2005

Comparing Distribution Functions via Empirical Likelihood

Yichuan Zhao
Georgia State University, yichuan@gsu.edu

Ian W. McKeague

The International Journal of Biostatistics

Volume 1, *Issue* 1	2005	*Article* 5

Comparing Distribution Functions via Empirical Likelihood

Ian W. McKeague* Yichuan Zhao†

*Columbia University, im2131@columbia.edu
†Georgia State University, dz2007@gmail.com

Comparing Distribution Functions via Empirical Likelihood

Ian W. McKeague and Yichuan Zhao

Abstract

This paper develops empirical likelihood based simultaneous confidence bands for differences and ratios of two distribution functions from independent samples of right-censored survival data. The proposed confidence bands provide a flexible way of comparing treatments in biomedical settings, and bring empirical likelihood methods to bear on important target functions for which only Wald-type confidence bands have been available in the literature. The approach is illustrated with a real data example.

KEYWORDS: Kaplan–Meier estimator, Nelson–Aalen estimator, plug-in, right censoring

1 Introduction

The purpose of this paper is to develop simultaneous confidence bands for the difference and the ratio of two distribution functions using the empirical likelihood approach. The proposed confidence bands provide an attractive graphical comparison of treatment and control groups in biomedical studies on the basis of independent right-censored survival time data from each group.

The graphical comparison of two survival distributions can be done in various ways. Empirical likelihood (EL) techniques have been used to provide confidence bands for Q-Q plots (Einmahl and McKeague, 1999), ratios of survival functions (McKeague and Zhao, 2002), and P-P plots (Claeskens et al., 2003); references to the earlier literature may be found in these papers. The simplest and most natural way to carry out such a comparison, however, is to target the difference and the ratio of the two distribution functions, which represent directly interpretable measures of treatment effect and relative risk. The difference is suitable when an absolute measure (of treatment effect) is needed, the ratio when a relative measure is needed. Parzen et al. (1997) constructed a Wald-type simultaneous confidence band for a difference between two distribution functions, but, as far as we know, there is no EL band available in the literature.

We develop our approach in terms of differences and ratios of linear functionals of the cumulative hazard functions:

$$\alpha(t) = \int_0^t g_1(s)\, dA_1(s) - \int_0^t g_2(s)\, dA_2(s)$$

and

$$\beta(t) = \int_0^t g_1(s)\, dA_1(s) \Big/ \int_0^t g_2(s)\, dA_2(s),$$

where $A_j(t)$ is the cumulative hazard function for group j, and $g_j(t) = S_j(t) = 1 - F_j(t)$ is the survival function for group j. The difference between the two distribution functions is then seen to be $\alpha(t) = F_1(t) - F_2(t) = S_2(t) - S_1(t)$, and the ratio of the two distribution functions is $\beta(t) = F_1(t)/F_2(t)$. The difference and the ratio of the cumulative hazard functions are obtained by taking $g_j \equiv 1$, and our approach extends essentially without change to that case as well. These various ways of comparing the two distributions provide greater flexibility than what is currently available.

The unknown survival functions $g_j = S_j$ in the above representations of $\alpha(t)$ and $\beta(t)$ can be seen as nuisance parameters in the EL statistic, so we take the approach of plugging-in the corresponding Kaplan–Meier estimates. Plug-in for unknown parameters in estimating equations has been used extensively in conjunction with EL, and typically perturbs the usual chi-squared limit distribution into a more complicated form, see, e.g., Hjort et al. (2005). The present case is no exception: we find that the empirical likelihood statistic with plug-in of the Kaplan–Meier estimates is not asymptotically distribution free; a bootstrap procedure is thus needed to determine critical values for the EL confidence bands.

An EL confidence interval for a linear functional $\int_0^\infty g(s)\,dA(s)$ of a cumulative hazard function A, where g is *known*, has been developed by Pan and Zhou (2002). Their approach is based on a Poisson extension of the likelihood (cf. Murphy, 1995), but we found that it is not easy to deal with the target functions $\alpha(t)$ and $\beta(t)$ using a likelihood of this form. Instead, our approach is based on the *standard* nonparametric likelihood for (S_1, S_2), cf. McKeague and Zhao (2002). The EL function (or nonparametric likelihood ratio) is constructed by substituting $A_j(t) = -\log S_j(t)$ in the estimating equation that defines the target function of interest, and we find that this leads to a mathematically tractable formulation.

The main results underlying our derivation of the proposed confidence bands are presented in Section 2. In Section 3 we develop the bootstrap procedure needed to construct the bands in practice. In Section 4 we give an illustrative example. Some concluding remarks are given in Section 5. Proofs are contained in the Appendix.

2 Main results

2.1 Preliminaries

We consider the standard two-sample framework with independent right censoring. That is, we have two independent samples of i.i.d. observations of the form (Z_{ji}, δ_{ji}), where $j = 1, 2$ indexes the sample, $i = 1, \ldots, n_j$ indexes the observations within each sample, and $Z_{ji} = X_{ji} \wedge Y_{ji}$, $\delta_{ji} = 1_{\{X_{ji} \le Y_{ji}\}}$. The distribution functions of X_{ji} and Y_{ji} are denoted F_j and G_j, respectively. The survival functions $S_j = 1 - F_j$ are assumed to be continuous. The total sample size is $n = n_1 + n_2$. We work with independent and non-negative X_{ji} and Y_{ji}. The nonparametric likelihood is given by

$$L(\tilde{S}_1, \tilde{S}_2) = \prod_{j=1}^{2} \prod_{i=1}^{n_j} \{\tilde{S}_j(Z_{ji}-) - \tilde{S}_j(Z_{ji})\}^{\delta_{ji}} \tilde{S}_j(Z_{ji})^{1-\delta_{ji}}, \tag{2.1}$$

where $\tilde{S}_j$ belongs to Γ, the space of all survival functions on $[0, \infty)$.

