Empirical likelihood tests for stochastic ordering based on censored and biased data

Hsin-wen Chang

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2014

ABSTRACT

Empirical likelihood tests for stochastic ordering based on censored and biased data

Hsin-wen Chang

In the classical two-sample comparison problem, it is often of interest to examine whether the distribution function is uniformly higher in one group than the other. This can be framed in terms of the notion of stochastic ordering. We consider testing for stochastic ordering based on two types of data: (1) right-censored and (2) size-biased data. We derive our procedures using the empirical likelihood method, and the proposed tests are based on maximally selected local empirical likelihood statistics. For (1), the proposed test is shown via a simulation study to have superior power to the commonly-used log-rank test under crossing-hazard alternatives. The approach is illustrated using data from a randomized clinical trial involving the treatment of severe alcoholic hepatitis. As for (2), simulations show that the proposed test outperforms the Wald test and the test overlooking size bias in all the cases considered. The approach is illustrated via a real data example of alcohol concentration in fatal driving accidents.

Table of Contents

List of Figures

List of Tables

Acknowledgments

First of all, I would like to express my deepest gratitude to my Ph.D. advisor, Dr. Ian W. McKeague, for his patient guidance and constant encouragement during the past few years. I am forever grateful for his invaluable advice on research, writing and career choices.

Second, I would like to thank my dissertation committee chair, Dr. Wei-Yann Tsai, for his continued professional support. I also owe thanks to other committee members, Dr. Ying Wei, Dr. Min Qian and Dr. Hammou El Barmi. Their constructive suggestions on this dissertation have been very useful.

I am grateful for having served as research assistant to the following two teams that provide me with funding: the team led by Dr. Shuang Wang and Dr. Frederica P. Perera at the Columbia Center for Children's Environmental Health; and the team led by Dr. John L. P. (Seamus) Thompson and Dr. Bruce Levin at the Statistical Analysis Center for Clinical Trials.

I would not have made it if not for the support of my wonderful classmates, Chih-Chi Hu and Xuezhou Mao, along with all other students, professors and staff in the Department of Biostatistics.

Last but not least, I thank my family for their emotional support, a necessity during my dissertation research.

To my family

Chapter 1

Introduction

The empirical likelihood (EL) method involves construction of a nonparametric likelihood ratio test for the parameter of interest. It utilizes a similar formation to the usual likelihood ratio test, except the optimization is with respect to a nonparametric likelihood instead (see Section 5.1 for more details). EL was first proposed by Thomas and Grunkemeier (1975) in establishing pointwise confidence interval for the survival function based on right-censored data. The method has been used to provide tests and confidence regions for parameters defined by estimating equations (Owen, 1988, 2001). The resulting confidence regions are range-preserving and transformation-respecting, and the tests have good power properties (see, e.g., Kitamura et al., 2012). Due to these favorable properties, there has been an increasing number of applications of EL in survival time data analysis and other areas of statistics in recent years.

In this dissertation, we study EL-based procedures for detecting an ordering between distribution functions. This can be framed in terms of the notion of *stochastic ordering*. Namely, a distribution function F_1 is said to be *stochastically larger* than another distribution function F_2 if $F_1(t) \leq F_2(t)$ for all $t \geq 0$; this is denoted as $F_1 \succeq F_2$. We consider the problem of testing the two-sided alternative

$$H_0 : F_1 = F_2 \text{ versus } H_1 : F_1 \succ F_2 \text{ or } F_2 \succ F_1$$

where $\succ$ denotes $\succeq$ with strict inequality for some t. Our approach will first be developed for testing the one-sided alternative

$$H_0 : F_1 = F_2 \text{ versus } H_1 : F_1 \succ F_2$$

and then extended to the two-sided alternative using the union-intersection principle. Such order-restricted inference incorporates a priori constraints into the construction of the parameter space and the hypotheses, thereby increasing efficiency (see, e.g., Silvapulle and Sen, 2005). A more detailed literature review is given in Section 1.3.

We develop EL-based tests for stochastic ordering in two different settings: right censorship and biased sampling models. We introduce the two topics as follows.

1.1 First topic

The first topic is motivated by the classical two-sample problem based on right-censored data. Specifically, when comparing a combination intervention of two drugs (or procedures) with one of the drugs (or procedures), the scientists may want to test whether the combination therapy results in a consistently higher/lower survival rate (throughout the follow-up period). Commonly used methods for such two-sample comparison include the log-rank test and weighted Kaplan–Meier (WKM) tests (Pepe and Fleming, 1989). However, the log-rank test is actually designed to detect an ordering between hazard functions instead of the more general stochastic ordering (and so is the more general class of weighted log-rank statistics), whereas the WKM statistics depend on an ad hoc weight function that needs to be specified throughout follow-up. Our proposed EL procedure targets stochastic ordering through the maximum likelihood paradigm, thereby providing a more powerful inference compared with the log-rank and WKM tests under crossing-hazard alternatives.

1.2 Second topic

Biased sampling models can be seen in numerous applications, such as in genetics (see, e.g., Fisher, 1934; Clark et al., 2005, ascertainment bias), animal studies (see, e.g., Patil and Rao, 1978, visibility bias), disease screening (Zelen and Feinleib, 1969; Duffy et al., 2008), and vaccine trials (Gilbert et al., 1999). In these models, samples from an underlying distribution function F is selected with probability proportional to a biasing or weight function $w(\cdot) \geq 0$. The resulting observed samples follow a biased version of F:

$$G(x) = \int_0^x \frac{w(u)}{W} dF(u),$$

where $W = \int_0^\infty w(u)\ dF(u) < \infty$ is the normalizing constant. Nonparametric inference about the unbiased distribution function has been studied extensively over the years (Vardi, 1982, 1985; Gill et al., 1988; Ramírez and Vidakovic, 2010), including EL-based procedures (Qin, 1993; El Barmi and Rothmann, 1998). However, the existing approaches do not consider more than one unbiased distribution for the groups.

For our second topic, we consider a two-sample setup where the biasing function and the unbiased distribution are allowed to differ between the samples. That is, samples from the j-th underlying distribution function F_j (for $j = 1, 2$) follow

$$G_j(x) = \int_0^x \frac{w_j(u)}{W_j} dF_j(u),$$

where $W_j = \int_0^\infty w_j(u)\ dF_j(u) < \infty$. Our contribution is to develop a nonparametric test for the presence of pointwise ordering between F_1 and F_2.

1.3 Literature review on stochastic ordering

Recent overviews of the literature on ordered restricted inference can be found in Silvapulle and Sen (2005) and Shaked and Shanthikumar (2007). For uncensored data, many nonparametric tests for stochastic ordering have been developed since the 1950s, such as the Jonckheere–Terpstra test (Terpstra, 1952; Jonckheere, 1954), likelihood ratio tests (Robertson and Wright, 1981; Dykstra et al., 1983; Franck, 1984; Wang, 1996), tests based on isotonic regression (Chacko, 1963; El Barmi and Mukerjee, 2005; Davidov and Herman, 2010) and on EL (El Barmi and McKeague, 2013). Some recent work involves summarizing localized test statistics via taking supremum (El Barmi and Mukerjee, 2005; Davidov and Herman, 2010) or integration (El Barmi and McKeague, 2013).

In right-censored data settings, two-sample tests for ordered alternatives often make use of the one-sided version of the omnibus tests, some common ones including the log-rank test (Mantel, 1966), Gehan's generalized Wilcoxon test (Gehan, 1965), and the Peto–Prentice–Wilcoxon test (Peto and Peto, 1972; Prentice, 1978). For k samples, one line of investigation is to form a trend test based on the pairwise two-sample statistics. This dates back to Gehan's Jonckheere–Terpstra-type test (Gehan, 1965) based on the two-sample generalized Wilcoxon test. Green (1979) showed that the same test can be rewritten in a form with simplified variance formula, which inspired subsequent development of the ordered weighted log-rank statistics (Crowley, 1979; Liu et al., 1993). Then Liu and Tsai (1999) pointed out some weaknesses of this class of tests and proposed a modified ordered log-rank test. Other classes of k-sample tests involve isotonic regression (Mau, 1988; El Barmi and Mukerjee, 2005) or the integrated weighted difference of Kaplan–Meier (KM) estimators (Pepe and Fleming, 1989, 1991; Chi, 2002).

In biased sampling setup, to our knowledge, nonparametric inference for stochastic ordering between the underlying distributions has not been properly addressed in the literature.

The dissertation is organized as follows. The first and second topics are investigated in Chapter 2 and 3, respectively. Tables, figures and appendices for each topic are provided at the end of each Chapter. A bibliography for both topics is provided in Chapter 4, and some supplementary materials are given in Chapter 5.

Chapter 2

Empirical likelihood based tests for stochastic ordering under right censorship

2.1 Introduction

When comparing survival patterns between two treatment groups in a randomized clinical trial (RCT), it is often of interest to examine whether there is a uniformly higher survival rate in one of the groups. For example, in a recent RCT involving patients with severe alcoholic hepatitis, the objective is to compare a combination therapy of prednisolone plus N-acetylcysteine with prednisolone alone. Testing whether the combination therapy has a consistently higher/lower survival rate (throughout the follow-up period) addresses the issue directly, as opposed to the standard practice of using an omnibus alternative (i.e., *any* difference between the survival functions). This chapter develops such a testing procedure that allows us to establish an ordering between two survival curves uniformly over time.

We frame our approach in terms of the classical notion of *stochastic ordering*. Namely, a survival function S_1 is said to be *stochastically larger* than another survival function S_2 if $S_1(t) \geq S_2(t)$ for all $t \geq 0$; this is denoted as $S_1 \succeq S_2$. We consider the problem of testing the two-sided alternative

$$H_0\colon S_1 = S_2 \text{ versus } H_1\colon S_1 \succ S_2 \text{ or } S_2 \succ S_1 \tag{2.1}$$

based on right-censored random samples from each population ($\succ$ denotes $\succeq$ with strict inequality for some t). Our approach will first be developed for testing the one-sided alternative

$$H_0\colon S_1 = S_2 \text{ versus } H_1\colon S_1 \succ S_2 \tag{2.2}$$

and then extended to the two-sided alternative using the union-intersection principle. Our approach also leads to a test for the null hypothesis of stochastic ordering ($S_1 \preceq S_2$ or $S_1 \succeq S_2$) versus the alternative of crossing survival functions.

Commonly used two-sample tests for censored data include the log-rank test and weighted Kaplan–Meier (WKM) tests (Pepe and Fleming, 1989), and these tests can be one-sided and two-sided. The log-rank test is based on an integrated weighted difference between hazard functions, and is thus designed to detect ordered hazards instead of more general stochastic ordering. Other tests based on weighted differences between hazard functions, such as the $\mathcal{K}$-class of weighted log-rank statistics (Gill, 1980), also share this property. The WKM class of tests targets stochastically ordered alternatives by estimating an integrated weighted difference between survival functions, but such test statistics depend on an ad hoc weight function that needs to be specified throughout follow-up.

We derive our procedure using the empirical likelihood (EL) method. EL involves forming a ratio of two nonparametric likelihoods subject to constraints on the parameters of interest. The method originates with Thomas and Grunkemeier (1975), who constructed pointwise confidence intervals for survival functions from right-censored data. EL has also been used to provide confidence regions for parameters defined by estimating equations (Owen, 1988, 2001), in numerous censored and uncensored settings. EL enjoys many appealing properties: highly accurate confidence regions, self-studentization and the possibility of Bartlett correctability. There is also evidence that EL-based tests have optimal power (see, e.g., Kitamura et al., 2012). On the other hand, order restricted inference is known to be challenging for EL (see, e.g., Owen, 2001, Ch. 10), and much less has been done in this direction. El Barmi (1996) explored EL tests for order-restricted hypotheses of the form $g(\theta) \leq 0$, where g is some smooth function and θ is a finite-dimensional parameter specified by estimating equations (see also Yu et al., 2011). Other recent contributions in this direction have been made by Andrews and Guggenberger (2009) and Canay (2010). As for order restrictions on distribution functions, El Barmi and McKeague (2013) studied EL-based tests for stochastic ordering, while Davidov et al. (2010) investigated EL-based tests for likelihood ratio ordering under a semiparametric biased sampling model. However, these tests are limited to uncensored data.

Our contribution is to provide a class of EL-based tests for stochastic ordering for right-censored data. First consider the one-sided alternative in (2.2). The idea is to construct a localized EL statistic for $H_0^t\colon S_1(t) = S_2(t)$ versus $H_1^t\colon S_1(t) > S_2(t)$ at each given t. The key step in this construction is to recast the stochastic ordering constraint into an inequality involving a single Lagrange multiplier. Then the proposed test rejects H_0 for large values of the maximally selected EL statistic. A maximally selected test statistic is used (as opposed to integral-type) because it is more sensitive to local differences between the survival functions. Kolmogorov–Smirnov type test statistics (not based on EL) for stochastic ordering have been proposed by El Barmi and Mukerjee (2005) and Davidov and Herman (2009). Besides localization, another possible approach might be to use the full nonparametric likelihood (Dykstra, 1982; Park et al., 2012a) and compute its ratio under $S_1 \succ S_2$ versus $S_1 = S_2$. However, we find the localization approach to be much more tractable. The localization approach has been used in Einmahl and McKeague (2003), Davidov and Herman (2012) and El Barmi and McKeague (2013) for testing various nonparametric hypotheses, except they considered an integral type test statistic and restricted attention to uncensored data. Park et al. (2012b) proposed a localized NPMLE under stochastic ordering (for right-censored data), but its asymptotic distribution is not known, so it is unclear how a formal test could be developed using their approach.

Various ways of formulating EL in right-censored data settings have been proposed. The standard approach for censored data (Thomas and Grunkemeier, 1975; Li, 1995) maximizes the censored data likelihood subject to contraint(s) on the parameter of interest. Wang and Jing (2001) instead used the nonparametric likelihood for uncensored data and plug-in of the Kaplan–Meier (KM) estimator of the censoring distribution. We use the former approach as it is tractable and more natural in our setting. There are in fact two different versions of EL for censored data, namely the binomial and Poisson versions (see, e.g., Murphy, 1995). We utilize the binomial version.

The chapter is organized as follows. In Section 2.2.1 we set up the general framework and notation to be used throughout the chapter. While our focus is on the two-sample test in Section 2.2.3, for clarity of exposition the one-sample test will be introduced first (in Section 2.2.2). Various extensions are discussed in Section 2.2.4: stochastic ordering in the null hypothesis, two-sided alternatives, and crossing survival functions. Section 2.3 presents results from a simulation study: the proposed two-sample EL test is shown to outperform the log-rank and WKM tests under different stochastically ordered alternatives, including alternatives with crossing hazards. Application of the proposed test to the randomized clinical trial (RCT) mentioned earlier is given in Section 2.4, and some concluding remarks are placed in Section 2.5.

2.2 EL tests for stochastic ordering under right censorship

2.2.1 Preliminaries

We begin by introducing notation for the one-sample case. Let X_i and C_i for $i = 1, \ldots, n$ be i.i.d. from unknown survival functions S and G, respectively; only $\min(X_i, C_i)$ and $I(X_i \le C_i)$ are observed. The lifetimes X_i and the censoring times C_i are assumed to be independent. Also, S is assumed to be continuous and $S(0) = G(0) = 1$. Order the uncensored lifetimes as $0 < T_1 < \ldots < T_m < \infty$. For each T_i $(i = 0, \ldots, m)$, let r_i be the number alive just before T_i, d_i be the number of deaths at T_i and h_i be the hazard at T_i. Let $N(t)$ be the number of observed lifetimes that are less than or equal to t. Then the nonparametric likelihood (depending on the unknown survival function) supported on the observed lifetimes is proportional to

$$L(S) \equiv \prod_{i=1}^{m} h_i^{d_i}(1 - h_i)^{r_i - d_i} \tag{2.3}$$

for $h_i \in [0, 1]$. The NPMLE for $S(t)$, namely the KM estimator $\hat{S}(t) = \prod_{i \le N(t)}(1 - d_i/r_i)$, is asymptotically normal with variance $S^2(t)\sigma^2(t)$, where $\sigma^2(t) = -\int_0^t dS(u)/\{S(u)S(u-)G(u-)\}$. This variance can be consistently estimated by the well-known Greenwood formula, $\hat{S}^2(t)\hat{\sigma}^2(t)/n$, where $\hat{\sigma}^2(t) = n\sum_{i \le N(t)}[d_i/\{r_i(r_i - d_i)\}]$.

For the two-sample case, we use a similar framework as in the one-sample setup with *a further subscript j indicating the j-th sample* in the corresponding notation. The nonparametric likelihood is proportional to $L(S_1, S_2) \equiv L(S_1)L(S_2)$. Additionally, the sample proportion n_j/n is assumed to converge to some $p_j > 0$, where $n = n_1 + n_2$. The $\hat{\sigma}^2(t)$ now equals the weighted average $n\{\hat{\sigma}_1^2(t)/n_1 + \hat{\sigma}_2^2(t)/n_2\}$, consistently estimating $\sigma_1^2(t)/p_1 + \sigma_2^2(t)/p_2$.

2.2.2 One-sample case

Suppose we wish to compare the survival function S with a given survival function S_0 for evidence of stochastic ordering. Formally, consider testing the null hypothesis $H_0\colon S = S_0$ versus the alternative hypothesis $H_1\colon S \succ S_0$. Our procedure is to first construct the test statistic for testing the "local" hypotheses $H_0^t\colon S(t) = S_0(t)$ versus $H_1^t\colon S(t) > S_0(t)$ for a given t, and then to deal with the general hypothesis based on some functional of the local statistics.

To construct the local test statistic at time t, consider the EL ratio

$$\mathcal{R}(t) = \frac{\sup\{L(S)\colon S(t) = S_0(t)\}}{\sup\{L(S)\colon S(t) \geq S_0(t)\}},$$

where we use the conventions $\sup\emptyset = 0$ and $0/0 = 1$. This follows the formulation of Thomas and Grunkemeier (1975) except with a one-sided alternative. Note that the numerator and denominator of $\mathcal{R}(t)$ maximize (2.3) over $(h_1, \ldots, h_m) \in [0,1]^m$ subject to the constraints

$$\prod_{i \leq N(t)} (1 - h_i) = S_0(t) \text{ or } \geq S_0(t), \tag{2.4}$$

respectively. We solve this constrained maximization problem using the Karush–Kuhn–Tucker (KKT) method (Boyd and Vandenberghe, 2004), a generalization of the Lagrange method that allows inequality constraints. As the constraints are placed only on the lifetimes up to t, the terms after t turn out to be the same in both the numerator and denominator and thus cancel out. Also, for some t the maximum is attained on the boundary of the constraint set, in which case $\mathcal{R}(t) = 1$. Specifically, in Appendix 2.6.1 we establish the following expression for the EL ratio:

$$\mathcal{R}(t) = \begin{cases} 1, & \hat{\lambda} \geq 0, \\ \displaystyle\prod_{i \leq N(t)} \frac{\hat{h}_i^{d_i}\left(1 - \hat{h}_i\right)^{r_i - d_i}}{\bar{h}_i^{d_i}\left(1 - \bar{h}_i\right)^{r_i - d_i}}, & \hat{\lambda} < 0, \end{cases}$$

where $\bar{h}_i = d_i/r_i$, $\hat{h}_i = d_i/(r_i + \hat{\lambda})$, and the Lagrange multiplier $\hat{\lambda} > D = \max_{i \leq N(t)}(d_i -$

r_i) is determined by the equality in (2.4) when h_i is replaced with $\hat{h}_i$. Here we have suppressed the dependence of $\hat{\lambda}$ and $\hat{h}_i$ on t.

Based on the above expression, we can derive large sample properties of the local EL test statistic, $-2\log\mathcal{R}(t)$. This is done by approximating $-2\log\mathcal{R}(t)$ via a Taylor expansion as a function of the difference between $\log\hat{S}(t)$ (recall from Section 2.2.1 that $\hat{S}(t)$ is the KM estimator) and $\log S_0(t)$. We then make use of asymptotic properties of $\hat{S}(t)$ to establish the weak convergence of $-2\log\mathcal{R}(t)$. The asymptotic null distribution turns out to be chi-bar square. Namely, for t such that $0 < S_0(t) < 1$ and $G(t) > 0$,

$$-2\log\mathcal{R}(t)\xrightarrow{d}Z_+^2$$

under H_0^t, where $Z \sim N(0,1)$ and $Z_+ = \max(Z,0)$. This result can be used to test the local hypotheses H_0^t versus H_1^t.

To test for the alternative of stochastic ordering, we propose the following maximally selected EL statistic:

$$K_n = \sup_{t\in[t_1,t_2]} \{-2\log\mathcal{R}(t)\}, \tag{2.5}$$

where $0 < t_1 < t_2 < \infty$ are to be specified. We suppress the dependence of K_n on t_1 and t_2. Guidance on the choice of $[t_1, t_2]$ is provided later.

Our first result gives the asymptotic null distribution of K_n. The proof is omitted, because it is similar to the two-sample case (presented in Appendix 2.6.2).

Theorem 2.1. *Suppose S_0 is continuous. Then under H_0, for t_1 and t_2 satisfying $S_0(t_1) < 1$ and $S_0(t_2)G(t_2) > 0$,*

$$K_n\xrightarrow{d}\sup_{x\in[x_1,x_2]}\left\{\frac{B_+^2(x)}{x(1-x)}\right\},$$

where B is a standard Brownian bridge on $[0,1]$, $B_+ = \max(B,0)$, $x_j = b(t_j)$ for $j = 1,2$, and $b(t) = \sigma^2(t)/\{1+\sigma^2(t)\}$.

To implement the test, we pre-specify one of the intervals $[t_1, t_2]$ or $[x_1, x_2] = [b(t_1), b(t_2)]$ and determine the other via $b(t)$ or $b^{-1}(x) = \inf\{t : b(t) \geq x\}$. However,

b is unknown, so one of the two intervals has to be estimated. If we fix $[t_1, t_2]$ and estimate $[x_1, x_2]$ (by $[\hat{x}_1, \hat{x}_2]$ say), then we cannot tabulate critical values in advance, because $[\hat{x}_1, \hat{x}_2]$ varies across different data sets. On the other hand, pre-determining $[x_1, x_2]$ allows "universal" critical values, and this is the approach we take. Both the choice of $[x_1, x_2]$ and details of implementation will be provided in the next subsection.

2.2.2.1 Calibrating the test

This section discusses issues in calibrating the test. The first one is the choice of $[x_1, x_2]$. Secondly, having chosen $[x_1, x_2]$, we explain how to estimate $[t_1, t_2]$ and implement the proposed EL test. Justification for this calibration procedure will be provided for the two-sample case in Appendix 2.6.3 (the justification is similar for the one-sample case), where a statistic K_n^* is defined for K_n with estimated $[t_1, t_2]$. Critical values for the test are then obtained via simulation in Section 2.3.

The choice of $[x_1, x_2]$ is important because the interval width can affect power of the EL test. In a similar context, this issue has been discussed by Davidov and Herman (2009); they proposed a (non-EL-based) test of stochastic ordering for uncensored data via localization, and point out that a narrower $[x_1, x_2]$ gives smaller critical values, but may fail to capture deviations (from H_0) outside the interval. Our simulation study (in Section 2.3) shows that the choice $x_1 = 0.2$ and $x_2 = 0.98$ performs well in terms of balancing power and accuracy, and this is what we recommend in practice.

Having specified $[x_1, x_2]$, we need to estimate $[t_1, t_2]$. Under suitable conditions on b^{-1}, t_l can be consistently estimated by $\hat{b}^{-1}(x_l) = \inf\{t : \hat{b}(t) \geq x_l\}$ for $l = 1, 2$, where

$$\hat{b}(t) = \frac{\hat{\sigma}^2(t)}{1 + \hat{\sigma}^2(t)}$$

is a consistent estimator of $b(t)$. We can then compute K_n^* accordingly, based on the estimates $\hat{t}_1$ and $\hat{t}_2$. To ensure stability of K_n^* in small samples, we further modify $[\hat{t}_1, \hat{t}_2]$ so that values of $-2 \log \mathcal{R}(t)$ outside the interval $[T_1, T_m]$ (recall from Section 2.2.1 that these are the smallest and largest observed lifetimes) are discarded.

Note that this modification makes no difference asymptotically, since $[T_1, T_m] \supset [\hat{b}^{-1}(x_1), \hat{b}^{-1}(x_2)]$ eventually.

