and $\nu = \nu + 1$.

This process is continued until sufficient accuracy is attained. This same algorithm can be used to compute $\hat{\boldsymbol{\lambda}}^{(1)}$ with the additional constraint that when $\hat{\lambda}_i^\nu < 0$ is step 2, it is set equal to 0. A careful inspection of this algorithm shows that $\hat{\lambda}_i^\nu$ that we seek in step 2 is the solution to

$$\prod_{j=1}^{n(t)} \left(1 - \frac{d_j}{n_{ij} + n(\lambda_i - \lambda_{i-1}^{\nu+1})}\right) = \prod_{j=1}^{n(t)} \left(1 - \frac{d_{i+1,j}}{n_{i+1,j} + n(\lambda_{i+1}^{\nu-1} - \lambda_i)}\right).$$

3.2 Simulations

We now present the results of two simulation studies. In the first simulation study, we assume that the censoring distributions are the same and compare the performance of $\mathcal{T}_{01}$ with the test statistic $\mathcal{S}_{01}$ of El Barmi and Mukerjee (2005) whose limiting distribution is known only when the censoring distributions are the same. This statistic is defined as the maximum of a sequence of (one-sided) two-sample Kolmogorov–Smirnov test statistics. In each case, 3000 data sets were used to approximate the power using $u_1 = 0.1$ and $u_2 = 0.9$ and assuming a common censoring distribution. The parameters are chosen in such a way to produce 10% or 25% censoring when sampling from the smallest distribution. Throughout we take $\tau_i = \hat{c}^{-1}(u_i), i = 1, 2,$ where $\hat{c}$ is as defined before. The critical values for $\mathcal{T}_{01}$ are taken from Davidov and Herman (2010) and those for $\mathcal{S}_{01}$ are obtained from its asymptotic distribution which is available in a closed form and is given in El Barmi and Mukerjee (2005). First, we note that the new procedure is invariant with respect to monotone transformations. To see this, suppose X_i has distribution F_i and let $\tilde{X}_i = g(X_i)$ where g is an increasing function. Let $X_{ij}, C_{ij}, Z_{ij}, \delta_{ij}, \theta_{ij}$ and T_j be as defined before and let $(\tilde{X}_{ij}, \tilde{C}_{ij}) = (g(X_{ij}), g(C_{ij}))$, $\tilde{Z}_{ij} = \min(\tilde{X}_{ij}, \tilde{C}_{ij})$ and $\tilde{\delta}_{ij} = I[\tilde{X}_{ij} \leq \tilde{C}_{ij}]$. Since g is increasing $\tilde{\delta}_{ij} = \delta_{ij}$ and if we let $\tilde{T}_j = g(T_j)$, then

$$\psi_{ij} \equiv \frac{P(\tilde{X}_i > \tilde{T}_j)}{P(\tilde{X}_i > \tilde{T}_{j-1})} = \frac{P(X_i > T_j)}{P(X_i > T_{j-1})} \equiv \theta_{ij} \quad \text{and}$$

$$n(g(t)) \equiv \sum_{j=1}^{m} I[\tilde{T}_j \leq g(t)] = \sum_{j=1}^{m} I[T_j \leq t] = n(t).$$

Table 5 Powers of the test $\mathcal{T}_{01}$ and the test $\mathcal{L}_{01}$ in Liu et al. (1993) using $\bar{F}_i, i = 1, 2, 3$ in Fig. 1 when $\alpha = 0.05$

Censoring distributions (rates of censoring)			$\mathcal{T}_{01}$	$\mathcal{L}_{01}$	$\mathcal{T}_{01}$	$\mathcal{L}_{01}$
Exp (10%)	Exp (10%)	Exp (10%)	0.450	0.097	0.576	0.143
Exp (10%)	Exp (10%)	Exp (25%)	0.377	0.091	0.543	0.136
Exp (10%)	Exp (25%)	Exp (25%)	0.364	0.086	0.527	0.128
Exp (25%)	Exp (25%)	Exp (25%)	0.352	0.083	0.513	0.125
Unif (10%)	Unif (10%)	Unif (10%)	0.360	0.079	0.469	0.110
Expl (10%)	Exp (10%)	Unif (10%)	0.373	0.095	0.522	0.111