The target functions $\alpha(t)$ and $\beta(t)$ may be written in the general form $\theta(t) = \theta(t, S_1, S_2, g_1, g_2)$ by substitution of the cumulative hazard functions $A_j(t) = -\log S_j(t)$. The empirical likelihood ratio for $\theta(t)$, with plug-in of estimators $\hat{g}_j$ for the g_j, is then given by

$$R(\tilde{\theta}(t), t, \hat{g}_1, \hat{g}_2) = \frac{\sup\{L(\tilde{S}_1, \tilde{S}_2) : \theta(t, \tilde{S}_1, \tilde{S}_2, \hat{g}_1, \hat{g}_2) = \tilde{\theta}(t),\ (\tilde{S}_1, \tilde{S}_2) \in \Gamma \times \Gamma\}}{\sup\{L(\tilde{S}_1, \tilde{S}_2) : (\tilde{S}_1, \tilde{S}_2) \in \Gamma \times \Gamma\}}, \tag{2.2}$$

which can be expressed more explicitly in the case of $\theta(t) = \alpha(t)$ as follows. The ordered uncensored survival times, i.e., the X_{ji} with corresponding $\delta_{ji} = 1$, are written $0 \leq T_{j1} \leq \cdots \leq T_{jN_j} < \infty$, and $r_{ji} = \sum_{k=1}^{n_j} 1_{\{Z_{jk} \geq T_{ji}\}}$ denotes the size of the risk set at $T_{ji}-$, $d_{ji} = \sum_{k=1}^{n_j} 1_{\{Z_{jk} = T_{ji}, \delta_{jk} = 1\}}$ denotes the number of "deaths" occurring at time T_{ji}. Define $K_j(t) = \#\{i : T_{ji} \leq t\}$ and $D_j = \max_{i: T_{ji} \leq t}((d_{ji} - r_{ji})/\hat{g}_j(T_{ji}) 1_{\{\hat{g}_j(T_{ji}) > 0\}})$. Using Lagrange's method [cf. Thomas and Grunkemeier (1975) or Li (1995)], it can be shown that

$$\begin{aligned} & -2 \log R(\alpha(t), t, \hat{g}_1, \hat{g}_2) \\ = & -2 \sum_{i=1}^{K_1(t)} \left((r_{1i} - d_{1i}) \log\left(1 + \frac{\lambda_n \hat{g}_1(T_{1i})}{r_{1i} - d_{1i}}\right) - r_{1i} \log\left(1 + \frac{\lambda_n \hat{g}_1(T_{1i})}{r_{1i}}\right)\right) \\ & -2 \sum_{i=1}^{K_2(t)} \left((r_{2i} - d_{2i}) \log\left(1 - \frac{\lambda_n \hat{g}_2(T_{2i})}{r_{2i} - d_{2i}}\right) - r_{2i} \log\left(1 - \frac{\lambda_n \hat{g}_2(T_{2i})}{r_{2i}}\right)\right) \end{aligned} \tag{2.3}$$

where the Lagrange multiplier $D_1 < \lambda_n < -D_2$ satisfies the equation

$$\sum_{i=1}^{K_1(t)} \log\left(1 - \frac{d_{1i}}{r_{1i} + \lambda_n \hat{g}_1(T_{1i})}\right) \hat{g}_1(T_{1i}) - \sum_{i=1}^{K_2(t)} \log\left(1 - \frac{d_{2i}}{r_{2i} - \lambda_n \hat{g}_2(T_{2i})}\right) \hat{g}_2(T_{2i}) = -\alpha(t). \tag{2.4}$$

Here we are suppressing the dependence of the Lagrange multiplier on t. The equation (2.4) has a unique solution λ_n provided $D_j < 0$, $\hat{g}_j(T_{ji}) > 0$, $i = 1, \ldots, N_j$, $j = 1, 2$, because as a function of λ_n the l.h.s. of (2.4) is continuous, strictly increasing and tends to $\pm\infty$ as $\lambda_n \uparrow -D_2$ or $\lambda_n \downarrow D_1$.

Similarly, in the case of $\theta(t) = \beta(t)$, we have

$$
\begin{aligned}
-2 \log R(\beta(t), t, \hat{g}_1, \hat{g}_2) \quad = \quad & -2 \sum_{j=1}^{2} \sum_{i=1}^{K_j(t)} \left\{ (r_{ji} - d_{ji}) \log \left(1 + \frac{(-\beta(t))^{j-1} \lambda_n \hat{g}_j(T_{ji})}{r_{ji} - d_{ji}} \right) \right. \\
& \left. - r_{ji} \log \left(1 + \frac{(-\beta(t))^{j-1} \lambda_n \hat{g}_j(T_{ji})}{r_{ji}} \right) \right\}, \qquad (2.5)
\end{aligned}
$$

where the Lagrange multiplier $D_1 < \lambda_n < -D_2/\beta(t)$ satisfies the equation

$$
\begin{aligned}
& \sum_{i=1}^{K_1(t)} \log \left(1 - \frac{d_{1i}}{r_{1i} + \lambda_n \hat{g}_1(T_{1i})} \right) \hat{g}_1(T_{1i}) \\
- \quad & \beta(t) \sum_{i=1}^{K_2(t)} \log \left(1 - \frac{d_{2i}}{r_{2i} \beta(t) \lambda_n \hat{g}_2(T_{2i})} \right) \hat{g}_2(T_{2i}) = 0. \qquad (2.6)
\end{aligned}
$$