2.2.3 Two-sample case

We now adapt our apporach to the two-sample case. The "local" hypotheses are $H_0^t\colon S_1(t) = S_2(t)$ versus $H_1^t\colon S_1(t) > S_2(t)$ for given t. The local EL ratio at time t is defined to be

$$\mathcal{R}(t) = \frac{\sup\{L(S_1, S_2)\colon S_1(t) = S_2(t)\}}{\sup\{L(S_1, S_2)\colon S_1(t) \geq S_2(t)\}}. \tag{2.6}$$

The numerator and denominator optimize $L(S_1)L(S_2)$ subject to the constraints on $\prod_{i \leq N(t)}(1 - h_i)$ for each sample. As before, an explicit form of the EL ratio can be obtained via the Lagrange method (see Appendix 2.6.1 for more details):

$$\mathcal{R}(t) = \begin{cases} 1, & \hat{\lambda} \geq 0, \\ \displaystyle\prod_{j=1}^{2} \prod_{i \leq N_j(t)} \frac{\hat{h}_{ij}^{d_{ij}} \left(1 - \hat{h}_{ij}\right)^{r_{ij} - d_{ij}}}{\bar{h}_{ij}^{d_{ij}} \left(1 - \bar{h}_{ij}\right)^{r_{ij} - d_{ij}}}, & \hat{\lambda} < 0, \end{cases} \tag{2.7}$$

where $\bar{h}_{ij} = d_{ij}/r_{ij}$, and $\hat{\lambda}$ and $\hat{h}_{ij}$ are given in Appendix 2.6.1. The local EL test statistic $-2 \log \mathcal{R}(t)$ is shown to converge in distribution to chi-bar square under H_0^t, a direct consequence of (2.18) in the proof of the next Theorem.

To test H_0 vs. H_1, we propose the maximally selected EL statistic K_n as in (2.5), except $\mathcal{R}(t)$ is now given in (2.7). The following result gives the asymptotic null distribution of K_n (see Appendix 2.6.2 for the proof).

Theorem 2.2. *Suppose H_0 holds and the common survival function S_0 is continuous. For t_1 and t_2 satisfying $S_0(t_1) < 1$ and $S_0(t_2)G_j(t_2) > 0$ for $j = 1, 2$,*

$$K_n \xrightarrow{d} \sup_{x \in [x_1, x_2]} \left\{ \frac{B_+^2(x)}{x(1 - x)} \right\},$$

where B is a standard Brownian bridge on $[0, 1]$, $B_+ = \max(B, 0)$, $x_j = b(t_j)$ for $j = 1, 2$, and $b(t) = \sigma^2(t)/\{1 + \sigma^2(t)\}$.

As in the one-sample case, we pre-specify $[x_1, x_2]$ and estimate $[t_1, t_2]$ when implementing the test. Justification for this calibration procedure will be provided in Appendix 2.6.3, where a statistic K_n^* is defined for K_n with estimated $[t_1, t_2]$. Issues discussed in Section 2.2.2.1 carry over as well.

2.2.4 Extensions

2.2.4.1 Stochastically ordered null

We have developed our test for the null hypothesis $S_1 = S_2$. Here we describe how our approach can be extended to the stochastically ordered null hypothesis $S_1 \preceq S_2$. The local EL ratio is

$$\mathcal{R}'(t) = \frac{\sup\{L(S_1, S_2)\colon S_1(t) \le S_2(t)\}}{\sup\{L(S_1, S_2)\}},$$

where the denominator maximizes over the union of the local (null and alternative) hypotheses and results in no constraint on $S_1(t)$ and $S_2(t)$. Since the KM estimator is the NPMLE, if $\hat{S}_1(t) \le \hat{S}_2(t)$ the numerator of $\mathcal{R}'(t)$ coincides with the unconstrained maximum and thus equals the denominator. If $\hat{S}_1(t) > \hat{S}_2(t)$, it can be shown that the numerator attains its maximum on the boundary $S_1(t) = S_2(t)$ of the constraint set (using log-concavity of (2.3)). We then have

$$\mathcal{R}'(t) = \begin{cases} 1, & \hat{S}_1(t) \le \hat{S}_2(t), \\ \dfrac{\sup\{L(S_1, S_2)\colon S_1(t) = S_2(t)\}}{\sup\{L(S_1, S_2)\}}, & \hat{S}_1(t) > \hat{S}_2(t). \end{cases}$$

Thus $\mathcal{R}'(t) = \mathcal{R}(t)$ by (2.7), since $\hat{\lambda} \ge 0$ is the same as $\hat{S}_1(t) \le \hat{S}_2(t)$ by Appendix 2.6.1. Hence K_n does not change under this broader null hypothesis. The same calibration method can be used because $S_1 = S_2$ is least favorable. The test is consistent in the "interior" of the stochastically ordered null — when $S_1(t) < S_2(t)$ for all t we have that K_n tends to zero in probability (since the indicator term in Lemma 2.3 vanishes for all $t \in [t_1, t_2]$ with probability tending to 1).

2.2.4.2 Two-sided testing

The one-sided tests introduced in the previous sections have immediate extensions to two-sided versions. The two-sided alternative in (2.1) is the union of the two one-sided alternatives ($S_1 \succ S_2$ or $S_2 \succ S_1$). Based on the union-intersection principle, the test statistic is the maximum of the two one-sided test statistics. The asymptotic null distribution of this test statistic is $\sup_{x\in[x_1,x_2]}[B^2(x)/\{x(1-x)\}]$, where B is a standard Brownian bridge, as in Theorem 2.2. The test can therefore be calibrated in much the same way as we did for the one-sided test.

2.2.4.3 Crossing survival functions

The possibility of crossing survival functions needs to be excluded for our (one-sided) test to be meaningful. This is because the one-sided test (asymptotically) rejects the null hypothesis if $S_1(t) > S_2(t)$ for some t and $S_1(t') < S_2(t')$ for some other t'. To address this issue, we recommend carrying out the one-sided test in *both* possible directions. If both tests reject, then there is evidence of crossing survival functions, excluding stochastic ordering. If only one of the tests rejects, then the interpretation is that there is evidence of stochastic ordering in that specific direction.

A formal test for crossing survival functions (against the null $S_1 \preceq S_2$ or $S_1 \succeq S_2$) can be devised using the intersection-union principle, taking the minimum of the two one-sided test statistics as the test statistic. The R code (see Section 5.2.2) for implementing the one-sided test is readily adapted for this purpose, with critical values obtained from simulating

$$\min\left[\sup_{x\in[x_1,x_2]}\left\{\frac{B_-^2(x)}{x(1-x)}\right\}, \sup_{x\in[x_1,x_2]}\left\{\frac{B_+^2(x)}{x(1-x)}\right\}\right],$$

where B_- is the negative part of the Brownian bridge B.

2.3 Simulation study

In this section, we report the results of a simulation study for the two-sample case. We restrict our attention to one-sided tests, but results for the two-sided tests are similar. We first tabulate selected critical values, and then compare the performance of K_n^* with the (one-sided) log-rank and WKM tests in terms of accuracy and power.

2.3.1 Critical values and accuracy

Quantiles of the limiting distribution in Lemma 2.4 of Appendix 2.6.3 are used as critical values for K_n^*. These are computed by simulation based on 100,000 replications of standard Brownian bridge over a fine grid on $[0, 1]$ (100,000 equidistant points), for selected values of x_1 and x_2 (see Table 2.1).

[Table 2.1 here]

To compute empirical significance levels, we simulate lifetimes from the piecewise exponential distribution displayed as solid line in upper left panel of Figure 2.1. We consider exponential censoring distribution: $G_1 = G_2 = \text{Exp}(\lambda)$, where λ is chosen to give a censoring rate (CR) of 10% or 25%. Our one-sided EL statistic (K_n^*) is compared with the one-sided log-rank statistic. Another class of tests for comparison is the one-sided WKM, and we follow recomendations of Pepe and Fleming (1989) and select the WKM statistic with the pooled variance estimator and the weight function denoted by $\hat{w}_c(t)$ in their paper.

Results on the size of our EL test are given in Table 2.2, where we use $[x_1, x_2] = [0.2, 0.98]$. The test is slightly conservative in small samples but approaches the nominal level as the sample size increases. Such conservativeness has been seen in other maximal deviation-type statistics for stochastic ordering (Davidov and Herman, 2009). The empirical significance levels of the one-sided log-rank test and the WKM test under the same settings are closer to the nominal level, but sometimes on the anticonservative side.

[Table 2.2 here]

2.3.2 Power comparisons

In this section, we compare the small sample power of the proposed test with the one-sided WKM and log-rank tests. Two models of lifetime distributions are considered, both with piecewise-constant hazards. In Model A, the hazard functions cross while the survival functions still remain stochastically ordered (see Figure 2.1, first column). In this case, the one-sided log-rank test can fail to detect the difference between the survival curves because it is designed to detect ordered hazards. In Model B, the two groups have different hazards initially but the same hazard later on, so the difference between the survival functions gradually wears off (see Figure 2.1, second column). This is a common phenomenon which is also seen in our real data example in Section 2.4. For both models, we consider exponential and uniform censoring distributions: $G_1 = G_2 = \text{Exp}(\lambda_1)$ or $\text{Uniform}(0, c_1)$, with λ_1 or c_1 chosen to give a CR of 10% or 25% for group 1.

[Figure 2.1 here]

Results are given in Table 2.3 for K_n^* using $[x_1, x_2]$=$[0.2, 0.98]$. Note that K_n^* outperforms the other tests in all the cases considered, especially in the crossing hazards scenario (Model A). The much lower power of WKM in Model A is surprising, because this test were shown to work well under crossing hazard alternatives in some previous simulation examples (Pepe and Fleming, 1989). The superior performance of our test may be due to two factors: first, our test is based on nonparametric likelihood, so it can be expected to be more powerful than tests that depend on an ad hoc weight function; second, we are using a maximal deviation-type statistic, rather than a weighted average, so our test may be more sensitive to local differences in the survival functions.

We have also investigated power under proportional hazards configurations, and our test closely matches the performance of the log-rank and WKM tests (results available upon request). These results show that for stochastically ordered alternatives, the proposed EL test can compete effectively with the log-rank and WKM tests, especially when the hazard functions cross.

[Table 2.3 here]

Table 2.4 gives size and power for various choices of x_1 and x_2 reflecting light or heavy truncation. It is clear from the last two rows that light truncation on the left results in both poor accuracy and power compared with the top row, which corresponds to our recommendation $[x_1, x_2]$=$[0.2, 0.98]$. Yet the performance is not very sensitive to the choice of x_2, so our preference is to choose x_2 close to 1 in order to reduce truncation.

[Table 2.4 here]

2.4 Application

A RCT for treatment of severe alcoholic hepatitis (Nguyen-Khac et al., 2011) is analyzed. The data are obtained by digitizing the published KM curves and reconstructing survival and censoring information using the algorithm developed by Guyot et al. (2012). The purpose of the trial was to assess whether a combination therapy of prednisolone plus *N*-acetylcysteine is better than prednisolone alone (the currently recommended treatment). A total of 174 patients were randomized to taking the combination ($n_1 = 85$) or only prednisolone ($n_2 = 89$), and the primary endpoint is their 6-month survival. The KM curves (see the top panel of Figure 2.2) suggest a stochastic ordering between the two groups.

[Figure 2.2 here]

Application of the one-sided EL test indicates that the combination therapy group has stochastically larger survival pattern than patients receiving only prednisolone ($K_n^* = 10.36$, $p = 0.018$). In comparison, the WKM and the one-sided log-rank tests yield p-values of 0.021 and 0.037, respectively. Examining the cumulative hazards plot (see the bottom panel of Figure 2.2), we can see that the slopes (i.e. hazards) of the two curves only differ noticeably during the initial 40 days. Such a scenario of an initial hazard difference has been considered in Model B of Section 2.3.2, where we show our EL test is better adapted to detecting a difference between the two treatment groups.

Nguyen-Khac et al. (2011) actually used the two-sided log-rank test and reported a p-value of 0.07. They concluded that the combination therapy does not improve the 6-month survival. In contrast, our two-sided EL test shows that the two treatment groups are significantly different and there is a uniformly higher survival rate in one of the groups ($p = 0.036$, computed by the supplementary R program that implements the two-sided EL test). In this case the EL test shows a more significant result that leads to a completely different conclusion than the log-rank test.

2.5 Discussion

We have developed a class of EL-based tests for both one- and two-sided stochastically ordered alternatives under right censoring. The proposed test statistic for one-sided alternatives is the maximally selected local EL statistic and is asymptotically distribution-free. The test statistic for two-sided alternatives is taken as the maximum of the two one-sided test statistics. A simulation study shows that our test can be much more powerful than the log-rank and WKM tests under alternatives with crossing hazards. We applied our test to a RCT involving patients with severe alcoholic hepatitis and found a more significant result than the log-rank and WKM tests.

Our test statistics utilize a data-dependent interval $[t_1, t_2]$, much like the data-dependent weight-function used in integral-type tests based on hazard or survival functions. Due to instability in the tails (caused by right-censoring), test statistics based on right-censored data invariably require such data-dependent tuning, and this feature cannot be avoided as far as we know. We could specify t_1 and t_2 in advance, but that would be inadvisable because of the instability in the test statistic arising when there are too few uncensored survival times outside the interval. However, in contrast to methods that rely on the selection of a complete weight function throughout follow-up (e.g., the WKM test), it is actually much easier and more transparent to select just the two tuning parameters (x_1 and x_2) needed in our case. Although t_1 and t_2 could be specified using a data-dependent rule (such as 5% of the data in each tail), this approach would have the disadvantage of needing tailor-made critical values for each dataset.

Our test targets stochastically ordered alternatives through construction of a nonparametric likelihood ratio (EL). It can be expected to be more powerful than commonly used two-sample tests that either are not tailored for such alternatives or depend on an ad hoc weight function. Moreover, it provides more information about the nature of the difference between S_1 and S_2 compared to the omnibus alternative $S_1 \neq S_2$, in which case the functional parameters S_1 and S_2 may be ordered in one direction at certain time points, but ordered in the reverse direction at other time points. Our test can also be used to detect crossing survival functions by applying it in *both* possible directions.

Our central contribution is the development of the first EL-based test for ordered survival functions in right-censored data settings, and we envision the test to be useful in clinical trials, in reliability engineering, and health policy applications. It would also be of interest to extend our approach to allow the testing of stochastic ordering in k-sample censored data settings, and to explore how it could be used for other types of ordering between distributions, such as increasing convex ordering, likelihood ratio ordering and uniform stochastic ordering (or hazard rate ordering).

Supplementary material

R programs implementing the procedures developed in this chapter are available in Section 5.2.2.

Acknowledgements

This chapter is joint work with Ian W. McKeague, who was partially supported by NIH Grant R01GM095722-01 and NSF Grant DMS-1307838. Computing resources for this chapter came from the Extreme Science and Engineering Discovery Environment (XSEDE) supported by NSF Grant OCI-1053575. The author thanks Hammou El Barmi and referees for numerous helpful comments.

2.6 Appendices

2.6.1 Derivation of the local EL statistic

We derive the local EL ratio (2.7) for the two-sample case. The one-sample case is similar and the proof is omitted.

First, we will obtain a closed-form expression for the denominator of (2.6) by the KKT method. After a log transformation, the optimization problem becomes minimizing

$$-\sum_{j=1}^{2}\sum_{i=1}^{m_j}\{d_{ij}(\log h_{ij}) + (r_{ij} - d_{ij})\log(1 - h_{ij})\}$$

over $(h_{11}, \dots, h_{m_1 1}, h_{12}, \dots, h_{m_2 2}) \in [0,1]^m$ ($m = m_1 + m_2$) subject to the constraints

$$\sum_{i \le N_2(t)} \log(1 - h_{i2}) - \sum_{i \le N_1(t)} \log(1 - h_{i1}) \le 0.$$

Since the domain $[0,1]^m$ is convex, the objective and constraint functions are convex and differentiable, and Slater's condition is satisfied, the KKT conditions are necessary and sufficient for optimality. More specifically, the Lagrangian is defined as a function $\mathcal{L} : [0,1]^m \times \mathbb{R} \to \mathbb{R}$ such that

$$\begin{aligned}
&\mathcal{L}(h_1, \ldots, h_m, \lambda) \\
&\equiv -\sum_{j=1}^{2} \sum_{i=1}^{m_j} \{d_{ij}(\log h_{ij}) + (r_{ij} - d_{ij}) \log(1 - h_{ij})\} \\
&\quad + \lambda \left\{ \sum_{i \le N_2(t)} \log(1 - h_{i2}) - \sum_{i \le N_1(t)} \log(1 - h_{i1}) \right\}.
\end{aligned}$$

The optimal solution is denoted as $(\hat{h}^1_{11}, \ldots, \hat{h}^1_{m_1 1}, \hat{h}^1_{12}, \ldots, \hat{h}^1_{m_2 2}, \hat{\lambda}^1)$, with the superscript indicating the correspondence of the denominator with H_1. The dependence of the solution on t is omitted here for simplicity but will appear in the proof of Theorem 2.2 (see Appendix 2.6.2) when the EL ratio is considered as a process indexed by t. The optimal solution must satisfy the KKT conditions:

$$\nabla_h \mathcal{L}(\hat{h}_{11}, \ldots, \hat{h}_{m_1 1}, \hat{h}_{12}, \ldots, \hat{h}_{m_2 2}, \hat{\lambda}^1) = 0, \tag{2.8}$$

$$\prod_{i \le N_1(t)} \left(1 - \hat{h}^1_{i1}\right) \ge \prod_{i \le N_2(t)} \left(1 - \hat{h}^1_{i2}\right), \tag{2.9}$$

$$\hat{\lambda}^1 \ge 0, \tag{2.10}$$

$$\hat{\lambda}^1 \left\{ \sum_{i \le N_2(t)} \log(1 - \hat{h}^1_{i2}) - \sum_{i \le N_1(t)} \log(1 - \hat{h}^1_{i1}) \right\} = 0, \tag{2.11}$$

which are known as stationarity, primal feasibility, dual feasibility, and complementary slackness, respectively. The stationarity condition yields $\hat{h}^1_{ij} = d_{ij}/r_{ij}$ for $i = N_j(t) + 1, \ldots, m_j$ and

$$\hat{h}^1_{ij} = \frac{d_{ij}}{r_{ij} + (-1)^{j-1} \hat{\lambda}^1}$$

for $i = 1, \ldots, N_j(t)$, for each $j = 1, 2$. Define $D_j = \max_{i=1,\ldots,N_j(t)} (d_{ij} - r_{ij})$. Since $(\hat{h}^1_{11}, \ldots, \hat{h}^1_{m_1 1}, \hat{h}^1_{12}, \ldots, \hat{h}^1_{m_2 2})$ should be in the domain $[0,1]^m$, we have that $D_1 \le \hat{\lambda}^1 \le -D_2$, where $D_j \le 0$ for $j = 1, 2$.

The numerator of $\mathcal{R}(t)$ can be handled in a similar fashion. Denoting the optimal solution to the Lagrangian by $(\hat{h}_{11}^0, \ldots, \hat{h}_{m_1 1}^0, \hat{h}_{12}^0, \ldots, \hat{h}_{m_2 2}^0, \hat{\lambda}^0)$, it turns out $\hat{h}_{ij}^0$ has the same form as $\hat{h}_{i,j}^1$ but with $\hat{\lambda}^1$ replaced by $\hat{\lambda}^0$, and $\hat{\lambda}^0$ only needs to satisfy $D_1 \leq \hat{\lambda}^0 \leq -D_2$ and

$$\prod_{i \leq N_1(t)} \left(1 - \hat{h}_{i1}^0\right) = \prod_{i \leq N_2(t)} \left(1 - \hat{h}_{i2}^0\right). \tag{2.12}$$

Note that the estimated hazards after time t under no constraints, namely $\hat{h}_{ij}^v$ for $v = 0, 1$ and $i = N_j(t) + 1, \ldots, m_j$, are the same in the numerator and denominator, and so these terms cancel out. This leads to

$$\mathcal{R}(t) = \prod_{j=1}^{2} \prod_{i \leq N_j(t)} \frac{\left(\hat{h}_{ij}^0\right)^{d_{ij}} \left(1 - \hat{h}_{ij}^0\right)^{r_{ij} - d_{ij}}}{\left(\hat{h}_{ij}^1\right)^{d_{ij}} \left(1 - \hat{h}_{ij}^1\right)^{r_{ij} - d_{ij}}}. \tag{2.13}$$

We next further simplify $\mathcal{R}(t)$ by analyzing the relationship between $\hat{\lambda}^0$ and $\hat{\lambda}^1$, namely by showing that $\hat{\lambda}^1 = 0$ when $\hat{\lambda}^0 < 0$ and $\hat{\lambda}^1 = \hat{\lambda}^0$ when $\hat{\lambda}^0 \geq 0$. Defining

$$a_j(\lambda) \equiv \prod_{i \leq N_j(t)} \left\{1 - \frac{d_{ij}}{r_{ij} + (-1)^{j-1}\lambda}\right\}$$

for $j = 1, 2$ and

$$a(\lambda) \equiv \frac{a_1(\lambda)}{a_2(\lambda)},$$

we can see that $a_j(0) = \hat{S}_j(t)$, $\hat{\lambda}^0$ satisfies $a(\hat{\lambda}^0) = 1$, and $\hat{\lambda}^1$ satisfies $a(\hat{\lambda}^1) \geq 1$. Notice that $a(\lambda)$ is strictly increasing in λ on $(D_1, -D_2)$, tending to 0 and ∞ as $\lambda \downarrow D_1$ and $\uparrow -D_2$, respectively. Also, condition (2.11) implies either $\hat{\lambda}^1 = 0$ or

$$\sum_{i \leq N_2(t)} \log(1 - h_{i2}) - \sum_{i \leq N_1(t)} \log(1 - h_{i1}) = 0 \tag{2.14}$$

must hold, and since (2.14) is equivalent to $\hat{\lambda}^1 = \hat{\lambda}^0$, we obtain that $\hat{\lambda}^1$ is either 0 or $\hat{\lambda}^0$. These observations along with (2.9) and (2.10) imply the following:

Case 1: If $\hat{\lambda}^0 < 0$, then by (2.10) we have $\hat{\lambda}^1 \neq \hat{\lambda}^0$. Since $\hat{\lambda}^1$ is either 0 or $\hat{\lambda}^0$, we obtain that $\hat{\lambda}^1 = 0$.

Case 2: If $\hat{\lambda}^0 > 0$, then by monotonicity of $a(\lambda)$ we have $a(0) < 1$. Suppose $\hat{\lambda}^1 = 0$, then $a(0) \geq 1$ by (2.9), which contradicts $a(0) < 1$. So we have $\hat{\lambda}^1 = \hat{\lambda}^0$.

Case 3: If $\hat{\lambda}^0 = 0$, then because $\hat{\lambda}^1$ is either 0 or $\hat{\lambda}^0$, we can see that $\hat{\lambda}^1 = \hat{\lambda}^0 = 0$.

Then from (2.13) we have

$$\mathcal{R}(t) = \begin{cases} 1, & \hat{\lambda}^0 \geq 0, \\ \displaystyle\prod_{j=1}^{2} \prod_{i \leq N_j(t)} \frac{\left(\hat{h}_{ij}^0\right)^{d_{ij}} \left(1 - \hat{h}_{ij}^0\right)^{r_{ij}-d_{ij}}}{\left(\frac{d_{ij}}{r_{ij}}\right)^{d_{ij}} \left(1 - \frac{d_{ij}}{r_{ij}}\right)^{r_{ij}-d_{ij}}}, & \hat{\lambda}^0 < 0. \end{cases}$$

This is exactly (2.7). We use the simplified notation $\hat{h}_{ij}$ and $\hat{\lambda}$ to replace $\hat{h}_{ij}^0$ and $\hat{\lambda}^0$, respectively.

Another version of (2.7) will be used in the proof of Theorem 2.2: replacing $\hat{\lambda}^0 \geq 0$ and $\hat{\lambda}^0 < 0$ in (2.7) by $\hat{S}_1(t) \leq \hat{S}_2(t)$ and $\hat{S}_1(t) > \hat{S}_2(t)$, respectively. This version is based on the equality of the events $\hat{\lambda}^0 < 0$ and $\hat{S}_1(t) > \hat{S}_2(t)$, which can be seen by noting that $a(\lambda)$ is strictly increasing, $a(\hat{\lambda}^0) = 1$ and $a(0) = \hat{S}_1(t)/\hat{S}_2(t)$.

2.6.2 Proof of Theorem 2.2

We will need the following lemma giving an asymptotic expansion of the localized EL statistic in terms of $\hat{S}_1(t)$ and $\hat{S}_2(t)$.

Lemma 2.3.

$$\begin{aligned} -2\log\mathcal{R}(t) = & \frac{n}{\hat{\sigma}^2(t)} \left\{\log \hat{S}_1(t) - \log \hat{S}_2(t)\right\}^2 I\left\{\hat{S}_1(t) > \hat{S}_2(t)\right\} \\ & + O_p(n^{-1/2}), \end{aligned}$$

where the O_p term holds uniformly in t over $[t_1, t_2]$.