For columns 4 and 5, $n_1 = n_2 = n_3 = 30$, for columns 6 and 7, $n_1 = n_2 = n_3 = 50$

In addition, the size of the risk set and the number of complete observations from the ith population at $\tilde{T}_j$ are n_{ij} and d_{ij}, respectively. Consequently

$$\prod_{j=1}^{m} \theta_{ij}^{n_{ij}-d_{ij}} (1-\theta_{ij})^{d_{ij}} = \prod_{j=1}^{m} \psi_{ij}^{n_{ij}-d_{ij}} (1-\psi_{ij})^{d_{ij}}$$

and $\bar{F}_i(T_j) = \prod_{\ell=1}^{j} \theta_{ij} = \prod_{\ell=1}^{j} \psi_{ij} = \bar{\tilde{F}}_i(\tilde{T}_j)$ where $\bar{\tilde{F}}_i$ is the SF corresponding to $\tilde{X}_i$. Therefore, the new test procedure is invariant under monotone transformations.

In the first part of this simulation study, we confine attention to Lehmann alternatives and assume that $\bar{F}_i(t) = [\bar{F}_0(t)]^{a_i}$ and $\bar{G}_i(t) = [\bar{F}_0(t)]^{b_i}$ for some cdf F_0 where $a_1 \geq a_2 \geq \ldots, \geq a_k$ and $b_1, b_2, \ldots, b_k$ are chosen to achieve the desired censoring size. Taking $g(t) = F_0(t)$, the invariance property under monotone transformations implies that it suffices in this case to sample from $\tilde{F}_i(t) = [1-(1-t)^{a_i}]I(0,1)(t) + I[1,\infty)(t)$. The corresponding censoring distribution is given by $\tilde{G}_i(t) = [1-(1-t)^{b_i}]I(0,1)(t) + I[1,\infty)(t)$. We use this procedure to evaluate the finite sample performance of the new test. Tables 2 and 3 give the results for $k = 3$ and $k = 4$. In all cases, $\mathcal{T}_{01}$ has greater power than $\mathcal{S}_{01}$ and has better agreement with the nominal level of the test.

In the second part of this simulation, we look at a variety of distributions and sample sizes. The censoring distributions are chosen from the same family in a way to produce the desired censoring rates. The results are given in Table 4 and in all cases $\mathcal{T}_{01}$ has again greater power than $\mathcal{S}_{01}$ and has better agreement with the nominal level of the test.

A careful inspection of the limiting distribution of $\mathcal{T}_{01}$ shows that it is based on the projection of the vector $\mathbf{V}(t)$ on the pointed isotonic cone defined by the linearly ordered set generated by $\{\log \bar{F}_i(t) : 1 \leq i \leq k,\ t \in \mathbf{R}\}$ in $\mathbf{R}^{k-1}$ after modding out the linear subspace corresponding to $\mathcal{H}_0^t$. On the other hand, the statistic $\mathcal{S}_{01}$ is the maximum of the statistics $\{\mathcal{S}_{01i}\}$ for sequentially testing $\mathcal{H}_{0i} : \bar{F}_1 = \cdots = \bar{F}_i$ vs $\mathcal{H}_{1i} : \bar{F}_1 = \cdots = \bar{F}_{i-1} \leq \bar{F}_i$, for $2 \leq i \leq k$, based on Hogg's (1962) suggestion. It can be shown that this corresponds to replacing the projection of $\mathbf{V}(t)$ on the isotonic cone by the projection on a strictly larger cone that is an orthant in $\mathbf{R}^{k-1}$. Heuristically, for a given level of significance, this amounts to a smaller rejection region in most directions resulting in a lower power.