If $\beta(t) > 0$, the equation (2.6) has a unique solution λ_n provided $\hat{g}_j(T_{ji}) > 0$, $i = 1, \ldots, K_j(t)$, $j = 1, 2$, because, as a function of λ_n, the l.h.s. of (2.6) is continuous, strictly increasing and tends to $\pm\infty$ as $\lambda_n \uparrow -D_2/\beta(t)$ or $\lambda_n \downarrow D_1$.

We assume throughout that $n_j/n \to p_j > 0$ as $n \to \infty$. The plug-in estimate of $g_j(t) = S_j(t)$ is specified by $\hat{g}_j(t) = S_{j,n_j}(t-)$, where $S_{j,n_j}(t)$ is the Kaplan–Meier estimator of $S_j(t)$. Define $H_j(s) = S_j(s)(1 - G_j(s))$, let τ_1 be such that $S_j(\tau_1) < 1$, and let $\tau_2 \geq \tau_1$ be such that $H_j(\tau_2) > 0$, $j = 1, 2$. For future convenience, we define

$$
\gamma_j(t) \quad = \quad \int_0^t \frac{dF_j(s)}{1 - G_j(s-)}, \qquad (2.7)
$$

$\sigma^2_{\text{diff}}(t) = \gamma_1(t)/p_1 + \gamma_2(t)/p_2$ and $\sigma^2_{\text{ratio}}(t) = \gamma_1(t)/p_1 + \beta(t)^2 \gamma_2(t)/p_2$. These functions appear in the limiting distributions of the likelihood ratio statistics and need to be estimated. It can be shown (see Lemma A.3) that $\hat{\sigma}^2_{\text{diff}}(t) = n[\hat{\gamma}_1(t)/n_1 + \hat{\gamma}_2(t)/n_2]$ and $\hat{\sigma}^2_{\text{ratio}}(t) = n[\hat{\gamma}_1(t)/n_1 + \hat{\beta}(t)^2 \hat{\gamma}_2(t)/n_2]$ are uniformly consistent estimators of $\sigma^2_{\text{diff}}(t)$ and $\sigma^2_{\text{ratio}}(t)$ over $[\tau_1, \tau_2]$, where

$$
\hat{\gamma}_j(t) \quad = \quad \int_0^t \frac{dF_{j,n_j}(s)}{1 - G_{j,n_j}(s-)}, \qquad (2.8)
$$

F_{j,n_j} and G_{j,n_j} are the Kaplan–Meier estimators of F_j and G_j, and $\hat{\beta}(t) = F_{1,n_1}(t)/F_{2,n_2}(t)$.

2.2 Confidence bands

We now state our main results and explain how they can be used to construct the proposed simultaneous confidence bands.

Theorem 2.1. *The process* $-2\log R(\alpha(t), t, \hat{g}_1, \hat{g}_2)$, $t \in [\tau_1, \tau_2]$, *converges in distribution to* $U_1^2(t)/\sigma^2_{\text{diff}}(t)$ *in* $D[\tau_1, \tau_2]$, *where* $U_1(t)$ *is a Gaussian process with mean zero and covariance function* $\text{cov}(U_1(s), U_1(t)) = S_1(s)S_1(t)\sigma_1^2(s \wedge t)/p_1 + S_2(s)S_2(t)\sigma_2^2(s \wedge t)/p_2$ *and* $\sigma_j^2(t) = \int_0^t dF_j(s)/(S_j(s)H_j(s-))$.

It follows that

$$\sup_{t\in[\tau_1,\tau_2]} -2\log R(\alpha(t), t, \hat{g}_1, \hat{g}_2) \quad \xrightarrow{\mathcal{D}} \quad \sup_{t\in[\tau_1,\tau_2]} \frac{U_1^2(t)}{\sigma^2_{\text{diff}}(t)}$$

by the continuous mapping theorem, and we obtain

$$\mathcal{B}_{\text{diff}} = \{(t, \tilde{\alpha}(t)) : -2\log R(\tilde{\alpha}(t), t, \hat{g}_1, \hat{g}_2) \leq c_\alpha[\tau_1, \tau_2], \quad t \in [\tau_1, \tau_2]\} \tag{2.9}$$

as an asymptotic $100(1-\alpha)\%$ confidence band for $\alpha(t)$ over $[\tau_1, \tau_2]$, where the critical value $c_\alpha[\tau_1, \tau_2]$ is the upper α-quantile of the distribution of $\sup_{t\in[\tau_1,\tau_2]} U_1^2(t)/\sigma^2_{\text{diff}}(t)$. A simulation method is developed in Section 3 to obtain this critical value.