Proof. We first find the asymptotic order of $\hat{\lambda}(t)$ uniformly for $t \in [t_1, t_2]$, then we derive an asymptotic expansion of $\hat{\lambda}(t)$ uniformly for $t \in [t_1, t_2]$. Next, by a Taylor series expansion, we approximate $-2\log\mathcal{R}(t)$ as a function of $\hat{\lambda}(t)$. Based on the two expansions, we obtain the desired result.

First, we find the asymptotic order of the Lagrange multiplier $\hat{\lambda}(t)$. Since $\hat{\lambda}(t)$ comes from the numerator of the EL ratio (2.6), it satisfies the equality constraint (2.12). McKeague and Zhao (2002) studied the same Lagrange multiplier derived from optimizing the nonparametric likelihood under an equality constraint on the ratio of two survival functions, so by their Lemma A.1,

$$\hat{\lambda}(t) = O_p(\sqrt{n}) \tag{2.15}$$

uniformly for $t \in [t_1, t_2]$.

Next we derive an asymptotic expansion of $\hat{\lambda}(t)$. The expansion is obtained by Taylor expanding the l.h.s. of

$$\sum_{i \le N_1(t)} \log\left\{1 - \frac{d_{i1}}{r_{i1} + \hat{\lambda}(t)}\right\} - \sum_{i \le N_2(t)} \log\left\{1 - \frac{d_{i2}}{r_{i2} - \hat{\lambda}(t)}\right\} = 0$$

and then rearranging terms. In detail, the j-th term ($j = 1, 2$) on the l.h.s., by a similar argument in Hollander et al. (1997, p. 225), has the expansion

$$\log \hat{S}_j(t) + \Delta_j \hat{\lambda}(t) \frac{\hat{\sigma}_j^2(t)}{n_j} + O_p(n_j^{-1}),$$

where $\Delta_j = 1$ for $j = 1$ and -1 for $j = 2$. Combining the two terms and using $n_j/n \to p_j$ gives

$$\log \hat{S}_1(t) - \log \hat{S}_2(t) + \hat{\lambda}(t) \frac{\hat{\sigma}^2(t)}{n} + O_p(n^{-1}) = 0.$$

Rearranging the terms, we have

$$\hat{\lambda}(t) = -\frac{n}{\hat{\sigma}^2(t)} \left\{\log \hat{S}_1(t) - \log \hat{S}_2(t) + O_p(n^{-1})\right\}. \tag{2.16}$$

Next, we find an asymptotic expansion of $-2\log\mathcal{R}(t)$ as a function of $\hat{\lambda}(t)$. We begin, based on (2.7), by writing $-2\log\mathcal{R}(t)$ as

$$-2\sum_{j=1}^{2}\sum_{i\le N_j(t)}\left[(r_{ij}-d_{ij})\log\left\{1+\frac{\Delta_j\hat\lambda(t)}{r_{ij}-d_{ij}}\right\}\right]$$
$$+2\sum_{j=1}^{2}\sum_{i\le N_j(t)}\left[r_{ij}\log\left\{1+\frac{\Delta_j\hat\lambda(t)}{r_{ij}}\right\}\right]$$

times an indicator $I(\hat\lambda(t)<0)$. The j-th term above, by a similar argument in Li (1995, p.102), has the expansion

$$\hat\lambda^2(t)\sum_{i\le N_j(t)}\frac{d_{ij}}{r_{ij}(r_{ij}-d_{ij})}+O_p(n_j^{-1/2})$$

for $j=1,2$. Using $n_j/n\to p_j$, and the fact that $\hat\lambda(t)<0$ is equivalent to $\hat S_1(t)>\hat S_2(t)$, we can combine the terms for $j=1,2$ and obtain

$$-2\log\mathcal{R}(t)=\left\{\hat\sigma^2(t)\frac{\hat\lambda^2(t)}{n}+O_p(n^{-1/2})\right\}I\left\{\hat S_1(t)>\hat S_2(t)\right\}.$$

This and (2.16) give the desired result.

□

Remark. Lemma 2.3 shows that $-2\log\mathcal{R}(t)$ is asymptotically equivalent to squaring the positive part of a scaled difference between the log of KM estimators from the two samples. The inclusion of only the positive part of the difference can be attributed to the stochastically ordered form of our alternative hypothesis. We have compared the small sample performance of K_n and its counterpart based on this squared difference (results not shown), and it turns out the latter tends to be too conservative.

The advantage of using the EL approach, as opposed to a test statistic derived from the first term in the expansion of Lemma 2.3, is that we expect higher-order accuracy (cf. Hall and La Scala, 1990). This is parallel to the parametric result in which the likelihood ratio test is asymptotically equivalent to the Wald test, but the former has better higher-order accuracy (see, e.g., Mukerjee, 1994).

We now complete the proof of Theorem 2.2.

We first obtain the weak convergence of $-2\log\mathcal{R}(t)$ as a process on $[t_1, t_2]$, based on Lemma 2.3 and large sample properties of the KM estimator. Then by a transformation of the limiting process and the continuous mapping theorem, we get the limiting distribution of K_n.

To obtain the limit process of $-2\log\mathcal{R}(t)$, we begin by finding the weak convergence of $\log\hat{S}_1 - \log\hat{S}_2$, as the asymptotic expansion of $-2\log\mathcal{R}(t)$ in Lemma 2.3 suggests. For each $j = 1, 2$, it has been shown (see, e.g., Andersen et al., 1993, p.191 and p.263) that

$$\sqrt{n_j}\left(\log\hat{S}_j - \log S_j\right) \xrightarrow{d} U_j$$

as $n \to \infty$ on $D[0, t_2]$, where $U_j(t)$ is a Gaussian martingale with $U_j(0) = 0$ and $\text{cov}(U_j(s), U_j(t)) = \sigma_j^2(\min(s,t))$. Therefore, under H_0, the continuous mapping theorem implies

$$\sqrt{n}\left(\log\hat{S}_1 - \log\hat{S}_2\right) \xrightarrow{d} \frac{U_1}{\sqrt{p_1}} - \frac{U_2}{\sqrt{p_2}} \equiv U, \tag{2.17}$$

where $U(t)$ is a Gaussian martingale with $U(0) = 0$ and $\text{cov}(U(s), U(t)) = \sigma^2(\min(s,t))$.

Next, we establish the weak convergence of $-2\log\mathcal{R}(t)$. By (2.17) and the continuous mapping theorem, we have

$$n\left\{\log\hat{S}_1(t) - \log\hat{S}_2(t)\right\}^2 I\left\{\hat{S}_1(t) > \hat{S}_2(t)\right\} \xrightarrow{d} U_+^2(t)$$

in $D[t_1, t_2]$, where $U_+ = \max(U, 0)$. Then by the uniform consistency of $\hat{\sigma}^2(t)$ with respect to $\sigma^2(t)$ and Slutsky's Lemma, we have

$$\frac{n}{\hat{\sigma}^2(t)}\left\{\log\hat{S}_1(t) - \log\hat{S}_2(t)\right\}^2 I\left\{\hat{S}_1(t) > \hat{S}_2(t)\right\} \xrightarrow{d} \frac{U_+^2(t)}{\sigma^2(t)}$$

in $D[t_1, t_2]$. This and Lemma 2.3 imply

$$-2\log\mathcal{R}(t) \xrightarrow{d} \frac{U_+^2(t)}{\sigma^2(t)} \tag{2.18}$$

in $D[t_1, t_2]$.

Lastly, the asymptotic null distribution of K_n is obtained as follows. First notice that

$$\frac{U(t)}{1+\sigma^2(t)} \quad \text{and} \quad B\left(\frac{\sigma^2(t)}{1+\sigma^2(t)}\right)$$

are both zero mean Gaussian processes with the same covariance function, so they have the same distribution. We then have $U_+^2(t)/\sigma^2(t)$ equal in distribution to

$$B_+^2\left(\frac{\sigma^2(t)}{1+\sigma^2(t)}\right)\frac{\left(1+\sigma^2(t)\right)^2}{\sigma^2(t)}.$$

This, together with (2.17) and the continuous mapping theorem, implies that $\sup_{t\in[t_1,t_2]}$ $\{-2\log\mathcal{R}(t)\}$ converges in distribution to

$$\sup_{t\in[t_1,t_2]}\left\{B_+^2\left(\frac{\sigma^2(t)}{1+\sigma^2(t)}\right)\frac{\left(1+\sigma^2(t)\right)^2}{\sigma^2(t)}\right\}.$$

The result follows from noticing that the r.h.s. of the above is the same as

$$\sup_{x\in[x_1,x_2]}\left\{\frac{B_+^2(x)}{x(1-x)}\right\},$$

where $x_1 = b(t_1)$ and $x_2 = b(t_2)$.

2.6.3 Validating the calibration procedure

The following result justifies the approach of pre-specifying $[x_1, x_2]$ and estimating $[t_1, t_2]$, as outlined in Section 2.2.2.1.

Lemma 2.4. *Suppose S_0 is continuous. Then under H_0, for $0 < x_1 < x_2 < 1$,*

$$K_n^* \xrightarrow{d} \sup_{x\in[x_1,x_2]}\left\{\frac{B_+^2(x)}{x(1-x)}\right\},$$

provided $b^{-1}(\cdot)$ is continuous at x_1 and x_2, where K_n^* is just K_n with t_1 and t_2 replaced by $\hat{t}_1 = \max\{\hat{b}^{-1}(x_1), T_{11}, T_{12}\}$ and $\hat{t}_2 = \min\{\hat{b}^{-1}(x_2), T_{m_1 1}, T_{m_2 2}\}$, respectively.

Proof. The idea is to obtain the joint convergence of $-2\log\mathcal{R}(t)$, $\hat{t}_1$ and $\hat{t}_2$, and then to apply the continuous mapping theorem.

First, we show the weak convergence of $-2\log\mathcal{R}(t)$. We will apply (2.18) in the proof of Theorem 2.2, but we need to translate the conditions to be in terms of x_1 and x_2 instead of t_1 and t_2. Given $0 < x_1 < x_2 < 1$ at which $b^{-1}(\cdot)$ is continuous, it suffices to show that $t_1 = b^{-1}(x_1)$ and $t_2 = b^{-1}(x_2)$ satisfy the conditions $S_0(t_1) < 1$

and $S_0(t_2)G_j(t_2) > 0$ for $j = 1, 2$. To show $S_0(t_1) < 1$, we simply use $b(t_1) = x_1 > 0$, which implies $\sigma^2(t_1) > 0$ and thus $S_0(t_1) < 1$. To show $S_0(t_2)G_j(t_2) > 0$ for $j = 1, 2$, we argue by contradiction. Suppose $S_0(t_2)G_j(t_2) = 0$ for some $j = 1, 2$. Since b is continuous (by continuity of S_0) and nondecreasing, we can pick an $\epsilon < 1 - x_2$ and δ small enough such that $x_2 \leq b(t_2 + \delta) < x_2 + \epsilon < 1$. Because b^{-1} is continuous at x_2, there is no "flat" of b around t_2, and thus δ can be chosen so that b is strictly increasing in $[t_2, t_2 + \delta]$. This and $S_0(t_2)G_j(t_2) = 0$ lead to $b(t_2 + \delta) = 1$, which contradicts $b(t_2 + \delta) < x_2 + \epsilon < 1$. So we have $S_0(t_2)G_j(t_2) > 0$ for $j = 1, 2$, as required.

Next, we show $\hat{t}_j \xrightarrow{P} t_j$ for $j = 1, 2$. The proof makes use of the theory of Z-estimators (see, e.g., van der Vaart, 2000, Theorem 5.9). Let $\Psi_n(t) = \hat{b}(t) - x_1$, $\Psi(t) = b(t) - x_1$, and $\Theta = [\tau_1, \tau_2]$ such that $[t_1, t_2] \subset \Theta \subset (0, \infty)$. We already know $\Psi_n(t_1) = o_p(1)$ and $\Psi(t_1) = 0$. It suffices to show that $\sup_{t \in \Theta} |\Psi_n(t) - \Psi(t)| \xrightarrow{P} 0$ and $\inf_{t:|t-t_1| \geq \epsilon} |\Psi(t)| > 0$ for all $\epsilon > 0$. The former is implied by the uniform consistency of $\hat{\sigma}^2$ (and thus b), and the latter by the continuity of b^{-1} at x_1. Therefore we have $\hat{t}_1 \xrightarrow{P} t_1$. The same argument applies to $\hat{t}_2$.

Lastly, the asymptotic null distribution of K_n^* is obtained as follows. From the weak convergence of $-2 \log \mathcal{R}(t)$ and $\hat{t}_j \xrightarrow{P} t_j$ for $j = 1, 2$, we have the joint convergence $[-2 \log \mathcal{R}(t), \hat{t}_1, \hat{t}_2]^T \xrightarrow{d} [U_+^2(t)/\sigma^2(t), t_1, t_2]^T$ in $D[t_1, t_2] \times \Theta^2$ (see, e.g., van der Vaart, 2000, Theorem 18.10 (v)). Then applying a similar argument in the last part of the proof for Theorem 2.2 and the continuous mapping theorem, we get the desired result.

□

Table 2.1: Critical values for K_n^* for selected x_1, x_2 and α.

x_1	0.1			0.15			0.2		
$x_2 \backslash \alpha$	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
0.975	11.822	8.255	6.648	11.672	8.074	6.489	11.542	7.953	6.365
0.98	11.912	8.329	6.720	11.758	8.159	6.556	**11.619**	**8.028**	**6.442**
0.985	11.996	8.415	6.807	11.851	8.253	6.658	11.739	8.131	6.532

Table 2.2: Empirical significance levels based on 10,000 replications.

CR	group	$\alpha = 0.05$			$\alpha = 0.01$		
	size	K_n^*	log-rank	WKM	K_n^*	log-rank	WKM
	50	0.040	0.057	0.055	0.007	0.013	0.011
10%	80	0.041	0.052	0.054	0.008	0.010	0.010
	200	0.045	0.051	0.048	0.009	0.011	0.011
	50	0.037	0.057	0.054	0.006	0.012	0.012
25%	80	0.041	0.051	0.056	0.008	0.009	0.010
	200	0.046	0.054	0.050	0.010	0.010	0.011

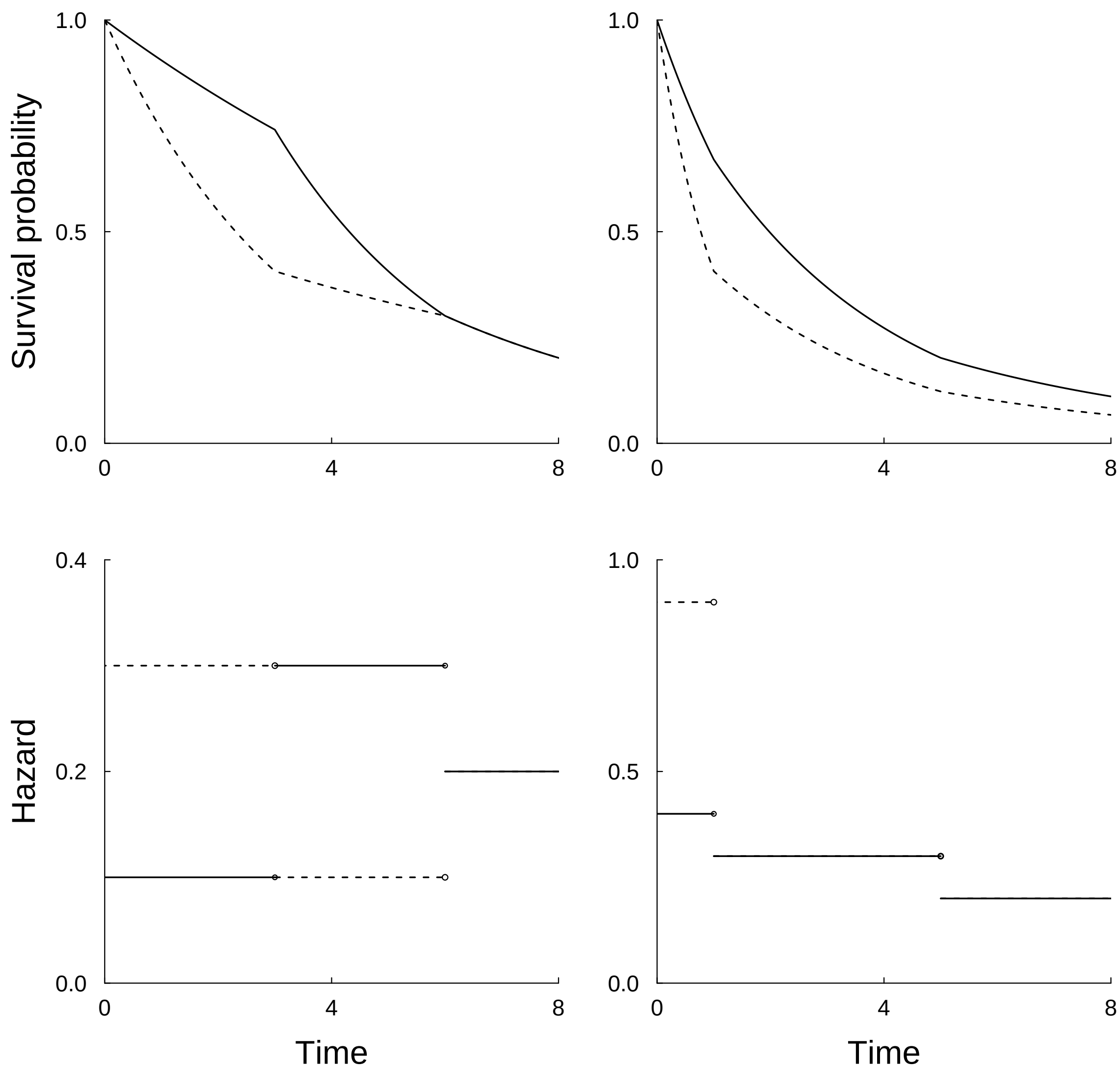

Figure 2.1: The piecewise exponential survival functions (top row) and the hazard functions (bottom row) in Model A (first column): S_1 (solid) and S_2 (dashed), and in Model B (second column): S_1 (solid) and S_2 (dashed).

Table 2.3: Power at $\alpha = 0.05$ based on 10,000 replications. **Model A**: survival functions as in Figure 2.1, upper left panel. **Model B**: survival functions as in Figure 2.1, upper right panel.

model	group size	test	exp. censoring 10%	exp. censoring 25%	unif. censoring 10%	unif. censoring 25%
Model A	50	K_n^*	0.851	0.833	0.849	0.834
		log-rank	0.318	0.379	0.314	0.373
		WKM	0.328	0.391	0.330	0.431
	80	K_n^*	0.975	0.968	0.975	0.971
		log-rank	0.416	0.503	0.415	0.501
		WKM	0.426	0.507	0.433	0.569
Model B	50	K_n^*	0.689	0.672	0.688	0.676
		log-rank	0.625	0.659	0.621	0.650
		WKM	0.521	0.583	0.521	0.613
	80	K_n^*	0.876	0.862	0.877	0.869
		log-rank	0.782	0.815	0.784	0.812
		WKM	0.660	0.729	0.675	0.775

Table 2.4: Size and power for various choices of x_1 and x_2 based on 10,000 replications, $\alpha = 0.05$, $n_1 = n_2 = 50$, and exponential censoring with censoring rate 10%. **Model A**: survival functions as in Figure 2.1, upper left panel. **Model B**: survival functions as in Figure 2.1, upper right panel. For size, only the solid survival functions are used.

x_1	x_2	critical	size		power	
		value	Model A	Model B	Model A	Model B
0.2	0.98	8.028	0.040	0.040	0.851	0.689
0.2	0.8	6.879	0.037	0.039	0.890	0.703
0.02	0.98	8.829	0.029	0.028	0.806	0.628
0.02	0.8	8.048	0.023	0.025	0.838	0.612

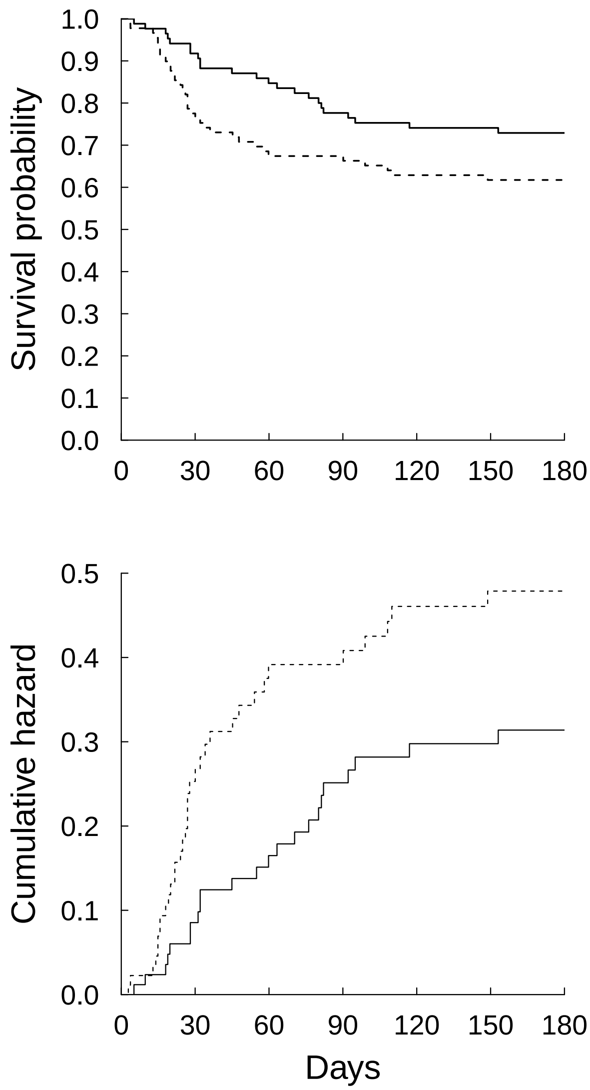

Figure 2.2: Estimates of survival functions (top) and cumulative hazards (bottom) for prednisolone plus *N*-acetylcysteine (solid line) versus prednisolone alone (dashed line).

Chapter 3

Empirical likelihood based tests for stochastic ordering in biased sampling models

3.1 Introduction

Sampling bias is found to be involved in numerous applications, such as in genetics (see, e.g., Fisher, 1934; Clark et al., 2005, ascertainment bias), animal studies (Patil and Rao, 1978), disease screening (Zelen and Feinleib, 1969; Duffy et al., 2008), and vaccine trials (Gilbert et al., 1999). Because of the bias, the observed (biased) distribution of the data is different from the underlying true (unbiased) distribution. Thus, inference cannot be conducted via standard statistical methods. This is especially challenging when comparing more than one group of data, each group having a different biasing mechanism and underlying distribution. This chapter aims to develop a procedure for such group comparison.

More specifically, we consider the setup where there are two underlying distribution functions of interest, F_1 and F_2. Due to sampling bias, instead of observing samples from F_j directly ($j = 1, 2$), we observe samples from a biased version of F_j:

$$G_j(x) = \int_0^x \frac{w_j(u)}{W_j} dF_j(u)$$

according to some known biasing or weight function $w_j(\cdot) > 0$, where $W_j = \int_0^\infty w_j(u) dF_j(u) < \infty$ is the normalizing constant. The data are also called size-biased because the biasing function $w_j(x)$ depends on the size x of the datum. For groups of size-biased data, the nonparametric maximum likelihood estimator (NPMLE) for the unbiased distribution function and its weak convergence have been established (Vardi, 1982, 1985; Gill et al., 1988). The NPMLE has been shown to converge weakly to a pinned Gaussian process. A two-sample test based on the NPMLEs from each sample can be constructed; however, only point-wise comparison is feasible, because the limiting process depends on the unbiased distribution.

In contrast to the common practice of comparing the groups in terms of the mean, median or some other univariate feature of the distributions, we compare the entire distribution functions. We focus our attention on detecting an ordering between the distribution functions, which can be framed in terms of the classical notion of *stochastic ordering*. Namely, F_1 is said to be *stochastically larger* than F_2 if $F_1(x) \leq F_2(x)$ for all $x \geq 0$; this is denoted as $F_1 \succeq F_2$. We investigate the problem of testing the two-sided alternative

$$H_0 : F_1 = F_2 \text{ versus } H_1 : F_1 \succ F_2 \text{ or } F_2 \succ F_1 \tag{3.1}$$

where $\succ$ denotes $\succeq$ with strict inequality for some x. Our approach will first be developed for testing the one-sided alternative

$$H_0 : F_1 = F_2 \text{ versus } H_1 : F_1 \succ F_2 \tag{3.2}$$

and then extended to the two-sided alternative using the union-intersection principle. Hypothesis testing for (3.2) has been investigated extensively in the literature (see, e.g., El Barmi and McKeague, 2013, and the references therein), though without consideration of sampling bias. Such order-restricted inference incorporates a priori constraints into the construction of the parameter space and the hypotheses, thereby increasing efficiency (see, e.g., Silvapulle and Sen, 2005).