In the second simulation study, we do not assume that the censoring distributions are the same and compare the power of $\mathcal{T}_{01}$ to that of a test developed in Liu et al. (1993) which we denote by $\mathcal{L}_{01}$. We note that this test is based on the sum of two-sample weighted log-rank statistics. Since the one-sided weighted log-rank tests are designed to detect uniform stochastic ordering or hazard rate ordering, which is more restrictive that stochastic order, they can fail to detect stochastic ordering. This is clearly shown in this simulation in which we take the hazard rates to be $\lambda_1(t) = 0.3I(0 < t < 1) + 0.1I(1 < t \leq 2) + 0.2I(t > 2)$, $\lambda_2(t) = 0.2$ and $\lambda_3(t) = 0.1I(0 < t < 1) + 0.3I(1 < t \leq 2) + 0.2I(t > 2)$. Clearly, these hazard rates cross but their corresponding SFs, $\bar{F}_i$, $i = 1, 2, 3$, are stochastically ordered (see Fig. 1). We take the censoring distributions to be either exponential and /or uniform, and they are chosen to produce the desired censoring rates. The conditions $\bar{F}(\tau_1) < 1$ and $\pi(\tau_2) > 0$ suggest a data-driven rule for choosing τ_1 and τ_2: $\tau_1 = \max_{1\leq i\leq k} \inf\{t, \hat{\bar{F}}_i(t) < 1\}$ and $\tau_2 = \min_{1\leq i\leq k} \sup\{t, \hat{\pi}_i(t) = \hat{\bar{F}}_i(t)\hat{\bar{G}}_i(t) > 0\}$ where each distribution is estimated by its Kaplan–Meier estimator. This is what we use in this simulation. In each case, 3000 data sets were simulated from these distributions, and for each one of these data sets, the value of $\mathcal{T}_{01}$ is computed and 3000 values of $\tilde{\mathcal{T}}_{01}$ were obtained by the method described above. The estimated power of $\mathcal{T}_{01}$ is estimated by the fraction of data sets whose value of $\mathcal{T}_{01}$ is greater than the corresponding 95% percentile of values of $\tilde{\mathcal{T}}_{01}$ (i.e., we use $\alpha = 0.05$). The results are given in Table 5, and they clearly show that the new test outperforms the test of Liu et al. (1993).

There are three reasons for low power of the $\mathcal{L}_{01}$ test: (i) The statistic $\mathcal{L}_{01}$ tests for uniform stochastic ordering only. This ordering is a meager subset of the stochastic ordering cone. As a result, it is powerful against uniform stochastic alternatives, but has very little power against stochastic ordering alternatives that are far removed from uniform stochastic ordering as demonstrated by simulations in Table 5. (ii) Squaring the two-sample log-rank test statistic amounts to using a two-sided Z-test for a one-sided alternative and (iii) the last reason for loss of power is the replacement of the isotonic cone by a larger cone that is an orthant as described above.

3.3 Example

Next, we illustrate the results in this paper using Data Set II from Kalbfleisch and Prentice (1980). This data set consists of survival times for patients with carcinoma of the oropharynx and several covariates. These patients were diagnosed with squamous carcinoma of the oropharynx, and they were classified by the degree to which the regional lymph nodes were affected by this disease into four populations. Lymph node deterioration indicates the seriousness of the carcinoma. As a result, we would expect the four populations to be stochastically ordered. In practice, one can take $\tau_2 = 1065$, the largest observation at which all the Kaplan–Meier estimators of the each distribution and its corresponding censoring distribution are positive and $\tau_1 = 105$, the smallest time at which all the Kaplan–Meier estimators are less that one. In this case $\mathcal{T}_{01} = 16.32$. The estimated p-value when we do not assume that the censoring