Implementation. We now explain how to compute the confidence band $\mathcal{B}_{\text{diff}}$. For fixed t, let $\phi(\lambda_n)$ denote the r.h.s. of (2.3). Its derivative is

$$\begin{aligned}\phi'(\lambda_n) &= \sum_{i=1}^{K_1(t)} \frac{2\lambda_n \hat{g}_1(T_{1i})d_{1i}}{(r_{1i} - d_{1i} + \lambda_n \hat{g}_1(T_{1i}))(r_{1i} + \lambda_n \hat{g}_1(T_{1i}))} \\ &+ \sum_{i=1}^{K_2(t)} \frac{2\lambda_n \hat{g}_2(T_{2i})d_{2i}}{(r_{2i} - d_{2i} - \lambda_n \hat{g}_2(T_{2i}))(r_{2i} - \lambda_n \hat{g}_2(T_{2i}))}\end{aligned}$$

As in McKeague and Zhao (2002), there exist exactly two roots $\lambda_L < 0 < \lambda_U$ for $\phi(\lambda_L) = \phi(\lambda_U) = c_\alpha[\tau_1, \tau_2]$ and $\{\lambda_n : \phi(\lambda_n) \leq c_\alpha[\tau_1, \tau_2]\} = [\lambda_L, \lambda_U]$.

The confidence set for $\alpha(t)$ is a closed interval $[\tilde{\alpha}_L, \tilde{\alpha}_U]$, where $\tilde{\alpha}_L = -\sum_{i=1}^{K_1(t)} \log\left(1 - d_{1i}/(r_{1i} + \lambda_L \hat{g}_1(T_{1i}))\right) \hat{g}_1(T_{1i}) + \sum_{i=1}^{K_2(t)} \log\left(1 - d_{2i}/(r_{2i} - \lambda_L \hat{g}_2(T_{2i}))\right) \hat{g}_2(T_{2i})$ and $\tilde{\alpha}_U$ is the same as $\tilde{\alpha}_L$, but with λ_L replaced by λ_U.

Next we state a parallel result for $\beta(t)$.

Theorem 2.2. *The process* $-2\log R(\beta(t), t, \hat{g}_1, \hat{g}_2)$, $t \in [\tau_1, \tau_2]$, *converges in distribution to* $U_2^2(t)/\sigma_{\text{ratio}}^2(t)$ *in* $D[\tau_1, \tau_2]$, *where* $U_2(t)$ *is a Gaussian process with mean zero and covariance function* $\text{cov}(U_2(s), U_2(t)) = S_1(s)S_1(t)\sigma_1^2(s\wedge t)/p_1 + S_2(s)S_2(t)\beta(s)\beta(t)\sigma_2^2(s\wedge t)/p_2$.

It follows that an asymptotic $100(1-\alpha)\%$ confidence band for $\beta(t)$ is given by

$$\mathcal{B}_{\text{ratio}} = \{(t, \tilde{\beta}(t)) : -2\log R(\tilde{\beta}(t), t, \hat{g}_1, \hat{g}_2) \leq K_\alpha[\tau_1, \tau_2], \quad t \in [\tau_1, \tau_2]\},$$

where $K_\alpha[\tau_1, \tau_2]$ is the upper α-quantile of the distribution of $\sup_{t\in[\tau_1,\tau_2]} U_2^2(t)/\sigma_{\text{ratio}}^2(t)$, which can be obtained by simulation, see Section 3.

3 Bootstrap procedure

The limiting distributions of the EL statistics in Section 2 are complicated and include unknown parameters, so we need to develop a Monte Carlo method to determine the critical values. To that end we adapt the Gaussian multiplier bootstrap procedure of Lin et al. (1993) for checking the adequacy of the Cox proportional hazards model.

First we define some (standard) counting process notation: $N_{ji}(t) = 1_{\{Z_{ji}\leq t, \delta_{ji}=1\}}$, $Y_{ji}(t) = 1_{\{Z_{ji}\geq t\}}$, and $Y_j(t) = \sum_{i=1}^{n_j} Y_{ji}(t)$ is the size of the risk set at $t-$. The processes $M_{ji}(s) = N_{ji}(s) - \int_0^s \alpha_j(s)Y_{ji}(s)\,ds$ are orthogonal locally square integrable martingales, where $\alpha_j(s)$ is hazard rate corresponding to F_j [see Andersen et al. (1993, II.4)].

From the proof of Theorem 2.1 in the Appendix, and using the martingale representation of $\sqrt{n_j}(S_{j,n_j}(t) - S_j(t))$, the process $U_1(t)/\sigma_{\text{diff}}(t)$ is seen to be asymptotically equivalent to

$$\frac{n^{1/2}}{\hat{\sigma}_{\text{diff}}(t)}\left(S_{1,n_1}(t)\int_0^t \frac{\sum_{i=1}^{n_1} dM_{1i}(s)}{Y_1(s)} + S_{2,n_2}(t)\int_0^t \frac{\sum_{i=1}^{n_2} dM_{2i}(s)}{Y_2(s)}\right).$$

DOI: 10.2202/1557-4679.1007

A version of this process that can be simulated is

$$U_1^*(t) = \frac{n^{1/2}}{\hat{\sigma}_{\text{diff}}(t)} \left(S_{1,n_1}(t) \int_0^t \frac{\sum_{i=1}^{n_1} G_{1i} dN_{1i}(s)}{Y_1(s)} + S_{2,n_2}(t) \int_0^t \frac{\sum_{i=1}^{n_2} G_{2i} dN_{2i}(s)}{Y_2(s)} \right),$$

where $G_{11}, \ldots, G_{1n_1}, G_{21}, \ldots, G_{2n_2}$ are i.i.d. $N(0,1)$ random variables independent of the data. Conditional on the data, the limiting distribution of $U_1^*(t)$ coincides with that of $U_1(t)/\sigma_{\text{diff}}(t)$ for almost all data sequences. This can be shown by noting that U_1^* is a Gaussian process with independent components whose covariance converges to that of U_1/σ_{diff} with probability 1, and by verifying tightness by applying the argument of Lin et al. (1993, Appendix 1) to each of the terms.