We derive our procedure using the empirical likelihood (EL) method. EL utilizes a similar construction to the usual likelihood ratio test, except the optimization is with respect to a nonparametric likelihood instead. The method has been used to provide tests and confidence regions for parameters defined by estimating equations (Owen, 1988, 2001). EL is more appealing than the Wald approach because it is self-studentized and provides more accurate confidence regions. There is also evidence that EL-based tests have optimal power (see, e.g., Kitamura et al., 2012). Due to these favorable properties, EL has been applied to biased sampling problems (Qin, 1993; El Barmi and Rothmann, 1998; Davidov et al., 2010). However, comparing the underlying distribution functions uniformly via simultaneous confidence bands and hypothesis testing has not been considered.

Our contribution is to provide an EL-based test for stochastic ordering between the unbiased distributions of size-biased samples. First consider the one-sided alternative in (3.2). The idea is to construct a localized EL statistic for $H_0^t : F_1(t) = F_2(t)$ versus $H_1^t : F_1(t) < F_2(t)$ at each given t. The proposed test rejects H_0 for large values of the maximally selected EL statistic. The localization approach has been used in Einmahl and McKeague (2003) and El Barmi and McKeague (2013) for testing various nonparametric hypotheses, except they considered an integral type test statistic and restricted attention to data without sampling bias. Various Kolmogorov–Smirnov type test statistics (not based on EL) for stochastic ordering have been proposed by El Barmi and Mukerjee (2005) and Davidov and Herman (2009), but these cannot deal with size-biased data, either.

The chapter is organized as follows. In Section 3.2.1 we set up the general framework and notation to be used throughout the chapter. While our focus is on the two-sample test in Section 3.2.3, for clarity of exposition the one-sample test will be introduced first (in Section 3.2.2). Extensions to two-sided and stratified testing are discussed in Section 3.2.4. Section 3.3 presents results from a simulation study: the proposed EL test performs better than the Wald test and the test ignoring size bias, in terms of both accuracy and power. Section 3.4 then provides an application of the proposed test to alcohol concentration records in fatal driving accidents. Finally, some concluding remarks are placed in Section 3.5.

3.2 EL tests for stochastic ordering in biased sampling model

3.2.1 Preliminaries

We begin by introducing notation for the one-sample case. Suppose the underlying distribution is supported on $\mathbb{R}^+$. Let the observed data X_i $(i = 1, \ldots, n)$ be i.i.d. from a weighted distribution function $G(x) = \int_0^x \frac{w(u)}{W} dF(u)$, where $w(\cdot) > 0$ is a known weight function and $W = \int_0^\infty w(u)dF(u) < \infty$ is the normalizing constant. Then the nonparametric likelihood (see, e.g., Owen, 2001, Ch. 6.1) for F is $L(F) = \prod_i dG(X_i) = \prod_i \frac{w(X_i)dF(X_i)}{W}$, which can be written as $\prod_{i=1}^n (w_i p_i)/W$ if we denote $w(X_i)$ as w_i and $dF(X_i)$ as p_i. The NPMLE has been shown to be $\tilde{F}(t) \equiv \sum_{i=1}^n \tilde{p}_i I_{X_i \le t}$, where $\tilde{p}_i = \tilde{W}/(nw_i)$ and $\tilde{W} = n/\sum_{i=1}^n (1/w_i)$.

For the two-sample case, we use a similar framework as in the one-sample setup with a *further subscript j indicating the j-th sample in the corresponding notation.* The nonparametric likelihood $L(F_1, F_2)$ is the product of the two one-sample likelihoods, $L(F_1)L(F_2)$. We assume the sample proportion $\kappa_j \equiv n_j/n > 0$ (for $j = 1, 2$) to remain fixed as the total sample size $n \to \infty$.

3.2.2 One-sample case

Consider testing $H_0 : F = F_0$ versus $H_1 : F \succ F_0$, where F_0 is a known distribution function. Our procedure is to first construct the statistic for testing the "local" hypotheses $H_0^t : F(t) = F_0(t)$ versus $H_1^t : F(t) < F_0(t)$ for a given t, and then to deal with the general hypotheses based on some functional of the local statistics.

To construct the local test statistic at t, consider the EL ratio

$$\mathcal{R}(t) = \frac{\sup\{L(F) : F(t) = F_0(t)\}}{\sup\{L(F) : F(t) \leq F_0(t)\}}, \tag{3.3}$$

where we use the conventions $\sup\emptyset = 0$ and $0/0 = 1$. A tractable form of the EL ratio can be obtained by first comparing $F_0(t)$ and the global optimum $\tilde{F}(t)$. When $\tilde{F}(t) \leq F_0(t)$, the denominator of $\mathcal{R}(t)$ achieves the global optimal value, $\prod_{i=1}^n (w_i\tilde{p}_i)/\tilde{W} = \prod_{i=1}^n (1/n)$. When $\tilde{F}(t) > F_0(t)$, we can show that the constrained maximum in the denominator is attained on the boundary of the constraint set, in which case $\mathcal{R}(t) = 1$ (see Supplementary Material Section 5.3.1). That is,

$$\mathcal{R}(t) = \begin{cases} 1 & \text{if } \tilde{F}(t) > F_0(t), \\ \dfrac{\sup\{L(F) : F(t) = F_0(t)\}}{n^{-n}} & \text{if } \tilde{F}(t) \leq F_0(t). \end{cases}$$

To find $\sup\{L(F) : F(t) = F_0(t)\}$, we follow a similar derivation in the literature of EL-based testing in biased sampling model (see, e.g., Qin, 1993; El Barmi and Rothmann, 1998). We have that $\sup\{L(F) : F(t) = F_0(t)\} = \prod_{i=1}^n (w_i\hat{p}_i)/\hat{W}$, where

$$\hat{p}_i = \frac{1}{n}\frac{1}{\hat{a} - 1 + w_i/\hat{W} + \hat{\lambda}(I_{X_i \leq t} - F_0(t))}$$

and $(\hat{a}, \hat{W}, \hat{\lambda})$ satisfy the estimating equations $\sum_{i=1}^n \hat{p}_i = 1$, $\sum_{i=1}^n \hat{p}_i(w_i - \hat{W}) = 0$ and $\sum_{i=1}^n \hat{p}_i(I_{X_i \leq t} - F_0(t)) = 0$ (see Appendix 3.6.1 for more details). This gives

$$\mathcal{R}(t) = \begin{cases} 1 & \text{if } \tilde{F}(t) > F_0(t), \\ \displaystyle\prod_{i=1}^n \frac{nw_i\hat{p}_i}{\hat{W}} & \text{if } \tilde{F}(t) \leq F_0(t). \end{cases}$$

Based on the above expression, we can derive large sample properties of the local EL test statistic, $-2\log\mathcal{R}(t)$. This is done via approximating $-2\log\mathcal{R}(t)$ by its Wald-type counterpart $U_n^2(t)I_{U_n(t)\geq 0}$, where

$$U_n(t) = \hat{\sigma}^{-\frac{1}{2}}(t,t)\left[\frac{W}{\sqrt{n}}\sum_{i=1}^n \frac{F_0(t) - I_{X_i \leq t}}{w_i}\right]$$

and $\hat{\sigma}(t,t) = \hat{W}^2 \sum_{i=1}^n [(I_{X_i \le t} - F_0(t))/w_i]^2/n$ (see Appendix 3.6.2 for more details). It can be shown that $U_n^2(t) I_{U_n(t) \ge 0}$ is asymptotically chi-bar square under H_0^t, thereby establishing the limiting null distribution of $-2\log\mathcal{R}(t)$. Namely, for t such that $0 < F_0(t) < 1$,

$$-2\log\mathcal{R}(t) \xrightarrow{d} Z_+^2$$

under H_0^t, where $Z \sim N(0,1)$ and $Z_+ = \max(Z,0)$. This result can be used to test the local H_0^t versus H_1^t.

To test for the alternative of stochastic ordering, we propose the following maximally selected EL statistic:

$$M_n = \sup_{t\in[t_1,t_2]} \left[-2\log\mathcal{R}(t)\right], \tag{3.4}$$

where $[t_1, t_2]$ is a proper subset of the support of F_0. Such restriction on the range of t is necessary in deriving the asymptotics. In practice we just take t_1 and t_2 to be the smallest and largest observations, respectively.

Our first result gives the asymptotic null distribution of M_n. The proof is omitted, because it is similar to the two-sample case (presented in Appendix 3.6.2).

Theorem 3.1. *Suppose* $E_{F_0}(1/w^2(X_i)) < \infty$. *Then under* H_0, *for* t_1 *and* t_2 *satisfying* $0 < F_0(t_l) < 1$ *(*$l = 1,2$*),*

$$M_n \xrightarrow{d} \sup_{t\in[t_1,t_2]} \left[U_+^2(t)\right],$$

where $U_+ = \max(U,0)$, $U(t)$ *is a mean 0 pinned Gaussian process with covariance* $\mathrm{cov}(U(s),U(t)) = \sigma(s,t)/\sqrt{\sigma(s,s)\sigma(t,t)}$, *and* $\sigma(s,t) = W^2 E_G[(I_{X_i\le s} - F_0(s))(I_{X_i\le t} - F_0(t))/w^2(X_i)]$.

Remark. When there is no size bias (i.e. $w_i = 1$), $U_n(t)$ reduces to

$$\frac{\sqrt{n}\,[F_0(t) - F_n(t)]}{\sqrt{F_n(t)\,(1-F_n(t))}}$$

where $F_n(t)$ denotes the empirical cdf. This implies

$$M_n \xrightarrow{d} \sup_{x\in[x_1,x_2]} \frac{B_+^2(x)}{x(1-x)},$$

where B is a standard Brownian bridge on $[0,1]$ and $x_l = F(t_l)$, $l = 1, 2$.

To obtain critical values based on the the limiting distribution in Theorem 3.1 (which depends on the unknown parameters), we propose a Gaussian multiplier bootstrap approach as follows.

3.2.2.1 Gaussian multiplier bootstrap calibration

This section constructs multiplier bootstrap approximations of the limiting distribution in Theorem 3.1. We begin with bootstrapping $U_n(t)$, whose functional $\sup_{t\in[t_1,t_2]}[U_n^2(t)I_{U_n(t)\geq 0}]$ is asymptotically equivalent to the test statistic M_n (see Section 3.2.2). Define a Gaussian multiplier bootstrap for $U_n(t)$ by

$$U_n^*(t) = \hat{\sigma}^{-\frac{1}{2}}(t,t)\left[\frac{\hat{W}}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\frac{F_0(t) - I_{X_i\leq t}}{w_i}\right],$$

where ξ_i $(i = 1, \ldots, n)$ are i.i.d. standard Gaussian random variables that are independent of the observed data. We can show bootstrap consistency of $U_n^*(t)$, thereby establishing consistency of

$$M_n^* \equiv \sup_{t\in[t_1,t_2]}\left[U_n^{*2}(t)I_{U_n^*(t)\geq 0}\right].$$

The result is provided in the following theorem. The proof is omitted, because it is similar to the two-sample case (presented in Appendix 3.6.3).

Theorem 3.2. *Assume the conditions of Theorem 3.1. Then conditionally on* $X_1, X_2, \ldots$, *for almost every sequence* $X_1, X_2, \ldots$,

$$M_n^* \xrightarrow{d} \sup_{t\in[t_1,t_2]}\left[U_+^2(t)\right].$$

Based on Theorem 3.2, to calibrate the test, we simulate M_n^* by repeatedly generating Gaussian random samples $\{\xi_i\}$ while holding the observed data fixed. We then compare the empirical quantiles of these bootstrap values M_n^* with our test statistic M_n. Similar multiplier bootstrap procedures have also been considered in Rémillard and Scaillet (2009) and Chernozhukov et al. (2013).

3.2.3 Two-sample case

We now adapt our approach to the two-sample case. The "local" hypotheses are $H_0^t : F_1(t) = F_2(t) \equiv F_0(t)$ versus $H_1^t : F_1(t) < F_2(t)$ for a given t. The local EL ratio at t is defined to be

$$\mathcal{R}(t) = \frac{\sup\{L(F_1, F_2) : F_1(t) = F_2(t)\}}{\sup\{L(F_1, F_2) : F_1(t) \leq F_2(t)\}}. \tag{3.5}$$

Similar to the one-sample case, we compare $\tilde{F}_1$ and $\tilde{F}_2$, derive $\sup\{L(F_1, F_2) : F_1(t) \leq F_2(t)\}$ accordingly, and solve for $\sup\{L(F_1, F_2) : F_1(t) = F_2(t)\}$ using the method of Lagrange multipliers (see Appendix 3.6.1 for more details). We then have

$$\mathcal{R}(t) = \begin{cases} 1 & \text{if } \tilde{F}_1(t) > \tilde{F}_2(t), \\ \displaystyle\prod_{j=1}^{2} \prod_{i=1}^{n_j} \frac{n_j w_{ij} \hat{p}_{ij}}{\hat{W}_j} & \text{if } \tilde{F}_1(t) \leq \tilde{F}_2(t), \end{cases} \tag{3.6}$$

where

$$\hat{p}_{ij} = \frac{1}{n} \frac{1}{\hat{a}_j - \kappa_j + (\kappa_j w_{ij})/\hat{W}_j + \hat{\lambda}(-1)^{j-1}(I_{X_{ij} \leq t} - \hat{F}_0(t))},$$

$(\hat{a}_1, \hat{a}_2, \hat{W}_1, \hat{W}_2, \hat{\lambda}, \hat{F}_0(t))$ satisfy (for $j = 1, 2$)

$$\sum_{i=1}^{n_j} \hat{p}_{ij} = 1, \sum_{i=1}^{n_j} \hat{p}_{ij} \left(w_{ij} - \hat{W}_j\right) = 0, \sum_{i=1}^{n_j} \hat{p}_{ij} \left(I_{X_{ij} \leq t} - \hat{F}_0(t)\right) = 0, \tag{3.7}$$

and $\hat{F}_0(t)$ denotes the estimate of the common underlying distribution function $F_0(t)$. The local EL test statistic $-2 \log \mathcal{R}(t)$ is shown to converge in distribution to chi-bar square under H_0^t. The derivation involves approximating $-2 \log \mathcal{R}(t)$ by its Wald-type counterpart $U_n^2(t) I_{U_n(t) \geq 0}$, where

$$U_n(t) = \hat{\sigma}^{-\frac{1}{2}}(t,t) \left[\frac{W_2}{\sqrt{n_2}\sqrt{\kappa_2}} \sum_{i=1}^{n_2} \frac{I_{X_{i2} \leq t} - F_0(t)}{w_{i2}} - \frac{W_1}{\sqrt{n_1}\sqrt{\kappa_1}} \sum_{i=1}^{n_1} \frac{I_{X_{i1} \leq t} - F_0(t)}{w_{i1}} \right] \tag{3.8}$$

and $\hat{\sigma}(t,t) = \sum_{j=1}^{2} (\hat{W}_j^2/\kappa_j) \sum_{i=1}^{n_j} [(I_{X_{ij} \leq t} - \hat{F}_0(t))/w_{ij}]^2/n_j$ (see Appendix 3.6.2 for more details).

To test H_0 vs. H_1, we propose the maximally selected EL statistic M_n as in (3.4), except $\mathcal{R}(t)$ is now given in (3.6). The following result gives the asymptotic null distribution of M_n (see Appendix 3.6.2 for the proof).

Theorem 3.3. *Suppose* $E_{F_j}(1/w_j^2(X_{ij})) < \infty$. *Then under* H_0, *for* t_1 *and* t_2 *satisfying* $0 < F_0(t_l) < 1$ *(*$l = 1, 2$*),*

$$M_n \xrightarrow{d} \sup_{t \in [t_1, t_2]} \left[U_+^2(t)\right]$$

where $U(t)$ *is a mean 0 pinned Gaussian process with covariance* $\mathrm{cov}(U(s), U(t)) = \sigma(s,t)/\sqrt{\sigma(s,s)\sigma(t,t)}$, *and* $\sigma(s,t) = \sum_{j=1}^{2}(W_j^2/\kappa_j)E_{G_j}[(I_{X_{ij}\le s} - F_0(s))(I_{X_{ij}\le t} - F_0(t))/w_j^2(X_{ij})]$.

Remark. When there is no size bias (i.e. $w_{ij} = 1$), $U_n(t)$ reduces to $\sigma_n^{-1/2}(t,t)\sqrt{n}\times [F_{n_2 2}(t) - F_{n_1 1}(t)]$, where $F_{n_j j}(t)$ denotes the empirical cdf of the j-th sample ($j = 1, 2$) and $\sigma_n(t,t) = \sum_{j=1}^{2}(1/\kappa_j) \times \sum_{i=1}^{n_j}(I_{X_{ij}\le t} - \hat{F}_0(t))^2/n_j$. This implies that M_n is asymptotically equivalent to

$$\sup_{t \in [t_1, t_2]} \left\{\sigma_n^{-1}(t,t) n \left[F_{n_2 2}(t) - F_{n_1 1}(t)\right]_+^2\right\},$$

which is the square of the one-sided scaled version of the commonly used two-sample Kolmogorov–Smirnov statistic, $\sup_{t\in[t_1,t_2]} \left[F_{n_2 2}(t) - F_{n_1 1}(t)\right]_+$.

As in the one-sample case, for calibration we use a Gaussian multiplier bootstrap approach, as illustrated in the following subsection.

3.2.3.1 Gaussian multiplier bootstrap calibration

This section constructs multiplier bootstrap approximations of the limiting distribution in Theorem 3.3. Similar to the one-sample case, we begin with bootstrapping $U_n(t)$, whose functional $\sup_{t\in[t_1,t_2]}[U_n^2(t) I_{U_n(t)\ge 0}]$ is asymptotically equivalent to the test statistic M_n (see Section 3.2.3). Define a Gaussian multiplier bootstrap for $U_n(t)$ by

$$U_n^*(t) = \hat{\sigma}^{-\frac{1}{2}}(t,t)\left[\frac{\hat{W}_2}{\sqrt{n_2}\sqrt{\kappa_2}}\sum_{i=1}^{n_2}\xi_{i2}\frac{I_{X_{i2}\le t} - \hat{F}_0(t)}{w_{i2}} - \frac{\hat{W}_1}{\sqrt{n_1}\sqrt{\kappa_1}}\sum_{i=1}^{n_1}\xi_{i1}\frac{I_{X_{i1}\le t} - \hat{F}_0(t)}{w_{i1}}\right],$$

where ξ_{ij} ($i = 1, \ldots, n_j$ and $j = 1, 2$) are i.i.d. standard Gaussian random variables that are independent of the observed data. We show bootstrap consistency of $U_n^*(t)$, thereby establishing consistency of

$$M_n^* \equiv \sup_{t\in[t_1,t_2]} \left[U_n^{*2}(t) I_{U_n^*(t)\geq 0}\right].$$

The result is provided in the following theorem (see Appendix 3.6.3 for the proof).

Theorem 3.4. *Assume the conditions of Theorem 3.3. Then conditionally on* $X_{11}, X_{21}, \ldots, X_{12}, X_{22}, \ldots$, *for almost every sequence* $X_{11}, X_{21}, \ldots, X_{12}, X_{22}, \ldots$,

$$M_n^* \xrightarrow{d} \sup_{t\in[t_1,t_2]} \left[U_+^2(t)\right].$$

Based on Theorem 3.4, to calibrate the test, we simulate M_n^* by repeatedly generating Gaussian random samples $\{\xi_{ij}\}$ while holding the observed data fixed. We then compare the empirical quantiles of these bootstrap values M_n^* with our test statistic M_n.

3.2.4 Extensions

3.2.4.1 Two-sided test

The one-sided tests introduced in the previous sections have an immediate extension to two-sided versions. The two-sided alternative in (3.1) is the union of the two one-sided alternatives ($F_1 \succ F_2$ or $F_2 \succ F_1$). Based on the union-intersection principle, the test statistic is the maximum of the two one-sided test statistics. The asymptotic null distribution of this test statistic is $\sup_{t\in[t_1,t_2]} U^2(t)$, where $U(t)$ is the Gaussian process specified in Theorem 3.3. The test can therefore be calibrated in much the same way as we did for the one-sided test.

3.2.4.2 Stratified statistic

We have assumed the observations are homogeneous within each sample, free of confounding factors. To account for discrete confounders, we can extend M_n to a stratified statistic. More specifically, if there are L independent strata each having homogeneous level(s) of the confounder(s), the relevant hypotheses are

$$H_0' : F_1^l = F_2^l \text{ for all } l \text{ versus } H_1' : F_1^l \succ F_2^l \text{ for some } l.$$

Based on the union-intersection principle, it is natural to use a stratified statistic $\sup_{l=1,\ldots,L} M_{n,l}$, where $M_{n,l}$ is the maximally selected EL statistic in each stratum l ($l = 1, \ldots, L$). The asymptotic null distribution of the stratified statistic is just $\sup_{l=1,\ldots,L} AM_l$, where AM_l is distributed as the limiting distribution in Theorem 3.3.

3.3 Simulation study

In this section, we report the result of a simulation study for the one-sided tests. Results for the two-sided tests are similar. We investigate the performance of M_n in terms of accuracy and power.

3.3.1 Accuracy

We consider $F_1 = F_2 = Beta(4, 3)$ and two scenarios for the weight functions: (A) $w_1(x) = x$ and $w_2(x) = \sqrt{x}$, (B) $w_1(x) = \sqrt{x}$ and $w_2(x) = x$. The scenarios are illustrated in Figure 3.1. From the figure, we can see that although the underlying F_1 and F_2 are the same, the observed distributions G_1 and G_2 are different due to size bias.

[Figure 3.1 here]

It would be interesting to see what happens if size bias is ignored; that is, one mistakes G_j as F_j. To this end, we compare M_n with its counterpart that sets $w_{ij} \equiv 1$ (see the remark after Theorem 3.3), which is related to the one-sided two-sample Kolmogorov–Smirnov statistic. We denote this statistic by M_n^{ign}. Another statistic for comparison is the one-sided Wald-type statistic $\sup_{t\in[t_1,t_2]}[U_n^2(t) I_{U_n(t)\geq 0}]$

(see Section 3.2.3), with W_j $(j = 1, 2)$ and $F_0(t)$ replaced by their consistent estimate $\hat{W}_j$ and $\hat{F}_0(t)$, respectively.

The size simulation results are given in Table 3.1. We can see that the empirical significance levels of our EL test are very close to the nominal level in all the cases considered. On the other hand, the Wald test is too conservative in Scenario B. As for the test ignoring size bias, its empirical significance levels are too large in Scenario A and too small in Scenario B. We conclude that the proposed EL test has better accuracy than the other tests.

[Table 3.1 here]

3.3.2 Power comparisons

In this section, we compare the small sample power of the proposed test with its counterpart ignoring size bias and the Wald test. Two models of underlying distribution functions are considered: (C) $F_1 = Beta(4, 3)$ versus $F_2 = Beta(4, 4)$, (D) $F_1 = Beta(3, 5)$ versus $F_2 = Beta(3, 7)$. For both models, we set $w_1(x) = \sqrt{x}$ and $w_2(x) = x$. The weight functions make the difference between G_1 and G_2 smaller than the difference between F_1 and F_2, as illustrated in Figure 3.2. As a result, the test ignoring size bias (i.e. comparing G_j instead of F_j) is expected to have lower power.

[Figure 3.2 here]

The power simulation results are summarized in Table 3.2. M_n outperforms the other tests in all the cases considered. The Wald test tends to have lower power. The much lower power of M_n^{ign} is alarming; this shows the importance of taking sampling bias into account. These results show that for size biased data, the proposed EL test should be used for testing stochastically ordered alternatives.

[Table 3.2 here]

3.4 Application

We are interested in the distribution of blood alcohol concentration (BAC) of drunk drivers. Our sample consists of drivers involved in fatal car accidents. Size bias arises because drivers with higher alcohol level are more likely to be involved in an accident and have their BAC recorded.

We consider comparing BAC of young and old drivers. The bias may be different between the young and old groups, as discussed by Ramírez and Vidakovic (2010) in a similar example. They argue that another factor, lack of experience, plays a more important role in the young group than in the old group, thereby downweighting the effect of BAC on sampling in the young group. They choose the weight functions $w_y(x) = \sqrt{x}$ and $w_o(x) = x$ for the young and old groups, respectively, and acknowledge that the choice is subjective and similar in spirit to Bayesian priors. We will follow their lead and use the same weight functions in the following analysis.