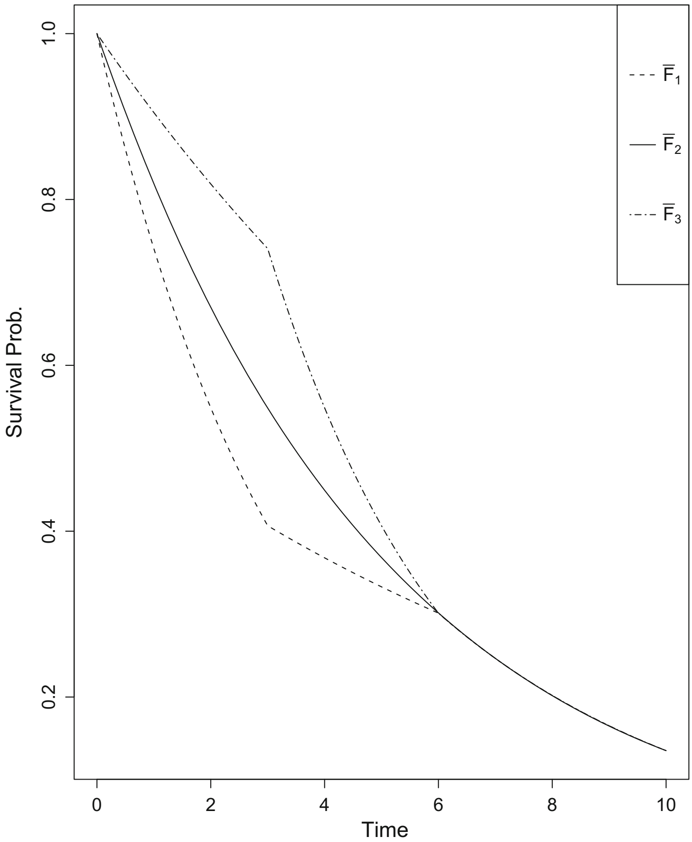

Fig. 1 Survival functions $\bar{F}_1$, $\bar{F}_2$ and $\bar{F}_3$

distributions are the same is 0.011 providing evidence that the distributions of the four populations are stochastically ordered. When we assume that the samples are censored by the same distribution, the p value is less that 0.01 based on Table 1 while the p value using El Barmi and Mukerjee (2005) is 0.024.

4 Concluding remarks

In this paper, we developed an empirical likelihood-based approach for testing the presence of stochastic ordering among k populations in the censored case. This new test is invariant under monotone transformations, and two simulation studies show that it is more powerful than a test developed for the same problem in El Barmi and Mukerjee (2005) and a test developed in Liu et al. (1993) at all the distributions that we considered. We also show that when all the samples are censored by the same distribution, its asymptotic distribution is exactly that of a test developed in the uncensored for the same problem in Davidov and Herman (2010). An advantage of this new test comes from the fact that it first considers the well-understood one-dimensional linear ordering problem for each t and then uses a supremum for an overall test. This method avoids the difficult problem of testing for the linear ordering at every t simultaneously. Previous tests used simplifying assumptions to achieve this, resulting in a loss of power.

We note that this new test extends naturally to testing whether $(F_1, F_2, \ldots, F_k)^T$ is isotonic with respect to a quasi-order on $\{1, 2, \ldots, k\}$. A relation $\preceq$ on $\{1, 2, \ldots, k\}$ is a *quasi-order* if it is reflexive and transitive and $(F_1, F_2, \ldots, F_k)^T$ is *isotonic* with respect to $\preceq$ if $F_i \preceq_{SO} F_j$ whenever $i \preceq j$. Examples of such ordered alternatives include tree ordering ($F_1 \preceq_{SO} F_i$, $i = 2, \ldots, k$) and umbrella ordering $F_1 \preceq_{SO} F_2 \cdots \preceq_{SO} F_{i_0}$ and $F_k \preceq_{SO} F_{k-1} \preceq_{SO} \cdots \preceq_{SO} F_{i_0+1}$, where i_0 is known. The localized empirical likelihood ratio can be computed by adapting the algorithm in Dykstra and Feltz (1989) to this situation. The limiting distribution of the resulting test statistic is obtained by taking $\mathcal{I}$ in (14) as the isotonic cone corresponding to $\preceq$.

A reviewer asked whether it is possible to compute Pitman efficiency for our test. Since Pitman efficiency is defined for test statistics that are asymptotically normal under a sequence of centering and scaling for a one-sided test, and the asymptotic distribution of the likelihood ratio test is a mixture of χ^2 distributions with different degrees of freedom that are derived from the asymptotic half-normal distributions of the projections of the observation on the isotonic cone. We have no idea how to even define a Pitman-type efficiency.

Acknowledgements The author is grateful to the editor, an associate editor and two referees for their insightful comments that have substantially improved the paper. The author is particularly grateful to the associate editor for many constructive comments and for pointing out the invariance property. He also wishes to thank Hari Mukerjee for his help and Zhiliang Ying for an interesting discussion on invariant tests and for suggesting to use Lehmann alternatives. This work was supported by a PSC-CUNY Grant.