The bootstrap resampling method is then used to obtain $c_\alpha[\tau_1, \tau_2]$: generate a large number, say $L = 3000$, independent copies $U_{11}^*, \ldots, U_{1L}^*$ of U_1^* and take $c_\alpha[\tau_1, \tau_2]$ to be the upper α-quantile of

$$\sup_{t \in [\tau_1, \tau_2]} U_{11}^{*2}(t), \cdots, \sup_{t \in [\tau_1, \tau_2]} U_{1L}^{*2}(t).$$

A similar method gives the critical value $K_\alpha[\tau_1, \tau_2]$: use

$$U_2^*(t) = \frac{n^{1/2}}{\hat{\sigma}_{\text{ratio}}(t)} \left(S_{1,n_1}(t) \int_0^t \frac{\sum_{i=1}^{n_1} G_{1i}\, dN_{1i}(s)}{Y_1(s)} + S_{2,n_2}(t) \hat{\beta}(t) \int_0^t \frac{\sum_{i=1}^{n_2} G_{2i}\, dN_{2i}(s)}{Y_2(s)} \right),$$

as the bootstrap approximation for $U_2(t)/\sigma_{\text{ratio}}(t)$.

4 Numerical example

The data come from a Mayo Clinic trial involving a treatment for primary biliary cirrhosis of the liver, see Fleming and Harrington (1991) for discussion. A total of $n = 312$ patients participated in the randomized clinical trial, 158 receiving the treatment (D-penicillamine) and 154 receiving a placebo. Censoring is heavy (187 of the 312 observations are censored). We use $\tau_1 = 304$ and $\tau_2 = 4427$, respectively.

Figure 1 displays the proposed confidence band $\mathcal{B}_{\text{diff}}$ for the difference of the distribution functions (placebo minus treatment). The corresponding difference of the Kaplan–Meier curves is also displayed. Note that the simultaneous band contains the horizontal line (difference = 0), so there is no evidence of a difference between treatment and placebo. As expected, the confidence band is much narrower in the left tail than in the right.

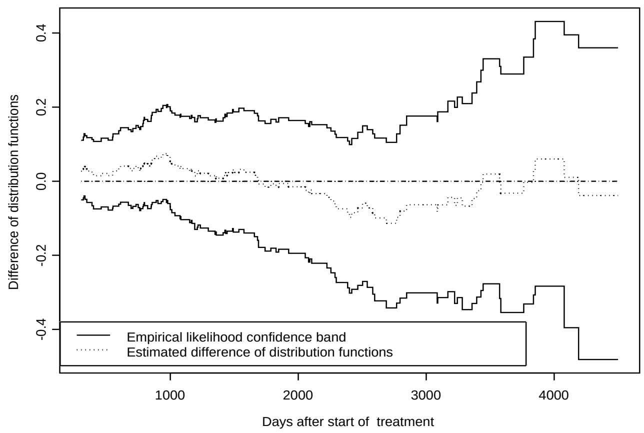

Figure 1: Mayo Clinic trial, 95% simultaneous confidence band for the difference of distribution functions (placebo−treatment).

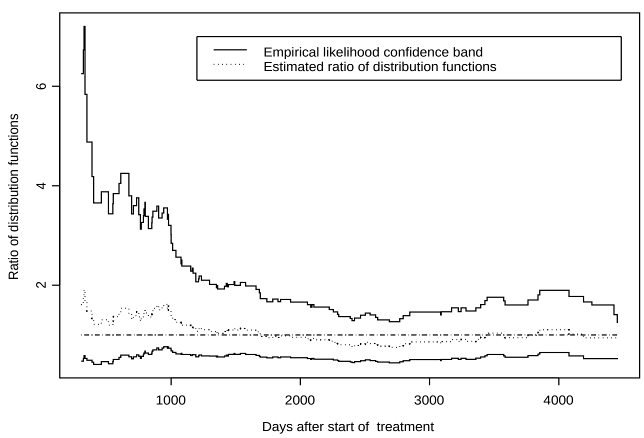

Figure 2: Mayo Clinic trial, 90% simultaneous confidence band for the ratio of distribution functions (placebo/treatment).

Figure 2 displays the proposed confidence band $\mathcal{B}_{\text{ratio}}$ for the ratio of the distribu-

DOI: 10.2202/1557-4679.1007

tion functions (placebo over treatment). The corresponding estimate of the ratio of the distribution functions is also displayed. Note that the simultaneous band contains the horizontal line (ratio = 1), so there is no evidence of a difference between treatment and placebo on the basis of this analysis. The lower bound of confidence band is greater than zero, which is within the range of parameter $\beta(t)$. As expected, the confidence band is much narrower in the right tail than in the left: the distribution function tends to zero in the left tail, so the variance of the ratio estimate blows up.