The BAC data are obtained online via the Fatality Analysis Reporting System (FARS) from the U.S. National Highway Traffic Safety Administration. To ensure sample homogenuity, we restrict our analysis to whole blood test results of male drives involved in interstate highway accidents in California during 2009—these criteria are satisfied by 125 drunk drivers. Using a cutoff of 30 years old as in Ramírez and Vidakovic (2010), there are 67 young and 58 old drivers out of the 125. Although the empirical cdfs (see top panel of Figure 3.3) show there is no obvious difference between the two observed distributions, the NPMLEs $\tilde{F}_j(\cdot)$ $(j = 1, 2)$ for the underlying distribution functions (see bottom panel of Figure 3.3) suggest that the young group may be slightly stochastically larger than the old group.

[Figure 3.3 here]

Application of the one-sided EL test indicates that, indeed, the young group has slightly stochastically larger BAC values than the old group ($M_n = 4.46$, $p = 0.109$). In comparison, the one-sided Wald test gives a slightly more conservative p-value of

0.168. The test ignoring size bias yields a very large p-value of 0.841, reflecting the fact that the two empirical cdfs almost overlap. We conclude that our EL test is better adapted to detecting group difference when the data are size biased.

We have also investigated how the result changes for different choices of the power term r in $w_y(x) = x^r$. A plot of various rs and the corresponding p-values is presented in Figure 3.4. We see that as r increases, the p-value becomes larger (i.e. less significant). For $r \leq 0.4$, our EL test shows significance at the 0.05 level; for $r < 0.5$, our EL test shows significance at the 0.1 level.

[Figure 3.4 here]

3.5 Discussion

We have developed an EL-based test for stochastically ordered alternatives in size bias models. The proposed test statistic M_n is the maximally selected local EL statistic. A simulation study shows that the EL test is more powerful than its counterpart ignoring size bias and the Wald test. We applied our test to blood alcohol measurements in fatal driving accidents and found a more significant result than the Wald test and the test ignoring sampling bias.

We calibrate the proposed EL test using a Gaussian multiplier bootstrap approach. Such multiplier bootstrap has been utilized by Rémillard and Scaillet (2009) and Chernozhukov et al. (2013). Other exchangeable bootstrap procedures (see, e.g., van der Vaart and Wellner, 1996, Ch. 3.6) could also be considered. For example, the empirical (or Efron's) bootstrap for $U_n(t)$ can be defined by replacing ξ_{ij} in $U_n^*(t)$ (see Section 3.2.3.1) with $M_{n_j i} - 1$ ($i = 1, \ldots, n_j$, $j = 1, 2$), where $M_{n_j i}$ is the number of times that X_{ij} is redrawn from the original sample. (See Supplementary Material for the proof of consistency.) Here note that despite resampling of the original observations, we keep $\hat{\sigma}^{-\frac{1}{2}}(t,t)$, $\hat{W}_j$ and $\hat{F}_0(t)$ intact. This is because computing them requires solving the estimating equations for each t, which could be time consuming if we repeat such computation for each bootstrap sample. The proposed multiplier bootstrap approach also avoids such recomputation. Although the aforementioned bootstrap procedures are asymptotically equivalent, a comparison among them could be done via analyzing higher-order properties (Hall, 1992), but this is beyond the scope of the present chapter.

Our key contribution is the development of the first EL-based test for ordered underlying distribution functions in biased sampling models. We envision the test to be useful in numerous applications involving length/size bias, such as the biostatistical examples provided in Section 3.1, reliability engineering (Oluyede and George, 2002), and marketing research (Nowell and Stanley, 1991). One future direction is to derive our test based on the multiplicative censorship model, which can be applied to prevalence cohort studies (Ning et al., 2013). Another direction is to deal with the situation where, in addition to the sample observed from G_j, we also have a random sample observed from F_j $(j = 1, 2)$. One possible test for stochastic ordering in this case is to use a convex combination of the two statistics M_n and M_m^{ign}, where M_n is computed based on samples (of total size n) from G_1 and G_2 and M_m^{ign} based on samples (of total size m) from F_1 and F_2.

3.6 Appendices

3.6.1 Derivation of the numerator of the local EL ratio

We derive $\sup\{L(F_1, F_2) : F_1(t) = F_2(t)\}$ for the two-sample case. The one-sample case is similar and the proof is omitted.

We first optimize $-\log L(F_1, F_2) = -\sum_{j=1}^{2}\sum_{i=1}^{n_j}\log p_{ij} + \sum_{j=1}^{2} n_j \log(\sum_{i=1}^{n_j} w_{ij}p_{ij})$ subject to the constraints $\sum_{i=1}^{n_j} p_{ij} = 1$, $\sum_{i=1}^{n_j} p_{ij}(I_{X_{ij}\le t} - F_0(t)) = 0$, and $\sum_{i=1}^{n_j} p_{ij}(w_{ij} - W_j) = 0$ for fixed W_j and $F_0(t)$, $j = 1, 2$. This is similar to the usual empirical likelihood for estimating equations, except that now the likelihood is of a weighted form.

The Lagrangian can be defined as a function $\mathcal{L} : [0,1]^n \times \mathbb{R}^3 \times [0,\infty)^2 \times [0,1] \to \mathbb{R}$ such that

$$\begin{aligned}
&\mathcal{L}(p_{11},\ldots,p_{n_1 1},p_{12},\ldots,p_{n_2 2},a_1,a_2,\lambda,W_1,W_2,F_0(t)) \\
&\equiv -\sum_{j=1}^{2}\sum_{i=1}^{n_j}\log p_{ij} + \sum_{j=1}^{2} n_j \log W_j + n\sum_{j=1}^{2} a_j \left(\sum_{i=1}^{n_j} p_{ij} - 1\right) \\
&\quad + n\sum_{j=1}^{2}\lambda_{1j}\left\{\sum_{i=1}^{n_j}\left(I_{X_{ij}\le t} - F_0(t)\right)p_{ij}\right\} + n\sum_{j=1}^{2}\lambda_{2j}\left\{\sum_{i=1}^{n_j}\left(w_{ij} - W_j\right)p_{ij}\right\}.
\end{aligned}$$

The solution is

$$\hat{p}_{ij} = \frac{1}{n}\frac{1}{\hat{a}_j + \hat{\lambda}_{1j}\left(I_{X_{ij}\le t} - F_0(t)\right) + \hat{\lambda}_{2j}\left(w_{ij} - W_j\right)},$$

where the Lagrange multipliers $\hat{a}_j, \hat{\lambda}_{1j}, \hat{\lambda}_{2j}$ $(j = 1, 2)$ satisfy

$$\begin{aligned}
\sum_{i=1}^{n_j}\frac{1}{n}\frac{1}{\hat{a}_j + \hat{\lambda}_{1j}\left(I_{X_{ij}\le t} - F_0(t)\right) + \hat{\lambda}_{2j}\left(w_{ij} - W_j\right)} &= 1, \\
\sum_{i=1}^{n_j}\frac{I_{X_{ij}\le t} - F_0(t)}{\hat{a}_j + \hat{\lambda}_{1j}\left(I_{X_{ij}\le t} - F_0(t)\right) + \hat{\lambda}_{2j}\left(w_{ij} - W_j\right)} &= 0, \\
\sum_{i=1}^{n_j}\frac{w_{ij} - W_j}{\hat{a}_j + \hat{\lambda}_{1j}\left(I_{X_{ij}\le t} - F_0(t)\right) + \hat{\lambda}_{2j}\left(w_{ij} - W_j\right)} &= 0.
\end{aligned}$$

Note that the dependence of the solution on t is omitted for simplicity.

Denote the resulting profile log-likelihood as

$$l(W_1, W_2, F_0(t)) = \sum_{j=1}^{2}\sum_{i=1}^{n_j}\log(\hat{p}_{ij}) - \sum_{j=1}^{2} n_j \log(W_j) + \sum_{j=1}^{2}\sum_{i=1}^{n_j}\log(w_{ij}).$$

We then optimize l over $(W_1, W_2, F_0(t))$. This leads to $\hat{\lambda}_{2j} = \kappa_j/\hat{W}_j$ and $\hat{\lambda}_{11} = -\hat{\lambda}_{12} \equiv \hat{\lambda}$, based on which we can re-write $\hat{p}_{ij}$ as

$$\frac{1}{n}\frac{1}{\hat{a}_j + \hat{\lambda}\Delta_j\left(I_{X_{ij}\le t} - \hat{F}_0(t)\right) + \frac{\kappa_j w_{ij}}{\hat{W}_j} - \kappa_j},$$

where $\Delta_j = 1$ for $j = 1$ and -1 for $j = 2$, and $(\hat{a}_1, \hat{a}_2, \hat{W}_1, \hat{W}_2, \hat{\lambda}, \hat{F}_0(t))$ satisfy (3.7).

3.6.2 Proof of Theorem 3.3

We first re-express a few quantities to facilitate proving the asymptotics. We simplify $\hat{p}_{ij}$ using the fact that $\sum_{i=1}^{n_j} p_{ij} \frac{\partial \mathcal{L}}{\partial p_{ij}} = 0$ implies $\hat{a}_j = \kappa_j$. This gives

$$\begin{aligned}\hat{p}_{ij} =& \frac{1}{n} \frac{1}{\frac{\kappa_j w_{ij}}{\hat{W}_j} + \hat{\lambda}\Delta_j \left(I_{X_{ij} \le t} - \hat{F}_0(t)\right)} \\ =& \frac{1}{n\hat{\eta}_{ij}} \frac{1}{1 + \hat{\lambda}\Delta_j g_{1ij}(\hat{F}_0(t), \hat{W}_j)},\end{aligned} \tag{3.9}$$

where $\hat{\eta}_{ij} = (\kappa_j w_{ij})/\hat{W}_j$ and $g_{1ij}(\hat{F}_0(t), \hat{W}_j) = (I_{X_{ij} \le t} - \hat{F}_0(t))/\hat{\eta}_{ij}$. We then show that $\hat{\lambda} \le 0$ iff $\tilde{F}_1(t) \le \tilde{F}_2(t)$ (see Supplementary Material (5.9)). This allows us to simplify $-2\log \mathcal{R}(t)$ as

$$2\sum_{j=1}^{2}\sum_{i=1}^{n_j} \log\left(1 + \hat{\lambda}\Delta_j g_{1ij}(\hat{F}_0(t), \hat{W}_j)\right) I_{\hat{\lambda} \le 0}. \tag{3.10}$$

Also, we re-write the last two estimating equations in (3.7) as

$$\begin{aligned}&Q_{1j}(\hat{F}_0(t), \hat{W}_j, \hat{\lambda}) = 0 \text{ and} \\ &Q_{2j}(\hat{F}_0(t), \hat{W}_j, \hat{\lambda}) = 0,\end{aligned} \tag{3.11}$$

for $j = 1, 2$, where $Q_{1j}(\hat{F}_0(t), \hat{W}_j, \hat{\lambda}) \equiv 1/n \times \sum_{i=1}^{n_j}\{g_{1ij}(\hat{F}_0(t), \hat{W}_j)/[1 + \Delta_j\hat{\lambda} g_{1ij}(\hat{F}_0(t), \hat{W}_j)]\}$, $Q_{2j}(\hat{F}_0(t), \hat{W}_j, \hat{\lambda}) \equiv 1/n \times \sum_{i=1}^{n_j}\{g_{2ij}(\hat{F}_0(t), \hat{W}_j)/[1 + \Delta_j\hat{\lambda} g_{1ij}(\hat{F}_0(t), \hat{W}_j)]\}$, and $g_{2ij}(\hat{F}_0(t), \hat{W}_j) = (w_{ij} - \hat{W}_j)/\hat{\eta}_{ij}$.

Using the estimating equations in (3.11), we show that $\left|\hat{F}_0(t) - F_0(t)\right|$, $\left|\hat{\lambda}\right|$ and $\left|\hat{W}_j - W_j\right|$ (for $j = 1, 2$) are $O_p(n^{-1/2})$ (see Supplementary Material). Here and in the sequel, the asymptotic o_p and O_p terms hold uniformly for $t \in [t_1, t_2]$. Based on these asymptotic orders, we apply Taylor's theorem to (3.10) and get

$$-2\log \mathcal{R}(t) = 2\sum_{j=1}^{2}\left[\hat{\lambda}\Delta_j \sum_{i=1}^{n_j} g_{1ij}(\hat{F}_0(t), \hat{W}_j) - \frac{\hat{\lambda}^2}{2}\sum_{i=1}^{n_j} g_{1ij}^2(\hat{F}_0(t), \hat{W}_j)\right] I_{\hat{\lambda} \le 0} + o_p(1). \tag{3.12}$$

We also expand $Q_{1j}(\hat{F}_0(t), \hat{W}_j, \hat{\lambda})$ (for $j = 1, 2$) around $(\hat{F}_0(t), \hat{W}_j, 0)$ and get

$$\begin{aligned} 0 =& Q_{1j}(\hat{F}_0(t), \hat{W}_j, 0) + \frac{\partial Q_{1j}(\hat{F}_0(t), \hat{W}_j, 0)}{\partial \lambda}(\hat{\lambda} - 0) + o_p(\left|\hat{\lambda}\right|) \\ =& \frac{1}{n}\sum_{i=1}^{n_j} g_{1ij}(\hat{F}_0(t), \hat{W}_j) - \frac{\Delta_j}{n}\sum_{i=1}^{n_j} g_{1ij}^2(\hat{F}_0(t), \hat{W}_j)\hat{\lambda} + o_p(n^{-\frac{1}{2}}), \end{aligned}$$

which implies

$$\hat{\lambda} = \left[\sum_{j=1}^{2}\frac{1}{n}\sum_{i=1}^{n_j} g_{1ij}^2(\hat{F}_0(t), \hat{W}_j)\right]^{-1}\left[\sum_{j=1}^{2}\Delta_j\frac{\hat{W}_j}{n_j}\sum_{i=1}^{n_j}\frac{I_{X_{ij}\le t} - \hat{F}_0(t)}{w_{ij}}\right] + o_p(n^{-\frac{1}{2}}) \quad (3.13)$$

and $\hat{\lambda}\Delta_j \sum_{i=1}^{n_j} g_{1ij}(\hat{F}_0(t), \hat{W}_j) = \hat{\lambda}^2 \sum_{i=1}^{n_j} g_{1ij}^2(\hat{F}_0(t), \hat{W}_j) + o_p(1)$. Substituting the latter into (3.12) gives

$$-2\log\mathcal{R}(t) = 2\sum_{j=1}^{2}\left[\frac{\hat{\lambda}^2}{2}\sum_{i=1}^{n_j} g_{1ij}^2(\hat{F}_0(t), \hat{W}_j)\right] I_{\hat{\lambda}\le 0} + o_p(1).$$

This, (3.13), $\left|\hat{F}_0(t) - F_0(t)\right| = O_p(n^{-1/2})$ and $\left|\hat{W}_j - W_j\right| = O_p(n^{-1/2})$ imply

$$-2\log\mathcal{R}(t) = U_n^2(t) I_{U_n(t)\ge 0} + o_p(1), \quad (3.14)$$

where $U_n(t)$ is defined as

$$\left[\sum_{j=1}^{2}\frac{1}{n}\sum_{i=1}^{n_j} g_{1ij}^2(\hat{F}_0(t), \hat{W}_j))\right]^{-\frac{1}{2}}\left[\sum_{j=1}^{2}(-\Delta_j)\frac{W_j}{\sqrt{n_j}\sqrt{\kappa_j}}\sum_{i=1}^{n_j}\frac{I_{X_{ij}\le t} - F_0(t)}{w_{ij}}\right]. \quad (3.15)$$

Based on (3.14), we can obtain the limiting distribution of $-2\log\mathcal{R}(t)$ by studying $U_n(t)$. We begin by finding the weak convergence of the second term in (3.15). By Donsker's Theorem, it can be shown that for each $j = 1, 2$, $(W_j/\sqrt{n_j\kappa_j}) \times \sum_{i=1}^{n_j}[(I_{X_{ij}\le t} - F_0(t))/w_{ij}]$ converges in distribution to a Gaussian process with zero mean and covariance function

$$\frac{(W_j)^2}{\kappa_j} E_{G_j}\left(\frac{I_{X_{ij}\le s} - F_0(s)}{w_j(X_{ij})}\frac{I_{X_{ij}\le t} - F_0(t)}{w_j(X_{ij})}\right)$$

in $l^\infty([t_1, t_2])$. Therefore, by independence between the two samples and the continuous mapping theorem, $\sum_{j=1}^{2}(-\Delta_j)(W_j/\sqrt{n_j\kappa_j})\sum_{i=1}^{n_j}[(I_{X_{ij}\le t} - F_0(t))/w_{ij}]$ converges in distribution to a Gaussian process with zero mean and covariance function being

$$\sigma(s,t) \equiv \sum_{j=1}^{2} \frac{(W_j)^2}{\kappa_j} E_{G_j} \left(\frac{I_{X_{ij} \le s} - F_0(s)}{w_j(X_{ij})} \frac{I_{X_{ij} \le t} - F_0(t)}{w_j(X_{ij})} \right).$$

On the other hand,

$$\sum_{j=1}^{2} \frac{1}{n} \sum_{i=1}^{n_j} g_{1ij}^2(\hat{F}_0(t), \hat{W}_j)) = \sigma(t,t) + o_p(1) \tag{3.16}$$

by Glivenko-Cantelli's Theorem and the fact that $\left|\hat{F}_0(t) - F_0(t)\right| = O_p(n^{-1/2})$ and $\left|\hat{W}_j - W_j\right| = O_p(n^{-1/2})$. Then by Slutsky's Lemma and (3.15), we have

$$U_n(t) \xrightarrow{d} U(t)$$

in $l^\infty([t_1, t_2])$, where $U(t)$ is a mean 0 Gaussian process with

$$\operatorname{cov}(U(s), U(t)) = \frac{\sigma(s,t)}{\sqrt{\sigma(s,s)}\sqrt{\sigma(t,t)}}.$$

This, (3.14) and the continuous mapping theorem imply

$$-2 \log \mathcal{R}(t) \xrightarrow{d} U_+^2(t).$$

Then applying continuous mapping theorem again, we obtain the desired result.

Remark. By (3.15) and (3.16), we have that $U_n(t)$ is asymptotically equivalent to

$$U_n^{**}(t) = \sigma(t,t)^{-\frac{1}{2}} \left[\sum_{j=1}^{2} (-\Delta_j) \frac{W_j}{\sqrt{n_j}\sqrt{\kappa_j}} \sum_{i=1}^{n_j} \frac{I_{X_{ij} \le t} - F_0(t)}{w_{ij}} \right].$$

This will be used in the following section for proving bootstrap consistency of $U_n^*(t)$ in Section 3.2.3.1.

3.6.3 Gaussian multiplier bootstrap consistency

To show bootstrap consistency of M_n^*, we start with $U_n^*(t)$. It is easier to first obtain bootstrap consistency of a similar process, $U_n^{**}(t)$ (see the remark above in Section

3.6.2). Next we establish asymptotic equivalence of $U_n^*(t)$ and $U_n^{**}(t)$, conditionally on the data almost surely. This implies bootstrap consistency of $U_n^*(t)$. Lastly, by continuous mapping theorem, we obtain the desired result for M_n^*.

To prove bootstrap consistency of $U_n^{**}(t)$, we make use of the multiplier central limit theorem. Specifically, we first show that for the j-th sample, the class $\mathcal{F}_j$ of functions

$$\left\{ f_{jt}(u) \equiv -\sigma(t,t)^{-\frac{1}{2}} \left[\Delta_j \frac{W_j}{\sqrt{\kappa_j}} \frac{I_{u\le t} - F_0(t)}{w_j(u)} \right], t \in [t_1, t_2] \right\}$$

is Donsker. This follows by the Donsker Preservation property (see, e.g., Kosorok, 2008, Corollary 9.32) since the class $\{I_{u\le t} - F_0(t), t \in [t_1, t_2]\}$ is Donsker, and

$$-\sigma(t,t)^{-\frac{1}{2}} \times \Delta_j \frac{W_j}{\sqrt{\kappa_j}} \frac{1}{w_j(u)}$$

is uniformly bounded on $[t_1, t_2]$. Secondly, we obtain $P\,\|f_{jt} - Pf_{jt}\|^2_{\mathcal{F}_j} < \infty$ by the assumption of finite $E_{F_j}(1/w_j^2(X_{ij}))$ in the theorem. These results and the multiplier central limit theorem (see, e.g., van der Vaart and Wellner, 1996, Theorem 2.9.7) then imply that conditionally on the data, $\sum_{i=1}^{n_j} \xi_{ij} f_{jt}(X_{ij})/\sqrt{n_j}$ converges in distribution to a Gaussian process with zero mean and covariance function being

$$[\sigma(s,s)\sigma(t,t)]^{-\frac{1}{2}} \frac{(W_j)^2}{\kappa_j} E_{G_j} \left(\frac{I_{X_{ij}\le s} - F_0(s)}{w_j(X_{ij})} \frac{I_{X_{ij}\le t} - F_0(t)}{w_j(X_{ij})} \right).$$

almost surely. Finally, by independence between the two samples and the continuous mapping theorem, we have that $U_n^{**}(t)$ converges conditionally in distribution to $U(t)$, given $X_{11}, X_{21}, \ldots, X_{12}, X_{22}, \ldots$, almost surely.

Now we show that conditionally on the data, $U_n^*(t)$ is asymptotically equivalent to $U_n^{**}(t)$ almost surely. That is, for all $\varepsilon > 0$

$$P_\xi \left(\sup_{t\in[t_1,t_2]} |U_n^*(t) - U_n^{**}(t)| > \varepsilon \middle| X_{11}, X_{21}, \ldots, X_{12}, X_{22}, \ldots \right) \xrightarrow{a.s.} 0.$$

Chebyshev's inequality enables us to prove the result above by showing that

$$
\begin{aligned}
&E_\xi\left(\sup_{t\in[t_1,t_2]}|U_n^*(t)-U_n^{**}(t)|^2\middle|X_{11},X_{21},\dots,X_{12},X_{22},\dots\right)\\
&\le\sum_{j=1}^{2}\frac{1}{n_j\kappa_j}\sum_{i=1}^{n_j}\frac{E_\xi\left(\xi_{ij}^2\right)}{w_{ij}^2}\sup_{t\in[t_1,t_2]}\left|\frac{\hat{W}_j\left(I_{X_{ij}\le t}-\hat{F}_0(t)\right)}{\sqrt{\hat{\sigma}(t,t)}}-\frac{W_j\left(I_{X_{ij}\le t}-F_0(t)\right)}{\sqrt{\sigma(t,t)}}\right|\\
&=\sum_{j=1}^{2}\frac{1}{n_j\kappa_j}\sum_{i=1}^{n_j}\frac{1}{w_{ij}^2}O(n^{-1})\xrightarrow{a.s.}0,
\end{aligned}
$$

where the $O(n^{-1})$ is due to the strong consistency of $\hat{W}_j$, $\hat{F}_0(t)$ and $\hat{\sigma}(t,t)^{-1/2}$ (see the remark in Section 5.3.2 of Supplementary Material).

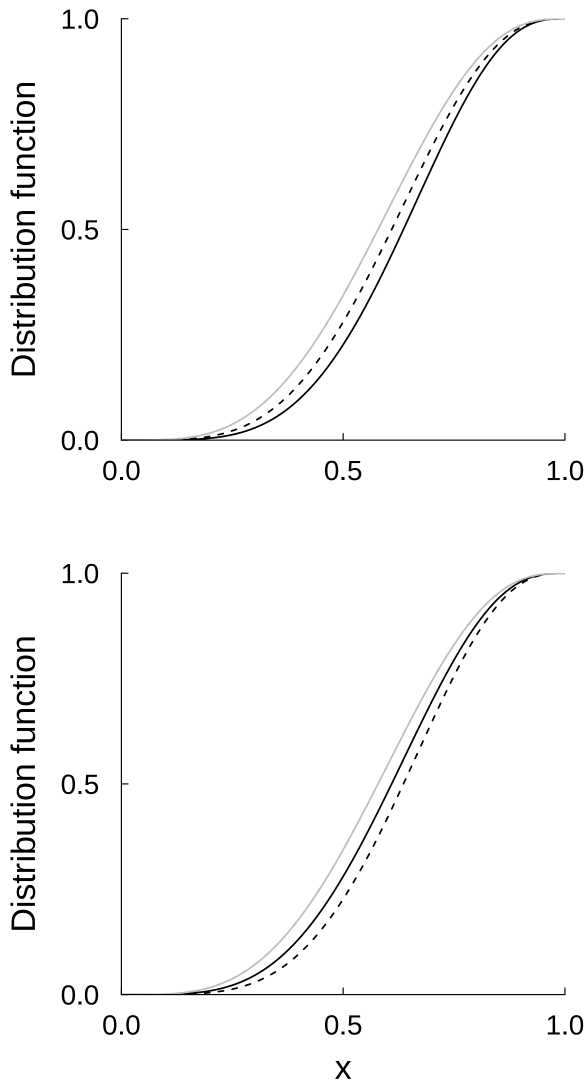

Figure 3.1: For computing empirical levels, the underlying (gray) and weighted (black) distribution functions in Scenario A (top) and Scenario B (bottom): F_1 and G_1 (solid) versus F_2 and G_2 (dashed).