References

Chang, H., El Barmi, H., McKeague, I. W. (2016). Tests for stochastic ordering under biased sampling. *Journal on Nonparametric Statistics*, *28*, 659–682.

Chang, H., McKeague, I. W. (2016). Empirical likelihood based tests for stochastic ordering under right censorship. *Electronic Journal of Statistics*, *10*, 2511–2536.

Davidov, O., Herman, A. (2010). Testing for order among K populations: Theory and practice. *The Canadian Journal of Statistics*, *38*, 97–115.

Dykstra, R. L., Feltz, C. J. (1989). Nonparametric maximum likelihood estimation of survival functions with general stochastic ordering and its dual. *Biometrika*, *76*, 331–341.

Dykstra, R. L., Madsen, R. W., Fairbanks, K. (1983). A nonparametric likelihood ratio test. *Journal of Statistical Computing and Simulation*, *18*, 247–264.

El Barmi, H. (2017). Testing for the uniform stochastic ordering via empirical likelihood under right censoring. *Statistica Sinica*, *27*, 645–664.

El Barmi, H., Johnson, M. S. (2006). A unified approach to testing for and against a set of linear inequality constraints in the product multinomial case. *Journal of Multivariate Analysis*, *97*, 1894–1912.

El Barmi, H., Kochar, S., Mukerjee, H. (2008). Order restricted inference for comparing the cumulative incidence of a competing risk over several populations. *IMS Collections Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, *1*, 50–61.

El Barmi, H., McKeague, I. W. (2013). Empirical likelihood based tests for stochastic ordering. *Bernoulli*, *19*, 295–307.

El Barmi, H., Mukerjee, H. (2005). Inference under a stochastic ordering constraint: The k-sample case. *Journal of the American Statistical Association*, *100*, 252–261.

Franck, W. E. (1984). A likelihood ratio test for stochastic ordering. *Journal of the American Statistical Association*, *80*, 686–691.

Hogg, R. V. (1962). Iterated tests of the equality of several distributions. *Journal of the American Statistical Association*, *57*, 579–585.

Kalbfleisch, J. D., Prentice, R. L. (1980). *The statistical analysis of failure time data*. New York: Wiley.

Kaplan, E. L., Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*, 457–481.

Kiefer, J. (1959). K-sample analogues of the Kolmogorov–Smirnov and Cramer–von Mises tests. *Annals of Mathematical Statistics*, *30*, 420–447.

Lee, Y. J., Wolf, D. A. (1976). A distribution free test for stochastic ordering. *Journal of the American Statistical Association*, *71*, 722–727.

Lehmann, E. L. (1955). Ordered families of distributions. *Annals of Mathematical Statistics*, *26*, 399–419.

Liu, P. Y., Green, S., Wolf, M., Cowrley, J. (1993). Testing against ordered alternative for censored survival data. *Journal of the American Statistical Association*, *88*, 153–160.

Mau, J. (1988). A generalization of a nonparametric test for stochastically ordered distributions to censored survival data. *Journal of the Royal Statistical Society, Series B*, *50*, 403–412.

Parzen, M. I., Wei, L. J., Ying, Z. (1997). Simultaneous confidence intervals for the difference of two survival functions. *Scandinavian Journal of Statistics*, *24*, 309–314.

Pepe, M. S., Fleming, T. R. (1989). Weighted Kaplan–Meier statistics: A class of distance tests for censored survival data. *Biometrics*, *45*, 497–507.

Præstgaard, T. J., Huang, J. (1996). Asymptotic theory for nonparametric estimation of survival curves under order restrictions. *Annals of Statistics*, *24*, 1679–1716.

Qin, J., Lawless, J. (1995). Estimating equations, empirical likelihood and constraints on parameters. *The Canadian Journal of Statistics*, *23*, 145–159.

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

Robertson, T., Wright, F. T., Dykstra, R. L. (1988). *Order restricted inferences*. New York: Wiley.

Shaked, M., Shanthikumar, G. J. (2007). *Stochastic orders*. New York: Springer.

Silvapulle, M. J., Sen, P. K. (2005). *Constrained statistical inference*. New York: Wiley.

Wang, Y. (1996). A likelihood ratio test against stochastic ordering in several populations. *Journal of the American Statistical Association*, *79*, 686–691.