5 Discussion

This article has developed an empirical likelihood approach for comparing the distributions of survival times from two independent samples of right censored survival data in terms of ratios, differences, and other functionals of the underlying distribution functions. Our methods have potential application in epidemiological cohort studies, and in randomized clinical trials for the comparison of treatment and placebo groups. Standard approaches to making such comparisons have been via pointwise confidence intervals expressed in terms of Kaplan–Meier estimates, with the standard errors computed by the bootstrap or Greenwood's formula in conjunction with the delta method; see, e.g., Brenner and Hakulinen (2005) on the estimation of relative survival rates of cancer patients. We have shown, on the other hand, that it is possible to obtain simultaneous confidence bands in this setting without relying on the Wald approach of centering the confidence band on a point estimate of the target function (cf., Parzen et al., 1997). Our approach gives an added perspective to that obtained from other EL-type confidence bands for the comparison of survival distributions: Q-Q plots (Einmahl and McKeague, 1999), ratios of survival functions (McKeague and Zhao, 2002), and P-P plots (Claeskens et al., 2003), the latter only being available in the absence of censoring.

Our proposed confidence bands are computationally intensive compared with the closed form of Wald-type bands because they require the solution of a nonlinear equation at each uncensored survival time, and rely on the Gaussian multiplier simulation technique. For this reason, a simulation study to assess their performance would be time-consuming, and we have not carried out such a study for the present article. Nevertheless, based on a previous simulation study of an EL-type confidence band in a survival analysis setting (Hollander et al., 1997), we expect that the present EL bands will have significant advantages over their Wald-type counterparts in terms of coverage accuracy and adaption to skewness in the sampling distribution of the point estimates.

Acknowledgements

Ian McKeague acknowledges partial support under NSF grant DMS-0505201. The authors are grateful to the editor and two referees for their helpful comments.

Appendix: Proofs

We need the following lemma to prove Theorem 2.1.

Lemma A.1. *Under the assumptions of Theorem 2.1, the Lagrange multiplier solving (2.4) satisfies* $\lambda_n = \lambda_n(t) = O_P(n^{1/2})$ *uniformly over* $[\tau_1, \tau_2]$.

Proof. First assume $\lambda_n(t) < 0$. Along similar lines as Li (1995, p. 101–102), it can be shown that

$$-\sum_{i=1}^{K_1(t)} \log\left(1 - \frac{d_{1i}}{r_{1i} + \lambda_n(t)\hat{g}_1(T_{1i})}\right)\hat{g}_1(T_{1i}) \;\geq\; \sum_{i=1}^{K_1(t)} \frac{d_{1i}}{r_{1i}}\left(\frac{n_1}{n_1 - |\lambda_n(t)|\hat{g}_1(T_{1i})}\right)\hat{g}_1(T_{1i}),$$

and

$$\sum_{i=1}^{K_2(t)} \log\left(1 - \frac{d_{2i}}{r_{2i} - \lambda_n(t)\hat{g}_2(T_{2i})}\right)\hat{g}_2(T_{2i}) \;\geq\; -\sum_{i=1}^{K_2(t)} \frac{d_{2i}}{r_{2i}}\left(\frac{n_2}{n_2 + |\lambda_n(t)|\hat{g}_2(T_{2i})}\right)\hat{g}_2(T_{2i}) + \sum_{i=1}^{K_2(t)}\left(\log\left(1 - \frac{d_{2i}}{r_{2i}}\right) + \frac{d_{2i}}{r_{2i}}\right)\hat{g}_2(T_{2i}).$$

Combining the above two inequalities and (2.4) we get

$$\begin{aligned}\alpha(t) &\geq \sum_{i=1}^{K_1(t)} \frac{d_{1i}\hat{g}_1(T_{1i})}{r_{1i}} \left(\frac{n_1}{n_1 - |\lambda_n|\hat{g}_1(T_{1i})}\right) \\ &- \sum_{i=1}^{K_2(t)} \frac{d_{2i}\hat{g}_2(T_{2i})}{r_{2i}} \left(\frac{n_2}{n_2 + |\lambda_n|\hat{g}_2(T_{2i})}\right) + \sum_{i=1}^{K_2(t)} \frac{d_{2i}\hat{g}_2(T_{2i})}{r_{2i}} \\ &+ \sum_{i=1}^{K_2(t)} \hat{g}_2(T_{2i}) \log\left(1 - \frac{d_{2i}}{r_{2i}}\right). \end{aligned} \tag{A.1}$$

Denote $\mathcal{T} = [\tau_1, \tau_2]$, $\hat{\theta}_j(t) = \sum_{i=1}^{K_j(t)} \hat{g}_j(T_{ji})d_{ji}/r_{ji}$, and $\theta_j(t) = \int_0^t g_j(s)dA_j(s) = F_j(t)$, for $t \in \mathcal{T}$. Let $\tilde{\theta}_j(t) = -\sum_{i=1}^{K_j(t)} \log(1 - d_{ji}/r_{ji})\hat{g}_j(T_{ji})$. Using $1/(1+x) \leq 1$ for $x \geq 0$ and $1/(1-x) \geq 1 + x$ for $0 \leq x < 1$, from (A.1) we obtain

$$\alpha(t) + \tilde{\theta}_2(t) - \hat{\theta}_1(t) \geq \sum_{i=1}^{K_1(t)} \frac{d_{1i}\hat{g}_1^2(T_{1i})}{r_{1i}} \frac{|\lambda_n|}{n_1}. \tag{A.2}$$