Table 3.1: Empirical significance levels based on 10,000 replications, each with 1000 bootstrap samples. **Scenario A**: distribution functions displayed in Figure 3.1, upper panel. **Scenario B**: distribution functions displayed in Figure 3.1, lower panel.

Scenario	group	$\alpha = 0.05$			$\alpha = 0.01$		
	size	M_n	M_n^{ign}	Wald	M_n	M_n^{ign}	Wald
A	50	0.053	0.153	0.053	0.012	0.044	0.012
	80	0.052	0.192	0.055	0.010	0.059	0.010
B	50	0.054	0.012	0.032	0.011	0.002	0.005
	80	0.055	0.010	0.032	0.011	0.001	0.005

Table 3.2: Power simulation results based on 10,000 replications, each with 1000 bootstrap samples. **Scenario C**: distribution functions displayed in Figure 3.2, first column. **Scenario D**: distribution functions displayed in Figure 3.2, second column.

Scenario	group	$\alpha = 0.05$			$\alpha = 0.01$		
	size	M_n	M_n^{ign}	Wald	M_n	M_n^{ign}	Wald
C	50	0.600	0.345	0.524	0.329	0.132	0.242
	80	0.791	0.484	0.736	0.530	0.229	0.440
D	50	0.757	0.405	0.674	0.494	0.176	0.365
	80	0.906	0.561	0.858	0.722	0.290	0.619

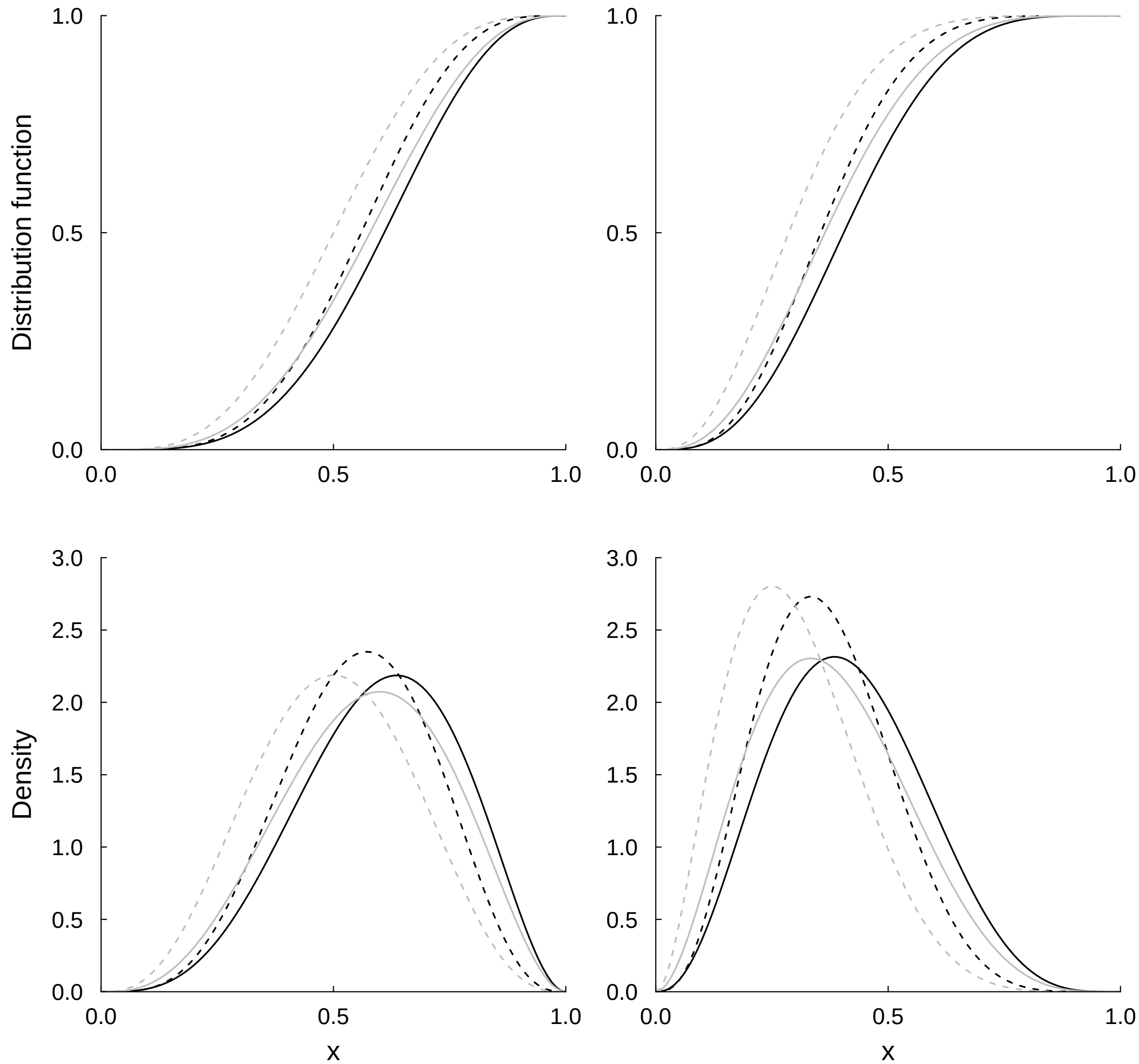

Figure 3.2: For power comparisons, the underlying (gray) and weighted (black) distribution (top row) and density (bottom row) functions in Scenario C (first column) and Scenario D (second column): F_1 and G_1 (solid) versus F_2 and G_2 (dashed).

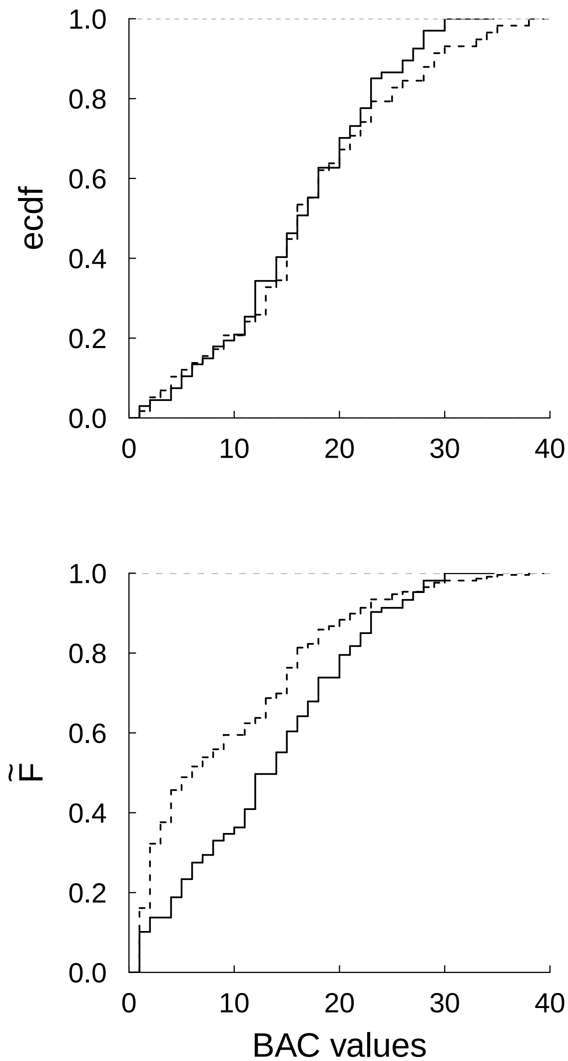

Figure 3.3: The empirical cdf (top) and NPMLE for the underlying distribution function (bottom) of BAC values for drivers of age less than 30 (solid) and at least 30 (dashed).

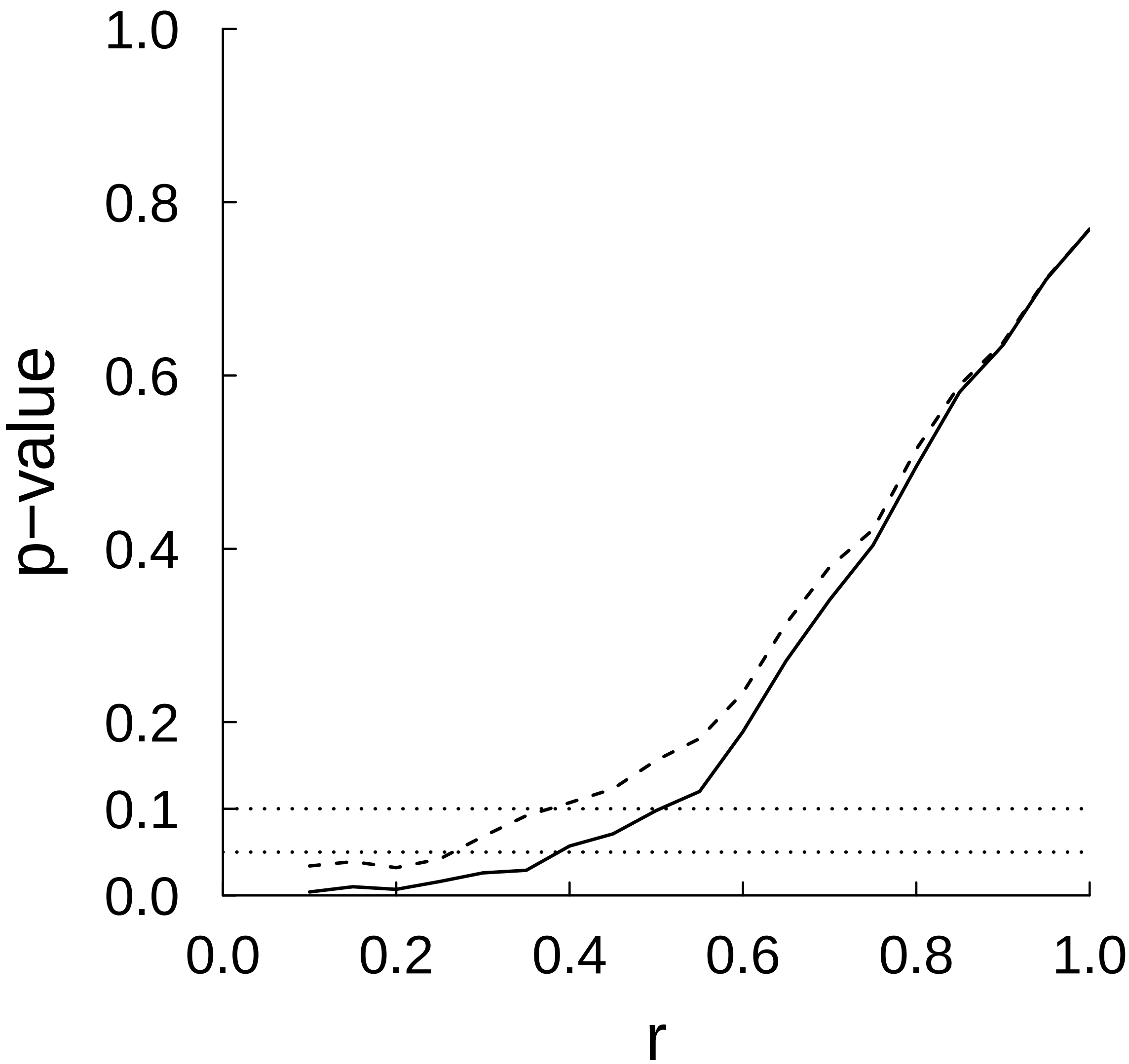

Figure 3.4: The p-value for M_n (solid) and Wald test (dashed) when the power term r in $w_y(x) = x^r$ changes. The horizontal dotted lines indicate the 0.05 and 0.1 significance levels.

Chapter 4

Bibliography

Bibliography

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer.

Andrews, D. W. K. and Guggenberger, P. (2009). Validity of subsampling and plug-in asymptotic inference for parameters defined by moment inequalities. *Econometric Theory*, 25:669–709.

Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.

Canay, I. A. (2010). EL inference for partially identified models: Large deviations optimality and bootstrap validity. *Journal of Econometrics*, 156(2):408–425.

Chacko, V. J. (1963). Testing homogeneity against ordered alternatives. *The Annals of Mathematical Statistics*, 34(3):945–956.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Annals of Statistics*, 41(6):2786–2819.

Chi, Y. (2002). Theory & methods: Ordered tests for right-censored survival data. *Australian & New Zealand Journal of Statistics*, 44(3):367–380.

Clark, A. G., Hubisz, M. J., Bustamante, C. D., Williamson, S. H., and Nielsen,

R. (2005). Ascertainment bias in studies of human genome-wide polymorphism. *Genome Research*, 15(11):1496–1502.

Crowley, J. (1979). Some extensions of the log rank test. In *Clinical Trials in 'Early' Breast Cancer*. New York: Springer-Verlag. Proceedings of a Symposium, Heidelberg, Germany.

Davidov, O., Fokianos, K., and Iliopoulos, G. (2010). Order-restricted semiparametric inference for the power bias model. *Biometrics*, 66(2):549–557.

Davidov, O. and Herman, A. (2009). New tests for stochastic order with application to case control studies. *Journal of Statistical Planning and Inference*, 139(8):2614–2623.

Davidov, O. and Herman, A. (2010). Testing for order among K populations: theory and examples. *Canadian Journal of Statistics*, 38(1):97–115.

Davidov, O. and Herman, A. (2012). Ordinal dominance curve based inference for stochastically ordered distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(5):825–847.

Davidov, O. and Iliopoulos, G. (2009). On the existence and uniqueness of the NPMLE in biased sampling models. *Journal of Statistical Planning and Inference*, 139(2):176–183.

Duffy, S. W., Nagtegaal, I. D., Wallis, M., Cafferty, F. H., Houssami, N., Warwick, J., Allgood, P. C., Kearins, O., Tappenden, N., O'Sullivan, E., and Lawrence, G. (2008). Correcting for lead time and length bias in estimating the effect of screen detection on cancer survival. *American Journal of Epidemiology*, 168(1):98–104.

Dykstra, R. L. (1982). Maximum likelihood estimation of the survival functions of stochastically ordered random variables. *Journal of the American Statistical Association*, 77(379):621–628.

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

Einmahl, J. H. J. and McKeague, I. W. (2003). Empirical likelihood based hypothesis testing. *Bernoulli*, 9(2):267–290.

El Barmi, H. (1996). Empirical likelihood ratio test for or against a set of inequality constraints. *Journal of Statistical Planning and Inference*, 55(2):191–204.

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

El Barmi, H. and Mukerjee, H. (2005). Inferences under a stochastic ordering constraint: The k-sample case. *Journal of the American Statistical Association*, 100(469):252–261.

El Barmi, H. and Rothmann, M. (1998). Nonparametric estimation in selection biased models in the presence of estimating equations. *Journal of Nonparametric Statistics*, 9(4):381–399.

Fisher, R. A. (1934). The effect of methods of ascertainment upon the estimation of frequencies. *Annals of Eugenics*, 6(1):13–25.

Franck, W. E. (1984). A likelihood ratio test for stochastic ordering. *Journal of the American Statistical Association*, 79(387):686–691.

Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily single-censored samples. *Biometrika*, 52:203–223.

Gilbert, P. B., Lele, S. R., and Vardi, Y. (1999). Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials. *Biometrika*, 86(1):27–43.

Gill, R. D. (1980). *Censoring and Stochastic Integrals*. Mathematisch Centrum.

Gill, R. D., Vardi, Y., and Wellner, J. A. (1988). Large sample theory of empirical distributions in biased sampling models. *The Annals of Statistics*, 16(3):1069–1112.

Green, S. (1979). *Estimation and Testing of Location for Arbitrarily Right-Censored Data*. PhD thesis, Dept. of Statistics, University of Wisconsin.

Guyot, P., Ades, A. E., Ouwens, M. J. N. M., and Welton, N. J. (2012). Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan–Meier survival curves. *BMC Medical Research Methodology*, 12(1):9.

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

Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review / Revue Internationale de Statistique*, 58(2):109–127.

Hollander, M., McKeague, I. W., and Yang, J. (1997). Likelihood ratio-based confidence bands for survival functions. *Journal of the American Statistical Association*, 92:215–226.

Jonckheere, A. R. (1954). A distribution-free k-sample test against ordered alternatives. *Biometrika*, 41(1/2):133–145.

Kitamura, Y., Santos, A., and Shaikh, A. M. (2012). On the asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica*, 80(1):413–423.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer: New York.

Li, G. (1995). On nonparametric likelihood ratio estimation of survival probabilities for censored data. *Statistics and Probability Letters*, 25:95–104.

Liu, P. Y., Green, S., Wolf, M., and Crowley, J. (1993). Testing against ordered alternatives for censored survival data. *Journal of the American Statistical Association*, 88(421):153–160.

Liu, P. Y. and Tsai, W. Y. (1999). A modified logrank test for censored survival data under order restrictions. *Statistics & Probability Letters*, 41:57–63.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer chemotherapy reports*, 50(3):163–170.

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 (Methodological)*, 50(3):403–412.

McKeague, I. W. and Zhao, Y. (2002). Simultaneous confidence bands for ratios of survival functions via empirical likelihood. *Statistics & Probability Letters*, 60:405–415.

Mukerjee, R. (1994). Comparison of tests in their original forms. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 56(1):118–127.

Murphy, S. A. (1995). Likelihood ratio-based confidence intervals in survival analysis. *Journal of the American Statistical Association*, 90(432):1399–1405.

Nguyen-Khac, E., Thevenot, T., Piquet, M.-A., Benferhat, S., Goria, O., Chatelain, D., Tramier, B., Dewaele, F., Ghrib, S., Rudler, M., Carbonell, N., Tossou, H., Bental, A., Bernard-Chabert, B., and Dupas, J.-L. (2011). Glucocorticoids plus *N*-acetylcysteine in severe alcoholic hepatitis. *New England Journal of Medicine*, 365(19):1781–1789.

Ning, J., Qin, J., Asgharian, M., and Shen, Y. (2013). Empirical likelihood-based confidence intervals for length-biased data. *Statistics in Medicine*, 32(13):2278–2291.

Nowell, C. and Stanley, L. R. (1991). Length-biased sampling in mall intercept surveys. *Journal of Marketing Research*, 28(4):475–479.

Oluyede, B. O. and George, E. O. (2002). On stochastic inequalities and comparisons of reliability measures for weighted distributions. *Mathematical Problems in Engineering*, 8(1):1–13.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249.

Owen, A. B. (2001). *Empirical likelihood*. Chapman & Hall/CRC.

Park, Y., Kalbfleisch, J. D., and Taylor, J. M. G. (2012a). Constrained nonparametric maximum likelihood estimation of stochastically ordered survivor functions. *Canadian Journal of Statistics*, 40(1):22–39.

Park, Y., Taylor, J. M. G., and Kalbfleisch, J. D. (2012b). Pointwise nonparametric maximum likelihood estimator of stochastically ordered survivor functions. *Biometrika*, 99(2):327–343.

Patil, G. P. and Rao, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, 34(2):179–189.

Pepe, M. S. and Fleming, T. R. (1989). Weighted Kaplan–Meier statistics: a class of distance tests for censored survival data. *Biometrics*, 45(2):497–507.

Pepe, M. S. and Fleming, T. R. (1991). Weighted Kaplan–Meier statistics: Large sample and optimality considerations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(2):341–352.

Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures (with discussion). *Journal of the Royal Statistical Society, Series A*, 135:185–206.

Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika*, 65(1):167–179.

Qin, J. (1993). Empirical likelihood in biased sample problems. *The Annals of Statistics*, 21(3):1182–1196.

Ramírez, P. and Vidakovic, B. (2010). Wavelet density estimation for stratified size-biased sample. *Journal of Statistical Planning and Inference*, 140(2):419–432.

Rémillard, B. and Scaillet, O. (2009). Testing for equality between two copulas. *Journal of Multivariate Analysis*, 100(3):377–386.

Robertson, T. and Wright, F. T. (1981). Likelihood ratio tests for and against a stochastic ordering between multinomial populations. *The Annals of Statistics*, 9(6):1248–1257.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer Science+Business Media, LLC.

Silvapulle, M. J. and Sen, P. K. (2005). *Constrained Statistical Inference*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Terpstra, T. J. (1952). The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. *Indagationes Mathematicae*, 14:327–333.

Thomas, D. R. and Grunkemeier, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. *Journal of the American Statistical Association*, 70:865–871.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge Series on Statistical and Probabilistic Mathematics. Cambridge University Press.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer-Verlag, New York.

Vardi, Y. (1982). Nonparametric estimation in the presence of length bias. *The Annals of Statistics*, 10(2):616–620.

Vardi, Y. (1985). Empirical distributions in selection bias models. *The Annals of Statistics*, 13(1):178–203.

Wang, Q.-H. and Jing, B.-Y. (2001). Empirical likelihood for a class of functionals of survival distribution with censored data. *Annals of the Institute of Statistical Mathematics*, 53:517–527.

Wang, Y. (1996). A likelihood ratio test against stochastic ordering in several populations. *Journal of the American Statistical Association*, 91(436):1676–1683.

Yu, W., El Barmi, H., and Ying, Z. (2011). Restricted one way analysis of variance using the empirical likelihood ratio test. *Journal of Multivariate Analysis*, 102(3):629–640.

Zelen, M. and Feinleib, M. (1969). On the theory of screening for chronic diseases. *Biometrika*, 56(3):601–614.

Chapter 5

Supplementary materials

5.1 Preliminaries: empirical likelihood in the usual setting

This section provides background on empirical likelihood (EL) in the usual setting, when there is no censoring nor sampling bias.

The EL is based on the nonparametric likelihood function (of a cdf F)

$$L(F) = \prod_{i=1}^{n} (F(X_i) - F(X_i-))$$

given $X_1, \ldots, X_n$ i.i.d. from some unknown cdf (assuming the X_i's are one-dimensional, for simplicity of illustration). To make inference about a parameter of interest, the idea of the likelihood ratio test is used. For example, if the parameter of interest is the mean $\mu \equiv E(X_1)$, the EL ratio is constructed for testing $H_0 : \mu = \mu_0$ versus $H_0 : \mu \neq \mu_0$:

$$\begin{aligned} \mathcal{R}(\mu_0) &= \frac{\sup\{\prod_{i=1}^{n} w_i | \sum_{i=1}^{n} w_i X_i = \mu_0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1\}}{\sup\{\prod_{i=1}^{n} w_i | w_i \geq 0, \sum_{i=1}^{n} w_i = 1\}} \\ &= \frac{\sup\{\prod_{i=1}^{n} w_i | \sum_{i=1}^{n} w_i X_i = \mu_0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1\}}{\left(\frac{1}{n}\right)^n} \\ &= \sup\left\{\prod_{i=1}^{n} n w_i | \sum_{i=1}^{n} w_i X_i = \mu_0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1\right\}. \end{aligned}$$

Note that in the denominator, the nonparametric likelihood is maximized globally when an equal weight of $1/n$ is placed on each observation X_i (i.e., the corresponding cdf is the empirical cdf). It has been shown (see, e.g., Owen, 2001, p. 16) that if $0 < \text{Var}(X_i) < \infty$, then

$$-2\log\left(\mathcal{R}(\mu_0)\right) \xrightarrow{d} \chi^2_{(1)} \tag{5.1}$$

as $n \to \infty$. This is a nonparametric version of the Wilks' theorem. Based on (5.1), hypothesis testing of H_0 vs. H_1 can be conducted, and a level $1-\alpha$ confidence interval for μ can be constructed by inverting the test:

$$\left\{\mu_0 : -2\log\left(\mathcal{R}(\mu_0)\right) \leq \chi^{2,1-\alpha}_{(1)}\right\},$$

where $\chi^{2,1-\alpha}_{(1)}$ is the $(1-\alpha)$-th quantile of the chi-square distribution $\chi^2_{(1)}$.

5.2 Supplementary Materials for "Empirical likelihood based tests for stochastic ordering under right censorship"

5.2.1 Supplementary Table

Table 5.1 provides results on the power comparison under unequal uniform or exponential censoring.