In the case that $\lambda_n(t) \geq 0$, a similar argument leads to

$$\begin{aligned}-\alpha(t) &\geq -\sum_{i=1}^{K_1(t)} \frac{d_{1i}\hat{g}_1(T_{1i})}{r_{1i}} \left(\frac{n_1}{n_1 + |\lambda_n|\hat{g}_1(T_{1i})}\right) + \sum_{i=1}^{K_1(t)} \frac{d_{1i}\hat{g}_1(T_{1i})}{r_{1i}} \\ &+ \sum_{i=1}^{K_1(t)} \hat{g}_1(T_{1i}) \log\left(1 - \frac{d_{1i}}{r_{1i}}\right) + \sum_{i=1}^{K_2(t)} \frac{d_{2i}\hat{g}_2(T_{2i})}{r_{2i}} \left(\frac{n_2}{n_2 - |\lambda_n|\hat{g}_2(T_{2i})}\right). \end{aligned} \tag{A.3}$$

From (A.3), in a similar fashion to (A.2), we obtain

$$-\alpha(t) + \tilde{\theta}_1(t) - \hat{\theta}_2(t) \geq \sum_{i=1}^{K_2(t)} \frac{d_{2i}\hat{g}_2^2(T_{2i})}{r_{2i}} \frac{|\lambda_n|}{n_2}. \tag{A.4}$$

Next, in terms of the Nelson–Aalen estimator $\hat{A}_j$ of A_j, we have

$$\begin{aligned}\sqrt{n_j}(\hat{\theta}_j(t) - \theta_j(t)) &= \sqrt{n_j}\left(\int_0^t S_{j,n_j}(s-)\, d\hat{A}_j(s) - \int_0^t S_j(s)\, dA_j(s)\right) \\ &= \sqrt{n_j}(S_j(t) - S_{j,n_j}(t)),\end{aligned}$$

from the Volterra integral equation that relates the Nelson–Aalen estimator and the Kaplan–Meier estimator (see Andersen et al., 1993, p. 92), and hence $(\sqrt{n_j}(\hat{\theta}_j(t) - \theta_j(t)), j = 1, 2) \xrightarrow{\mathcal{D}} (S_j(t)U_j(t), j = 1, 2)$ in $D[\tau_1, \tau_2] \times D[\tau_1, \tau_2]$, where the $U_j(t)$ are independent Gaussian martingales with mean zero and $\mathrm{var}(U_j(s)) = \sigma_j^2(s)$ (Andersen et al., 1993, p. 263). By $n_j/n \to p_j > 0$, it follows that

$$\sqrt{n}\{[\hat\theta_1(t) - \theta_1(t)] - [\hat\theta_2(t) - \theta_2(t)]\} \xrightarrow{\mathcal{D}} \frac{S_1(t)U_1(t)}{\sqrt{p_1}} + \frac{S_2(t)U_2(t)}{\sqrt{p_2}},$$

or in terms of $\theta_1(t) - \theta_2(t) = \alpha(t)$, we have

$$\sqrt{n}(\hat\theta_1(t) - \hat\theta_2(t) - \alpha(t)) \xrightarrow{\mathcal{D}} \frac{S_1(t)U_1(t)}{\sqrt{p_1}} + \frac{S_2(t)U_2(t)}{\sqrt{p_2}}. \tag{A.5}$$

Using $|\log(1-x) + x| \le x^2$ for $0 \le x < 1$, we have

$$\begin{aligned}
\sqrt{n_j}|\hat\theta_j(t) - \tilde\theta_j(t)| &= \sqrt{n_j}\left|\sum_{i=1}^{K_j(t)} \log\left(1 - \frac{d_{ji}}{r_{ji}}\right)\hat g_j(T_{ji}) + \sum_{i=1}^{K_j(t)} \frac{d_{ji}}{r_{ji}}\hat g_j(T_{ji})\right| \\
&\le \left(\max_{i \le K_j(t)} \frac{d_{ji}\sqrt{n_j}}{r_{ji}}\right)\sum_{i=1}^{K_j(t)} \frac{d_{ji}\hat g_j(T_{ji})}{r_{ji}} \\
&= \left(\max_{i \le K_j(t)} \frac{\sqrt{n_j}}{r_{ji}}\right)\hat\theta_j(t) \xrightarrow{P} 0 \qquad (A.6)
\end{aligned}$$

uniformly in $t \in \mathcal{T}$, where in the last equality we use $d_{ji} = 1$ a.s., which is a consequence of the continuity of S_j, and in the final step we used the Glivenko–Cantelli theorem and the uniform convergence in probability of $\hat\theta_j(t)$ to $\theta_j(t)$ on $\mathcal{T}$.

Combining (A.5) and (A.6), we have

$$\begin{aligned}
\sqrt{n}(\alpha(t) - \hat\theta_1(t) + \tilde\theta_2(t)) &\xrightarrow{\mathcal{D}} \frac{S_1(t)U_1(t)}{\sqrt{p_1}} + \frac{S_2(t)U_2(t)}{\sqrt{p_2}}, \\
\sqrt{n}(\alpha(t) - \tilde\theta_1(t) + \hat\theta_2(t)) &\xrightarrow{\mathcal{D}} \frac{S_1(t)U_1(t)}{\sqrt{p_1}} + \frac{S_2(t)U_2(t)}{\sqrt{p_2}}.
\end{aligned}$$

Thus the l.h.s. of (A.2) and (A.4) are $O_P(n^{-1/2})$ uniformly for $t \in \mathcal{T}$.