[Table 5.1 here]

5.2.2 Supplementary R functions

We provide documentation and R code for two R functions that compute critical values and implement the one-sided/two-sided EL tests.

5.2.2.1 Critical value computation

Name of function

`critical_value_compute`

Description

Compute critical values for the one-sided/two-sided two-sample EL tests.

Usage

`critical_value_compute(x1, x2, one_sided = 1, alpha = 0.05, file = '', nsimu = 10000, wgrid = 100000, seed = 1370)`

Arguments

`x1`	a pre-specified x_1 number ($0 < x_1 < 1$).
`x2`	a pre-specified x_2 number ($x_1 < x_2 < 1$).
`one_sided`	1 if one-sided; 0 if two-sided.
`alpha`	vector of significance levels of interest.
`file`	a character string naming a file to store the simulated values of $\sup\{B_+^2(x)/(x(1-x))\}$ (if `one_sided = 1`) or $\sup\{B^2(x)/(x(1-x))\}$ (if `one_sided = 0`); this is for the next function that implement the one-sided/two-sided EL tests. `''` indicates no storing.
`nsimu`	number of values of $\sup\{B_+^2(x)/(x(1-x))\}$ (if `one_sided = 1`) or $\sup\{B^2(x)/(x(1-x))\}$ (if `one_sided = 0`) to be simulated.
`wgrid`	number of grid points on $[0, 1]$ where the Brownian bridge is simulated.
`seed`	seed for random number generation.

Details

Simulate `nsimu` sample paths of standard Brownian bridge $B(x)$ on $[0, 1]$. After each draw, compute $\sup\{B_+^2(x)/(x(1-x))\}$ (if `one_sided = 1`) or $\sup\{B^2(x)/(x(1-x))\}$ (if `one_sided = 0`), where the supremum is taken over the [`x1`,`x2`] interval. Use an array to store the values. Then compute quantiles of the stored values giving critical values for the EL test at specified `alpha` levels.

Appoximate run time: 5 minutes on a 3.40 GHz desktop computer based on `nsimu=10000`.

Value

A column vector with `i`-th element given by the (1-`alpha[i]`)-quantile of the simulated values of $\sup\{B_+^2(x)/(x(1-x))\}$ (if `one_sided = 1`) or $\sup\{B^2(x)/(x(1-x))\}$ (if `one_sided = 0`).

Note

First load the required function `division00`, a function that makes $0/0 = 1$. This is to follow the convention of setting the EL ratio equal to 1 when both the numerator and denominator equal 0.

Examples

```
#quick check using a small number of replications and a coarse grid
critical_value_compute(nsimu=100,wgrid=100)
#         [,1]
#[1,] 7.377502
#save simulated values from the limiting distribution
critical_value_compute(file='supB.Rdata')
```

R code

```
#(1) A function that makes 0/0=1
division00=function(x,y){ #not dealing with Inf/Inf
#Arguments:
#x = numerator
#y = denominator
#x, y should be vectors of the same length
#Output:
#out = x/y; when x[i]=0 and y[i]=0, out[i]=1
  out=x/y
```

```
  out[as.logical((x==0)*(x==y))]=1
  return(out)
}
#(2) A function that computes critical values
critical_value_compute=function(x1=0.2,x2=0.98,one_sided=1,
   alpha=0.05,file="",nsimu=10000,wgrid=10^5,seed=1370){
  xaxis=seq(0,1,length=(wgrid+1)) #A fine grid on [0,1] for
     the standard Brownian bridge
  bgrid=length(x1)
  X=array(0,c(nsimu,bgrid,bgrid)) #An array that will contain
      sup(B_+^2(x)/(x(1-x))) or sup(B^2(x)/(x(1-x))) values
  set.seed(seed)
    for (i in 1:nsimu){
        #Simulate standard Brownian bridge B(x) on [0,1]
        w=cumsum(c(0,rnorm(wgrid)/sqrt(wgrid)))
        bbridge=w-xaxis*w[(wgrid+1)]
        #Compute sup(B_+^2(x)/(x(1-x))) or sup(B^2(x)/(x(1-x)
           ))
        for (j in 1:bgrid){
      for (m in 1:bgrid){
        if (one_sided) {
          #Compute sup(B_+^2(x)/(x(1-x))) for 1-sided test
              X[i,j,m]=max(division00(((bbridge^2)*(bbridge
                 >0)),xaxis*(1-xaxis))[xaxis>=x1[j] & xaxis<=
                 x2[m]])
        } else {
          #Compute sup(B^2(x)/(x(1-x))) for 2-sided test
              X[i,j,m]=max(division00((bbridge^2),xaxis*(1-
```

```
                    xaxis))[xaxis>=x1[j] & xaxis<=x2[m]])
            }
              }
            }
            rm(w)
            rm(bbridge)
        }
    if (file != ""){
        save(X, file=file) #Save the file for the p-value function
            in the other supplementary R code for EL test
            implementation
    }
    nq=length(alpha)
    #Compute the quantiles when bgrid=1
    Xcrit=matrix(0,nrow=nq,ncol=bgrid)
        for (i in 1:nq){
              Xcrit[i,]=quantile(X[,,1],probs=1-alpha[i])
        }
    return(Xcrit)
}
```

5.2.2.2 The one-sided/two-sided EL tests

Name of function

`ELtest_so`

Description

Implement the one-sided EL test. Implementation of the two-sided EL test using two applications of the function will be demonstrated in the Example.

Usage

`ELtest_so(data, g1, xi1 = 0.2, xi2 = 0.98, file)`

Arguments

`data` a matrix with 3 columns; column 1 contains the survival times, column 2 contains the censoring indicators, and column 3 contains the grouping variable.

`g1` the group with stochastically larger survival function for the one-sided test (should take one of the values from column 3 of `data`).

`x1` a pre-specified x_1 number ($0 < x_1 < 1$).

`x2` a pre-specified x_2 number ($x_1 < x_2 < 1$).

`file` a character string naming a file that stores the simulated values from the limiting distribution. Cannot be `''`. Use the same non-empty `file` as in `critical_value_compute`.

Details

See Chapter 2.

Value

A list with components named `test` and `p_value` containing the test statistic and the associated p-value.

Note

Need to have the R package `survival` installed first.

The function should be used after having run `critical_value_compute` with `one_sided = 1` and a nonempty `file` argument.

Several intermediate functions have to be loaded first for computing quantities associated with the test statistic, including `division00` (see Section 5.2.2.1), `product`, `a_1`, `lambda0`, `sigma2_hat`, `xi_to_tau`, `teststat`, and `pvalue`.

Examples

```
###Create an example dataset
data=matrix(c(1, 4, 4, 4, 10, 18, 28, 28, 1, 18, 19, 63, 63, 0, 0, 1,
1, 0, 1, 1, 0, 1, 1, 0, 1, 1, c(rep(1,8)), rep(2,5)), ncol=3)
###Before testing, compute critical values first
```

```
#one-sided
critical_value_compute(file='supB.Rdata')
#two-sided
critical_value_compute(one_sided=0,file='supB2.Rdata')
###One-sided EL test
ELtest_so(data,1,0.2,0.98,file='supB.Rdata')
###Two-sided EL test:
#Do the one-sided test first
test1=ELtest_so(data,1,0.2,0.98,file='supB.Rdata')$test
#Switch the two groups and do the one-sided test again
test2=ELtest_so(data,2,0.2,0.98,file='supB.Rdata')$test
#Two-sided EL statistic is the maximum of the 2 one-sided statistics
(test_2sided=max(test1,test2))
pvalue(test_2sided,file='supB2.Rdata') #the p-value
```

R code

```
library(survival)
#(1) a function that makes 0/0=1
division00=function(x,y){ #not dealing with Inf/Inf
#Arguments:
#x = numerator
#y = denominator
#x, y should be vectors of the same length
#Output:
#out = x/y; when x[i]=0 and y[i]=0, out[i]=1
  out=x/y
  out[as.logical((x==0)*(x==y))]=1
  return(out)
}
```

```
#(2) a function that makes 0*Inf=0
product=function(x,y){
#Arguments:
#x, y are vectors of the same length
#Output:
#out = x*y; when x[i]=0 or y[i]=0, out[i]=0
  out=x*y
  out[(x==0 | y==0)]=0
  return(out)
}
#(3) a function that computes a(\lambda)-1 (to solve for \hat
   {lambda}^0 later)
#see the Appendix of Chapter 2 for more info
a_1=function(lambda,fit1 ,fit2 ,t){
#Arguments:
#lambda = \lambda in the displayed equation above (A.7) in
   the Appendix
#fit1 = survfit applied on data from the 1st treatment group
#(e.g. survfit(Surv(data[,1],data[,2])~1), where data[i,1] is
    min(X_{i1},C_{i1}) and data[i,2]=I(X_{i1}<=C_{i1})
#fit2 = survfit applied on data from the 2nd treatment group
#t = the given t>=0 in local EL statistic -2logR(t)
#Output:
#out = a number a(\lambda)-1
  if (sum(fit1$n.risk==fit1$n.event)==0){
    h1=division00(fit1$n.event,(fit1$n.risk+lambda))
  } else {
    h1=division00(fit1$n.event,(fit1$n.risk+lambda))*as.
```

```
      numeric(fit1$n.risk!=fit1$n.event)+division00(fit1$n.
      event,(fit1$n.event+lambda))*as.numeric(fit1$n.risk==
      fit1$n.event)
  }
  if (sum(fit2$n.risk==fit2$n.event)==0){
    h2=division00(fit2$n.event,(fit2$n.risk-lambda))
  } else {
    h2=division00(fit2$n.event,(fit2$n.risk-lambda))*as.
      numeric(fit2$n.risk!=fit2$n.event)+division00(fit2$n.
      event,(fit2$n.event-lambda))*as.numeric(fit2$n.risk==
      fit2$n.event)
  }
  num=(1-h1)[fit1$time<=t]
  denom=(1-h2)[fit2$time<=t]
  return(division00(prod(num),prod(denom))-1)
}
#(4) a function that solves for \hat{lambda}^0
lambda0=function(t,fit1,fit2){ #t can be a vector
#Arguments:
#t = the given t>=0 in localized EL statistic -2logR(t); can
   be a vector,
#for computing a vector of \hat{lambda}^0's for different t's
#fit1 = survfit applied on data from the 1st treatment group
#(e.g. survfit(Surv(data[,1],data[,2])~1), where data[i,1] is
    min(X_{i1},C_{i1}) and data[i,2]=I(X_{i1}<=C_{i1})
#fit2 = survfit applied on data from the 2nd treatment group
#Output:
#out = \hat{lambda}^0
```

```
  out=1:length(t)*0
  for (i in 1:length(t)){
    if (sum(fit1$time[fit1$n.event!=0]<=t[i])==0 | sum(fit2$
       time[fit2$n.event!=0]<=t[i])==0) {out[i]=NA} #doesn't
       matter since -2logR is set to 0 for these cases later
    else {
  #D1 and -D2 are lower and upper bounds for \hat{lambda}^0;
     see Appendix in Chapter 2
      D1=max((fit1$n.event-fit1$n.risk)[fit1$time<=t[i] &
         fit1$n.event!=0])
      D2=max((fit2$n.event-fit2$n.risk)[fit2$time<=t[i] &
         fit2$n.event!=0])
  if (D1!=(-D2)){
      out[i]=uniroot(a_1,interval=c(D1,-D2),tol=0.0001,fit1=
         fit1,fit2=fit2,t=t[i])$root
  } else {
  out[i]=D1
  }
    }
  }
  return(out)
}
#(5) a function that computes \hat{sigma}^2(t)
sigma2_hat=function(t,fit,fit1,fit2){ #t can be a vector!
#Arguments:
#t = the given t>=0 in localized EL statistic; can be a
   vector,
#for computing a vector of \hat{sigma}^2(t)'s for different t
```

```
    's
#fit = survfit applied on data from both treatment groups
#(e.g. survfit(Surv(data[,1],data[,2])~1), where data[i,1] is
    min(X_{ij},C_{ij}) and data[i,2]=I(X_{ij}<=C_{ij})
#fit1 = survfit applied on data from the 1st treatment group
#fit2 = survfit applied on data from the 2nd treatment group
#Output:
#out = \hat{sigma}^2(t)
  n=fit$n
  out=1:length(t)*0
  for (i in 1:length(t)){
    #d1 = d_{i1} in Chapter 2, r1 = r_{i1} and so on
    d1=fit1$n.event[fit1$time<=t[i] & fit1$n.event!=0]
    r1=fit1$n.risk[fit1$time<=t[i] & fit1$n.event!=0]
    d2=fit2$n.event[fit2$time<=t[i] & fit2$n.event!=0]
    r2=fit2$n.risk[fit2$time<=t[i] & fit2$n.event!=0]
    out[i]=n*sum(division00(d1,r1*(r1-d1)))+n*sum(division00(
        d2,r2*(r2-d2)))
  }
  return(out)
}
#(6) a function that computes b(t) minus a number xi=x_1 or x
    _2, to solve for t_1 and t_2 given x_1 and x_2 later using
    uniroot
xi_to_tau=function(tau,fit,fit1,fit2,xi){
#Arguments:
#tau = argument of the function b()
#fit = survfit applied on data from both treatment groups
```

```
#(e.g. survfit(Surv(data[,1],data[,2])~1), where data[i,1] is
    min(X_{ij},C_{ij}) and data[i,2]=I(X_{ij}<=C_{ij})
#fit1 = survfit applied on data from the 1st treatment group
#fit2 = survfit applied on data from the 2nd treatment group
#xi = pre-specified x_1 or x_2
#Output:
#b(t)-xi
  #when sigma2_hat= Inf, we want xi_to_tau=sigma2_hat/1+
     sigma2_hat=1
  #so we use division00 to prevent R from giving NA's
  return(division00(1/(1+sigma2_hat(tau,fit,fit1,fit2)),1/
     sigma2_hat(tau,fit,fit1,fit2))-xi)
}
#(7) a function that computes K_n^*
teststat=function(tau1,tau2,fit,fit1,fit2){
#Arguments:
#tau1 = \hat{t}_1 in Chapter 2
#tau2 = \hat{t}_2 in Chapter 2
#fit = survfit applied on data from both treatment groups
#(e.g. survfit(Surv(data[,1],data[,2])~1), where data[i,1] is
    min(X_{ij},C_{ij}) and data[i,2]=I(X_{ij}<=C_{ij})
#fit1 = survfit applied on data from the 1st treatment group
#fit2 = survfit applied on data from the 2nd treatment group
#Output:
#testvec = the vector of -2 \log \mathcal{R} (t)'s for t \in
   [\hat{t}_1,\hat{t}_2]
#test = K_n^*
  T_11=min(fit1$time[fit1$n.event!=0])
```

```
T_21=min(fit2$time[fit2$n.event!=0])
T_1m=max(fit1$time[fit1$n.event!=0])
T_2m=max(fit2$time[fit2$n.event!=0])
Td_sort=fit$time[fit$n.event!=0 & fit$time<=tau2]
teststat_pre=1:length(Td_sort)*0
for (j in 1:length(Td_sort)){
  t=Td_sort[j]
  if (t>=max(T_11,T_21)) {
    lambda0_hat=lambda0(t,fit1 ,fit2)
    if (lambda0_hat<0) { #else teststat_pre is 0
      d1=fit1$n.event[fit1$time<=t & fit1$n.event!=0]
      r1=fit1$n.risk[fit1$time<=t & fit1$n.event!=0]
      d2=fit2$n.event[fit2$time<=t & fit2$n.event!=0]
      r2=fit2$n.risk[fit2$time<=t & fit2$n.event!=0]
      A1=r1-d1
      A2=r2-d2
      B1=d1/(r1+lambda0_hat)*as.numeric(r1!=d1)+d1/(d1+
          lambda0_hat)*as.numeric(r1==d1)
      B2=d2/(r2-lambda0_hat)*as.numeric(r2!=d2)+d2/(d2-
          lambda0_hat)*as.numeric(r2==d2)
      teststat_pre[j]=-2*sum(d1*log(B1))-2*sum(A1*log(1-B1)
          )-2*sum(d2*log(B2))-2*sum(A2*log(1-B2))+2*sum(d1*
          log(d1/r1))+2*sum(d2*log(d2/r2))+2*sum(product(A1,
          log(1-d1/r1)))+2*sum(product(A2,log(1-d2/r2)))
    }
  }
}
test_pre=teststat_pre[Td_sort>=tau1]
```

```
  return(list(testvec=test_pre,test=max(test_pre)))
}
#(8) a function that computes p-value of the test
pvalue=function(test,file){
#Arguments:
#test = K_n^* value, computed from the previous teststat
   function
#file = a character string naming a file that stores the
   simulated
#values from the limiting distribution. Cannot be empty. Use
   the
#same non-empty file as in the function critical_value_
   compute.
#Output:
#the p-value
  load(file)
  return(mean(X[,1,1]>test))
}
#(9) a function that implements the test
ELtest_so=function(data,g1,xi1,xi2,file){
#Arguments:
#data = a matrix with 3 columns; column 1 is the survival
   times, column 2 is the censoring indicators,
#and column 3 is the grouping variable
#g1 = the group with the longer survival (should take a value
    from column 3 of the data matrix)
#xi1 = pre-specified x_1
#xi2 = pre-specified x_2
```

```
#file = a character string naming a file that stores the
    simulated
#values from the limiting distribution. Cannot be empty. Use
    the
#same non-empty file as in the function critical_value_
    compute.
#Output:
#test = K_n^*
#p_value = p-value based on K_n^*
  dat=Surv(data[,1],data[,2])
  fit=survfit(dat~1)
  fit1=survfit(dat[data[,3]==g1]~1)
  fit2=survfit(dat[data[,3]!=g1]~1)
#Computing tau1 = \hat{t}_1 and tau2 = \hat{t}_2
    tau1=uniroot(xi_to_tau,interval=c(0,max(fit$time)+1),tol
        =0.0001,fit=fit,fit1=fit1,fit2=fit2,xi=xi1)$root
    if (xi_to_tau((max(fit$time)+1),fit,fit1,fit2,0)<xi2) {
      tau2=max(fit$time)
    } else {
      tau2=uniroot(xi_to_tau,interval=c(0,max(fit$time)+1),
          tol=0.0001,fit=fit,fit1=fit1,fit2=fit2,xi=xi2)$root
    }
#Computing K_n^*
  ELtest=teststat(tau1,tau2,fit,fit1,fit2)
  load(file)
  return(list(test=ELtest$test,p_value=pvalue(ELtest$test,
      file)))
}
```

5.3 Supplementary Materials for "Empirical likelihood based tests for stochastic ordering in biased sampling models"

In Section 5.3.1 we provide an alternative derivation of the local EL statistic using the Karush–Kuhn–Tucker (KKT) method (Boyd and Vandenberghe, 2004), a generalization of the Lagrange method that allows inequality constraints. Section 5.3.2 derives asymptotic orders of $\left|\hat{F}_0(t) - F_0(t)\right|$, $\left|\hat{\lambda}\right|$ and $\left|\hat{W}_j - W_j\right|$. Section 5.3.3 proves consistency of the empirical bootstrap mentioned in Section 3.5.

5.3.1 Alternative derivation of the local EL statistic

We derive the local EL ratio (3.6) for the two-sample case. The one-sample case is similar and the proof is omitted.

First, we obtain a closed-form expression for the denominator of (3.5). After a log transformation, the optimization problem becomes minimizing

$$-\sum_{j=1}^{2}\sum_{i=1}^{n_j}\log p_{ij} + \sum_{j=1}^{2} n_j \log\left(\sum_{i=1}^{n_j} w_{ij}p_{ij}\right)$$

over $(p_{11}, \ldots, p_{n_1 1},\, p_{12},\, \ldots, p_{n_2 2}) \in [0,1]^n$ subject to the constraints

$$\sum_{i=1}^{n_1} p_{i1} I_{X_{i1}\le t} - \sum_{i=1}^{n_2} p_{i2} I_{X_{i2}\le t} \le 0$$

and $\sum_{i=1}^{n_j} p_{ij} = 1$ for $j = 1, 2$. Treating $\zeta_{ij} = \log p_{ij}$ as the optimization variable, we can show that the domain is convex, the constraint and objective functions are convex (by a similar proof as in Davidov and Iliopoulos, 2009, p.179) and differentiable, and Slater's condition is satisfied. Therefore, a unique optimal value exists and the KKT conditions are necessary and sufficient for optimality. As ζ_{ij} and p_{ij} are one-to-one, the stationarity condition yields the same results, and thus we can proceed using p_{ij} instead of ζ_{ij} as the optimization variable. Details of the KKT analysis are as follows.

The Lagrangian is defined as a function $\mathcal{L} : [0,1]^n \times \mathbb{R}^3 \to \mathbb{R}$ such that

$$\begin{aligned} &\mathcal{L}(p_{11}, \dots, p_{n_1 1}, p_{12}, \dots, p_{n_2 2}, a_1, a_2, \lambda) \\ \equiv &- \sum_{j=1}^{2} \sum_{i=1}^{n_j} \log p_{ij} + \sum_{j=1}^{2} n_j \log \left(\sum_{i=1}^{n_j} w_{ij} p_{ij} \right) \\ &+ n \sum_{j=1}^{2} a_j \left(\sum_{i=1}^{n_j} p_{ij} - 1 \right) + n\lambda \left\{ \sum_{i=1}^{n_1} p_{i1} I_{X_{i1} \le t} - \sum_{i=1}^{n_2} p_{i2} I_{X_{i2} \le t} \right\}. \end{aligned}$$

It is possible that we have multiplie feasible solutions that give the same (unique) optimal value, so we adapt the convention of choosing the solution with the largest λ (and if there are multiple solutions with the same λ, choose the one with the largest a_1 and a_2) in the sequel. The optimal solution is denoted as $(\hat{p}^1_{11}, \dots, \hat{p}^1_{n_1 1}, \hat{p}^1_{12}, \dots, \hat{p}^1_{n_2 2}, \hat{a}^1_1, \hat{a}^1_2, \hat{\lambda}^1)$, with the superscript indicating the correspondence of the denominator with H_1. The dependence of the solution on t is omitted for simplicity. The optimal solution must satisfy the KKT conditions:

$$\nabla_{\boldsymbol{p}} \mathcal{L}(\hat{p}^1_{11}, \dots, \hat{p}^1_{n_1 1}, \hat{p}^1_{12}, \dots, \hat{p}^1_{n_2 2}, \hat{a}^1_1, \hat{a}^1_2, \hat{\lambda}^1) = 0,$$

$$\sum_{i=1}^{n_j} \hat{p}^1_{ij} = 1 \text{ for } j = 1, 2, \tag{5.2}$$

$$\sum_{i=1}^{n_1} \hat{p}^1_{i1} I_{X_{i1} \le t} \le \sum_{i=1}^{n_2} \hat{p}^1_{i2} I_{X_{i2} \le t}, \tag{5.3}$$

$$\hat{\lambda}^1 \ge 0, \tag{5.4}$$

$$\hat{\lambda}^1 \left\{ \sum_{i=1}^{n_1} \hat{p}^1_{i1} I_{X_{i1} \le t} - \sum_{i=1}^{n_2} \hat{p}^1_{i2} I_{X_{i2} \le t} \right\} = 0, \tag{5.5}$$

which are known as stationarity, primal feasibility (5.2 and 5.3), dual feasibility, and complementary slackness, respectively. The stationarity condition yields

$$-\frac{1}{\hat{p}^1_{ij}} + \frac{n_j w_{ij}}{\sum_{i=1}^{n_j} w_{ij} \hat{p}^1_{ij}} + n\hat{a}^1_j + n\hat{\lambda}^1 \Delta_j I_{X_{ij} \le t} = 0, \tag{5.6}$$

where $\Delta_j = 1$ for $j = 1$ and -1 for $j = 2$

The numerator of $\mathcal{R}(t)$ can be handled in a similar fashion. Denoting the optimal solution to the Lagrangian by $(\hat{p}^0_{11}, \dots, \hat{p}^0_{n_1 1}, \hat{p}^0_{12}, \dots, \hat{p}^0_{n_2 2}, \hat{a}^0_1, \hat{a}^0_2, \hat{\lambda}^0)$, the solution satisfies (5.6), (5.2) (with superscript 1 replaced by 0) and

$$\sum_{i=1}^{n_1} \hat{p}_{i1}^0 I_{X_{i1} \le t} - \sum_{i=1}^{n_2} \hat{p}_{i2}^0 I_{X_{i2} \le t} = 0. \tag{5.7}$$

We then have

$$\mathcal{R}(t) = \prod_{j=1}^{2} \prod_{i=1}^{n_j} \frac{\hat{p}_{ij}^0 / \sum_{i=1}^{n_j} w_{ij} \hat{p}_{ij}^0}{\hat{p}_{ij}^1 / \sum_{i=1}^{n_j} w_{ij} \hat{p}_{ij}^1}. \tag{5.8}$$

We next further simplify $\mathcal{R}(t)$ by showing that it depends only on a single Lagrange multiplier, $\hat{\lambda}^0$. The key is to study the condition (5.5), which implies that $\hat{\lambda}^1$ either satisfies $\sum_{i=1}^{n_1} \hat{p}_{i1}^1 I_{X_{i1} \le t} - \sum_{i=1}^{n_2} \hat{p}_{i2}^1 I_{X_{i2} \le t} = 0$ (i.e., $\hat{\lambda}^1 = \hat{\lambda}^0$) or $\hat{\lambda}^1 = 0$. The case when both hold is equivalent to $\hat{\lambda}^0 = 0$ and results in $\mathcal{R}(t) = 1$. On the other hand, when only one of $\hat{\lambda}^1 = 0$ and $\sum_{i=1}^{n_1} \hat{p}_{i1}^1 I_{X_{i1} \le t} - \sum_{i=1}^{n_2} \hat{p}_{i2}^1 I_{X_{i2} \le t} = 0$ is true, we have the following two cases:

Case 1: If $\hat{\lambda}^0 < 0$, then by (5.4) we have $\hat{\lambda}^1 \neq \hat{\lambda}^0$. Since $\hat{\lambda}^1$ is either 0 or $\hat{\lambda}^0$, we obtain that $\hat{\lambda}^1 = 0$.

Case 2: If $\hat{\lambda}^0 > 0$, then $\hat{\lambda}^0$ satisfies (5.2)–(5.6). Suppose $\hat{\lambda}^1 = 0$, then this contradicts with our convention of choosing the largest $\hat{\lambda}^1$ satisfying (5.2)–(5.6). Thus, $\hat{\lambda}^1$ cannot be 0 but must be $\hat{\lambda}^0$.

Then from (5.8) we have

$$\mathcal{R}(t) = \begin{cases} 1, & \hat{\lambda}^0 > 0, \\ \prod_{j=1}^{2} \prod_{i=1}^{n_j} \frac{\hat{p}_{ij}^0 / \sum_{i=1}^{n_j} w_{ij} \hat{p}_{ij}^0}{\bar{p}_{ij} / \sum_{i=1}^{n_j} w_{ij} \bar{p}_{ij}}, & \hat{\lambda}^0 \le 0, \end{cases} \tag{5.9}$$

where $\bar{p}_{ij} = \frac{1/w_{ij}}{\sum_{i=1}^{n_j} 1/w_{ij}}$ is the optimal solution corresponding to $\hat{\lambda}^1 = 0$. We also obtain the equivalence between $\hat{\lambda}^0 > 0$ and $\tilde{F}_1(t) > \tilde{F}_2(t)$ (i.e. $\hat{\lambda}^1 > 0$), which is used in (3.6).

To obtain an explicit expression of $\hat{p}_{ij}^0$, we consider an alternative formulation of the optimization in the numerator as follows. Let $F_0(t) = \sum_{i=1}^{n_j} p_{ij} I_{X_{ij} \le t}$ and

$W_j = \sum_{i=1}^{n_j} p_{ij} w_{ij}$ be fixed first. This adds the additional constraints

$$\sum_{i=1}^{n_j} p_{ij} \left(I_{X_{ij} \leq t} - F_0(t) \right) = 0 \text{ and} \tag{5.10}$$

$$\sum_{i=1}^{n_j} p_{ij} \left(w_{ij} - W_j \right) = 0 \tag{5.11}$$

for $j = 1, 2$, where (5.10) replaces the previous condition (5.7). The Lagrangian can be defined as a function $\mathcal{L} : [0,1]^n \times \mathbb{R}^3 \times [0,\infty)^2 \times [0,1] \to \mathbb{R}$ such that

$$\begin{aligned} &\mathcal{L}(p_{11}, \ldots, p_{n_1 1}, p_{12}, \ldots, p_{n_2 2}, a_1, a_2, \lambda, W_1, W_2, F_0(t)) \\ \equiv & -\sum_{j=1}^{2} \sum_{i=1}^{n_j} \log p_{ij} + n \sum_{j=1}^{2} a_j \left(\sum_{i=1}^{n_j} p_{ij} - 1 \right) \\ & + n \sum_{j=1}^{2} \lambda_{1j} \left\{ \sum_{i=1}^{n_j} \left(I_{X_{ij} \leq t} - F_0(t) \right) p_{ij} \right\} + n \sum_{j=1}^{2} \lambda_{2j} \left\{ \sum_{i=1}^{n_j} \left(w_{ij} - W_j \right) p_{ij} \right\}. \end{aligned}$$

The stationarity condition yields

$$\hat{p}_{ij}^{0alt} = \frac{1}{n} \frac{1}{\hat{a}_j^{0alt} + \hat{\lambda}_{1j}^{0alt} \left(I_{X_{ij} \leq t} - \hat{F}_0(t) \right) + \hat{\lambda}_{2j}^{0alt} \left(w_{ij} - W_j \right)}.$$

Denote the resulting profile log-likelihood as

$$l(W_1, W_2, F_0(t)) = \sum_{j=1}^{2} \sum_{i=1}^{n_j} \log(\hat{p}_{ij}^{0alt}) - \sum_{j=1}^{2} n_j \log(W_j) + \sum_{j=1}^{2} \sum_{i=1}^{n_j} \log(w_{ij}).$$

We then optimize l over $(W_1, W_2, F_0(t))$. This leads to $\hat{\lambda}_{2j}^{0alt} = \kappa_j / \hat{W}_j^{0alt}$ and $\hat{\lambda}_{11}^{0alt} = -\hat{\lambda}_{12}^{0alt}$, based on which we can re-write $\hat{p}_{ij}^{0alt}$ as

$$\frac{1}{n} \frac{1}{\hat{a}_j^{0alt} - \kappa_j + \hat{\lambda}_{11}^{0alt} \Delta_j \left(I_{X_{ij} \leq t} - \hat{F}_0(t) \right) + \frac{\kappa_j w_{ij}}{\hat{W}_j^{0alt}}}.$$

Note that by setting $\hat{\lambda}_{11}^{0alt} = \hat{\lambda}^0$ and $\hat{a}_j^{0alt} - \kappa_j - \hat{\lambda}^0 \Delta_j \hat{F}_0(t) = \hat{a}_j^0$, we can see $\hat{p}_{ij}^{0alt}$ satisfies (5.6), (5.2) and (5.7), and thus $\hat{p}_{ij}^{0alt} = \hat{p}_{ij}^0$. Then we have

$$\hat{p}_{ij}^0 = \hat{p}_{ij}^{0alt} = \frac{1}{n} \frac{1}{\hat{a}_j^{0alt} - \kappa_j + \frac{\kappa_j w_{ij}}{\hat{W}_j^{0alt}} + \hat{\lambda}_{11}^{0alt} \Delta_j \left(I_{X_{ij} \leq t} - \hat{F}_0(t) \right)}, \tag{5.12}$$

where $(\hat{a}_1^{0alt}, \hat{a}_2^{0alt}, \hat{W}_1^{0alt}, \hat{W}_2^{0alt}, \hat{\lambda}^{0alt}, \hat{F}_0(t))$ satisfy the conditions (5.2) (with superscript 1 replaced by 0), (5.10) and (5.11). This, (5.9) and the equivalence between $\hat{\lambda}^0 > 0$ and $\tilde{F}_1(t) > \tilde{F}_2(t)$ then give (3.6). Note that in (3.6) we use the simplified notation $\hat{p}_{ij}$, $\hat{a}_j$, $\hat{W}_j$, and $\hat{\lambda}$ to replace $\hat{p}_{ij}^0$, $\hat{a}_j^{0alt}$, $\hat{W}_j^{0alt}$, and $\hat{\lambda}_{11}^{0alt}$, respectively.

5.3.2 Asymptotic orders of $\left|\hat{F}_0(t) - F_0(t)\right|$, $\left|\hat{\lambda}\right|$ and $\left|\hat{W}_j - W_j\right|$

We establish the asymptotic orders of $\left|\hat{F}_0(t) - F_0(t)\right|$, $\left|\hat{\lambda}\right|$ and $\left|\hat{W}_j - W_j\right|$ (for $j = 1, 2$). The orders we prove hold uniformly.

First establish the asymptotic orders of $\left|\hat{\lambda}\right| \hat{W}_j$ and $\left|\hat{F}_0(t) - F_0(t)\right|$. Let $\hat{\lambda} = \theta \left|\hat{\lambda}\right|$ such that $|\theta| = 1$ and let $\theta_j = \Delta_j \theta$. Denote $\Delta_j \hat{\lambda} g_{1ij}(\hat{F}_0(t), \hat{W}_j)$ by V_{ij}. Substituting $1/(1 + V_{ij}) = 1 - V_{ij}/(1 + V_{ij})$ into $\theta_j Q_{1j}(\hat{F}_0(t), \hat{W}_j, \hat{\lambda}) = 0$, we get

$$\begin{aligned} 0 &= \frac{1}{n}\theta_j \sum_{i=1}^{n_j} g_{1ij}(\hat{F}_0(t), \hat{W}_j)\left(1 - \frac{V_{ij}}{1 + V_{ij}}\right) \\ &= \theta_j \frac{1}{n}\sum_{i=1}^{n_j} g_{1ij}(\hat{F}_0(t), \hat{W}_j) - \frac{\left|\hat{\lambda}\right|}{n}\sum_{i=1}^{n_j} \frac{\theta_j g_{1ij}(\hat{F}_0(t), \hat{W}_j) g_{1ij}(\hat{F}_0(t), \hat{W}_j)\theta_j}{1 + V_{ij}} \\ &= \theta_j \frac{1}{n}\sum_{i=1}^{n_j} g_{1ij}(\hat{F}_0(t), \hat{W}_j) - \left|\hat{\lambda}\right| \frac{1}{n}\sum_{i=1}^{n_j} \frac{g_{1ij}^2(\hat{F}_0(t), \hat{W}_j)}{1 + V_{ij}}. \end{aligned} \tag{5.13}$$

Note that $1 + V_{ij} > 0$ by $\hat{p}_{ij} > 0$ for all i, j. Thus we can obtain the following (in)equalities:

$$\begin{aligned} \left|\hat{\lambda}\right| \frac{1}{n}\sum_{i=1}^{n_j} g_{1ij}^2(\hat{F}_0(t), \hat{W}_j) &\le \left|\hat{\lambda}\right| \frac{1}{n}\sum_{i=1}^{n_j} \frac{g_{1ij}^2(\hat{F}_0(t), \hat{W}_j)}{1 + V_{ij}}\left(1 + \max_{1\le i\le n_j} V_{ij}\right) \\ &\le \left|\hat{\lambda}\right| \frac{1}{n}\sum_{i=1}^{n_j} \frac{g_{1ij}^2(\hat{F}_0(t), \hat{W}_j)}{1 + V_{ij}}\left(1 + \left|\hat{\lambda}\right| \max_{1\le i\le n_j} \left|g_{1ij}(\hat{F}_0(t), \hat{W}_j)\right|\right) \\ &= \theta_j \frac{1}{n}\sum_{i=1}^{n_j} g_{1ij}(\hat{F}_0(t), \hat{W}_j)\left(1 + \left|\hat{\lambda}\right| \max_{1\le i\le n_j} \left|g_{1ij}(\hat{F}_0(t), \hat{W}_j)\right|\right), \end{aligned}$$

where the last equality follows from (5.13). This implies

$$\left|\hat{\lambda}\right| \hat{W}_j S_j(\hat{F}_0(t)) \le \theta_j \bar{g}_{1j}(\hat{F}_0(t))\left(1 + \left|\hat{\lambda}\right| \hat{W}_j Z_j(\hat{F}_0(t))\right), \tag{5.14}$$

where $S_j(\hat{F}_0(t)) = 1/n \times \sum_{i=1}^{n_j} [(I_{X_{ij} \leq t} - \hat{F}_0(t))/(\kappa_j w_{ij})]^2$, $\bar{g}_{1j}(\hat{F}_0(t)) = 1/n \times \sum_{i=1}^{n_j} [(I_{X_{ij} \leq t} - \hat{F}_0(t))/(\kappa_j w_{ij})]$, and $Z_j(\hat{F}_0(t)) = \max_{1 \leq i \leq n_j} \left| (I_{X_{ij} \leq t} - \hat{F}_0(t))/(\kappa_j w_{ij}) \right|$. Assuming $E_{F_j}(1/w_j^2(X_{ij})) < \infty$, we can show the uniform convergence of $S_j(\hat{F}_0(t))$ and $\bar{g}_{1j}(\hat{F}_0(t))$ by the Glivenko-Cantelli's Theorem and the Donsker's Theorem (see, e.g., van der Vaart, 2000, Ch. 19), based on which we have

$$S_j(\hat{F}_0(t)) = \frac{1}{\kappa_j W_j} E_{F_j} \frac{\left(I_{X_{ij} \leq t} - \hat{F}_0(t) \right)^2}{w_{ij}} + o(1) \tag{5.15}$$

and

$$\bar{g}_{1j}(\hat{F}_0(t)) = \left(F_0(t) - \hat{F}_0(t) \right) \frac{1}{n_j} \sum_{i=1}^{n_j} \frac{1}{w_{ij}} + O_p(n^{-\frac{1}{2}}). \tag{5.16}$$

As for $Z_j(\hat{F}_0(t))$, we can bound it by

$$\frac{2}{\kappa_j} \max_{1 \leq i \leq n_j} \left| \frac{1}{w_{ij}} \right| = O(1) o(n^{\frac{1}{3}}) = o(n^{\frac{1}{3}}), \tag{5.17}$$

where the first $o(n^{1/3})$ order is obtained via a similar proof as in Lemma 11.2 of Owen (2001). From these uniform convergence results and (5.14), we have

$$\left| \hat{\lambda} \right| W_1 \frac{S_1(\hat{F}_0(t))}{1 + \left| \hat{\lambda} \right| W_1 Z_1(\hat{F}_0(t))} \leq \theta_1 \left(F_0(t) - \hat{F}_0(t) \right) \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{1}{w_{i1}} + O_p(n^{-\frac{1}{2}}), \tag{5.18}$$

$$\left| \hat{\lambda} \right| W_2 \frac{S_2(\hat{F}_0(t))}{1 + \left| \hat{\lambda} \right| W_2 Z_2(\hat{F}_0(t))} \leq -\theta_1 \left(F_0(t) - \hat{F}_0(t) \right) \frac{1}{n_2} \sum_{i=1}^{n_2} \frac{1}{w_{i2}} + O_p(n^{-\frac{1}{2}}). \tag{5.19}$$

Multiplying (5.18) by $n_1 / \sum_{i=1}^{n_1} (1/w_{i1})$ and (5.19) by $n_2 / \sum_{i=1}^{n_2} (1/w_{i2})$, adding up the two terms, and using the fact that $\sum_{i=1}^{n_j} (1/w_{ij})/n_j$ is $O(1)$ (SLLN), we can bound

$$\sum_{j=1}^{2} \frac{S_j}{Z_j + \frac{1}{\left| \hat{\lambda} \right| \hat{W}_j}} \frac{n_j}{\sum_{i=1}^{n_j} \frac{1}{w_{ij}}}$$

above by an $O_p(n^{-1/2})$ term. This and (5.15) imply $Z_j + 1/(\left| \hat{\lambda} \right| \hat{W}_j)$ must grow faster than $n^{1/2}$ (in probability). Then by (5.17) we obtain

$$\left| \hat{\lambda} \right| \hat{W}_j = O_p(n^{-\frac{1}{2}}) \tag{5.20}$$

for $j = 1, 2$. This, together with (5.15) and (5.17), imply that the l.h.s. of (5.18) and (5.19) are both $O_p(n^{-1/2})$. Then (5.18) and (5.19) imply that both $\theta_1(\hat{F}_0(t) - F_0(t))$ and $\theta_1(F_0(t) - \hat{F}_0(t))$ are bounded above by $O_p(n^{-1/2})$ terms. And thus

$$\left|\hat{F}_0(t) - F_0(t)\right| = O_p(n^{-\frac{1}{2}}).$$

Next we establish the order of $\left|\hat{W}_j - W_j\right|$ and $\left|\hat{\lambda}\right|$. Let $\mathbf{g}_{ij}(\hat{F}_0(t), W_j) = [(w_{ij} - W_j)/(\kappa_j w_{ij}), (I_{X_{ij} \le t} - \hat{F}_0(t))/(\kappa_j w_{ij})]^T$ and let $\hat{\boldsymbol{\lambda}} = [0, \hat{\lambda}]^T = \boldsymbol{\theta}\|\hat{\boldsymbol{\lambda}}\|$ such that $\|\boldsymbol{\theta}\| = 1$ and let $\boldsymbol{\theta}_j = \Delta_j \boldsymbol{\theta}$. Then (3.11) and $\sum_{i=1}^{n_j} \hat{p}_{ij} = 1$ imply

$$\begin{bmatrix} \hat{W}_j - W_j \\ 0 \end{bmatrix} = \frac{1}{n} \sum_{i=1}^{n_j} \frac{\hat{W}_j \mathbf{g}_{ij}(\hat{F}_0(t), W_j)}{1 + \Delta_j \hat{W}_j \hat{\boldsymbol{\lambda}}^T \mathbf{g}_{ij}(\hat{F}_0(t), W_j)}. \tag{5.21}$$

By (5.20) and a similar reasoning as in (5.17), we have that $\max_{1 \le i \le n_j} \|\Delta_j \hat{W}_j \hat{\boldsymbol{\lambda}}^T \mathbf{g}_{ij}(\hat{F}_0(t), W_j)\| \le \|\Delta_j \hat{W}_j \hat{\boldsymbol{\lambda}}^T\| \max_{1 \le i \le n_j} \|\mathbf{g}_{ij}(\hat{F}_0(t), W_j)\| = O_p(n^{-1/2}) o(n^{1/3}) = o_p(1)$. Then we can expand the r.h.s. of (5.21) as

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n_j} \left\{ \hat{W}_j \mathbf{g}_{ij}(\hat{F}_0(t), W_j) \left[1 - \Delta_j \hat{W}_j \hat{\boldsymbol{\lambda}}^T \mathbf{g}_{ij}(\hat{F}_0(t), W_j) + o_p(\|\hat{W}_j \hat{\boldsymbol{\lambda}}^T \mathbf{g}_{ij}(\hat{F}_0(t), W_j)\|) \right] \right\} \\
=&\hat{W}_j \bar{\mathbf{g}}_j(\hat{F}_0(t), W_j) - \Delta_j \hat{W}_j^2 \mathbf{S}_j(\hat{F}_0(t), W_j) \hat{\boldsymbol{\lambda}} \\
&+ \hat{W}_j \bar{\mathbf{g}}_j(\hat{F}_0(t), W_j) o_p(\max_{1 \le i \le n_j} \|\hat{W}_j \hat{\boldsymbol{\lambda}}^T \mathbf{g}_{ij}(\hat{F}_0(t), W_j)\|),
\end{aligned}$$

where $\bar{\mathbf{g}}_j(\hat{F}_0(t), W_j) = 1/n \times \sum_{i=1}^{n_j} \mathbf{g}_{ij}(\hat{F}_0(t), W_j)$ and $\mathbf{S}_j(\hat{F}_0(t), W_j) = 1/n \times \sum_{i=1}^{n_j} [\mathbf{g}_{ij}(\hat{F}_0(t), W_j) \mathbf{g}_{ij}^T(\hat{F}_0(t), W_j)]$. By a similar reasoning as in (5.15) and (5.16), we can show that $\bar{\mathbf{g}}_j(\hat{F}_0(t), W_j) = O_p(n^{-1/2})$ and $\mathbf{S}_j(\hat{F}_0(t), W_j) = O(1)$. Then $\left|\hat{W}_j - W_j\right| \le \hat{W}_j \|\bar{\mathbf{g}}_j(\hat{F}_0(t), W_j)\| + \|\mathbf{S}_j(\hat{F}_0(t), W_j)\| \left|\hat{\lambda}\right| \hat{W}_j^2 + \hat{W}_j \|\bar{\mathbf{g}}_j(\hat{F}_0(t), W_j)\| o_p(1)$. This and (5.20) imply $\left|1 - W_j/\hat{W}_j\right| = O_p(n^{-1/2})$, which implies

$$\left|\hat{W}_j - W_j\right| = O_p(n^{-\frac{1}{2}}).$$

Then by (5.20) again, we obtain

$$\left|\hat{\lambda}\right| = O_p(n^{-\frac{1}{2}}).$$

Remark. In (5.16), we can replace $O_p(n^{-\frac{1}{2}})$ by $O(n^{-\frac{1}{2}})$. This is due to the law of iterated logarithms and the fact that the envelope of the Donsker class has bounded second moment (see, e.g., Kosorok, 2008, p. 31). As a result, we can show that all the subsequent O_p and o_p terms can be replaced by O and o, respectively. This will be used when we prove bootstrap consistency in Section 3.6.3 and 5.3.3 .

5.3.3 Empirical bootstrap consistency

We follow the same steps in Appendix 5.3.3, except that now $U_n^*(t)$ is replaced with

$$U_n^{*E}(t) = \hat{\sigma}^{-\frac{1}{2}}(t,t)\times$$
$$\left[\frac{\hat{W}_2}{\sqrt{n_2}\sqrt{\kappa_2}}\sum_{i=1}^{n_2}(M_{n_2 i}-1)\frac{I_{X_{i2}\le t}-\hat{F}_0(t)}{w_{i2}} - \frac{\hat{W}_1}{\sqrt{n_1}\sqrt{\kappa_1}}\sum_{i=1}^{n_1}(M_{n_1 i}-1)\frac{I_{X_{i1}\le t}-\hat{F}_0(t)}{w_{i1}}\right],$$

where $(M_{n_j 1},\ldots,M_{n_j n_j}) \sim Multinomial(n_j; 1/n_j,\ldots,1/n_j)$ $(j=1,2)$ is independent of the observed data, and that $U_n^{**E}(t)$ is again defined by replacing $\hat{W}_j$ and $\hat{F}_0(t)$ in $U_n^{*E}(t)$ with their respective limits W_j and $F_0(t)$.

To establish bootstrap consistency of $U_n^{*E}(t)$, we follow the proof of consistency for $U_n^*(t)$ in Appendix 5.3.3, but use the empirical bootstrap consistency theorem directly (see, e.g., van der Vaart and Wellner, 1996, Theorem 3.6.2) instead of the multiplier central limit theorem.

Next we establish the asymptotic equivalence of $U_n^{*E}(t)$ and $U_n^{**E}(t)$, conditionally on the data almost surely. We can prove the result by showing that

$$E_M\left(\sup_{t\in[t_1,t_2]}\left|U_n^{*E}(t)-U_n^{**E}(t)\right|^2 \middle| X_{11},X_{21},\ldots,X_{12},X_{22},\ldots\right)$$
$$\le \sum_{j=1}^{2}\frac{1}{n_j\kappa_j}\sum_{i=1}^{n_j}\frac{E_M\left(M_{n_j i}-1\right)^2}{w_{ij}^2}\sup_{t\in[t_1,t_2]}\left|\frac{\hat{W}_j\left(I_{X_{ij}\le t}-\hat{F}_0(t)\right)}{\sqrt{\hat{\sigma}(t,t)}}-\frac{W_j\left(I_{X_{ij}\le t}-F_0(t)\right)}{\sqrt{\sigma(t,t)}}\right|$$
$$=\sum_{j=1}^{2}\frac{1}{n_j\kappa_j}\sum_{i=1}^{n_j}\frac{1-1/n_j}{w_{ij}^2}O(n^{-1})\xrightarrow{a.s.}0.$$

Table 5.1: Power at $\alpha = 0.05$ for unequal censoring based on 10,000 replications. **Model A**: survival functions as in Figure 2.1, upper left panel. **Model B**: survival functions as in Figure 2.1, upper right panel.

model	group size	test	exp. censoring 10%	exp. censoring 25%	unif. censoring 10%	unif. censoring 25%
Model A	50	K_n^*	0.850	0.827	0.848	0.831
		log-rank	0.325	0.397	0.319	0.386
		WKM	0.334	0.414	0.340	0.464
	80	K_n^*	0.974	0.966	0.975	0.970
		log-rank	0.424	0.528	0.423	0.519
		WKM	0.432	0.532	0.443	0.613
Model B	50	K_n^*	0.684	0.657	0.684	0.669
		log-rank	0.642	0.690	0.627	0.677
		WKM	0.549	0.648	0.545	0.683
	80	K_n^*	0.870	0.852	0.874	0.862
		log-rank	0.798	0.836	0.795	0.836
		WKM	0.696	0.792	0.704	0.843