Applying Lenglart's inequality to the martingale integral $\int_0^t \hat g_j^k(s)\, d(\hat A_j - A_j)(s)$ (cf. Andersen et al., 1993, p. 190), where $k \ge 1$, shows that it converges to zero uniformly in probability over $t \in \mathcal{T}$. Since $S_{j,n_j}(t)$ is a uniformly consistent estimator of $S_j(t)$, we have that $\hat g_j^k(t)$ is a uniformly consistent estimator of $g_j^k(t)$. Thus, $\int_0^t (\hat g_j^k(s) - g_j^k(s))\, dA_j(s) \to 0$ uniformly in probability over $t \in \mathcal{T}$, and

$$\sum_{i=1}^{K_j(t)} \frac{d_{ji}\hat g_j^k(T_{ji})}{r_{ji}} \xrightarrow{P} \int_0^t g_j^k(s)dA_j(s) \tag{A.7}$$

DOI: 10.2202/1557-4679.1007

uniformly over $t \in \mathcal{T}$, for $j = 1, 2$ and any fixed $k \geq 1$.

Using $n_j/n \to p_j > 0$, (A.2), (A.4), (A.7), we find that $\lambda_n = O_P(n^{1/2})$ uniformly for $t \in \mathcal{T}$. □

Proof of Theorem 2.1. Let $f_j(\lambda_n) = \sum_{i=1}^{K_j(t)} \log(1 - d_{ji}/(r_{ji} + \lambda_n \hat{g}_j(T_{ji})))\hat{g}_j(T_{ji})$, and recall $\tilde{\theta}_j(t) = -\sum_{i=1}^{K_j(t)} \log(1 - d_{ji}/r_{ji})\hat{g}_j(T_{ji})$. Then $f_j(0) = \tilde{\theta}_j(t)$, $f_j'(0) = \tilde{\gamma}_j(t)/n_j$, where

$$\tilde{\gamma}_j(t) \quad = \quad n_j \sum_{i:T_{ji} \leq t} \frac{\hat{g}_j^2(T_{ji}) d_{ji}}{r_{ji}(r_{ji} - d_{ji})}. \tag{A.8}$$

is a uniformly consistent estimator of $\gamma_j(t)$ over $t \in [\tau_1, \tau_2]$. This is proved in Lemma A.3.

For any $\lambda_n = O_P(n^{1/2})$, by a Taylor expansion we have

$$f_j((-1)^{j-1}\lambda_n) \quad = \quad -\tilde{\theta}_j(t) + \frac{\tilde{\gamma}_j(t)(-1)^{j-1}\lambda_n}{n_j} + \frac{f_j''(\xi_{jn})\lambda_n^2}{2}, \tag{A.9}$$

where $|\xi_{jn}| \leq |\lambda_n|$. By (A.7) with $k = 3$, $f_j''(\xi_{jn})\lambda_n^2 = O_P(n_j^{-1})$. Using $n_j/n \to p_j > 0$, (2.4) and (A.9) we then obtain

$$-\alpha(t) \quad = \quad -\tilde{\theta}_1(t) + \tilde{\theta}_2(t) + \frac{\tilde{\gamma}_1(t)\lambda_n}{n_1} + \frac{\tilde{\gamma}_2(t)\lambda_n}{n_2} + O_P(1). \tag{A.10}$$

It follows from (A.10) that

$$\lambda_n \quad = \quad -n\hat{\sigma}_{\text{diff}}^{-2}(\alpha(t) - \tilde{\theta}_1(t) + \tilde{\theta}_2(t) + O_P(n^{-1})). \tag{A.11}$$

Since $\lambda_n \hat{g}_j(T_{ji})/(r_{ji} - d_{ji}) = o_P(1)$ and $\lambda_n \hat{g}_j(T_{ji})/r_{ji} = o_P(1)$ uniformly in $i = 1, \ldots, K_j(\tau_2)$, using a Taylor expansion for (2.3) and (A.11) we obtain

$$\begin{aligned} -2\log R(\alpha(t), t, \hat{g}_1, \hat{g}_2) \quad &= \quad \lambda_n^2 \left(\sum_{i=1}^{K_1(t)} \frac{\hat{g}_1^2(T_{1i}) d_{1i}}{r_{1i}(r_{1i} - d_{1i})} + \sum_{i=1}^{K_2(t)} \frac{\hat{g}_2^2(T_{2i}) d_{2i}}{r_{2i}(r_{2i} - d_{2i})} \right) \\ &- \quad \frac{2\lambda_n^3}{3} \sum_{i=1}^{K_1(t)} \hat{g}_1^3(T_{1i}) \left(\frac{1}{(r_{1i} - d_{1i})^2} - \frac{1}{r_{1i}^2} \right) \end{aligned}$$

$$
\begin{aligned}
&+ \quad \frac{2\lambda_n^3}{3}\sum_{i=1}^{K_2(t)} \hat{g}_2^3(T_{2i})\left(\frac{1}{(r_{2i}-d_{2i})^2}-\frac{1}{r_{2i}^2}\right)\\
&+ \quad \frac{\lambda_n^4}{2}\sum_{i=1}^{K_1(t)} \hat{g}_1^4(T_{1i})\left(\frac{1}{(r_{1i}-d_{1i})^3}-\frac{1}{r_{1i}^3}\right)\\
&+ \quad \frac{\lambda_n^4}{2}\sum_{i=1}^{K_2(t)} \hat{g}_2^4(T_{2i})\left(\frac{1}{(r_{2i}-d_{2i})^3}-\frac{1}{r_{2i}^3}\right)+o_P(1)\\
&= \quad n\hat{\sigma}_{\text{diff}}^{-2}(\alpha(t)-\tilde{\theta}_1(t)+\tilde{\theta}_2(t)+O_P(n^{-1}))^2+o_P(1),
\end{aligned}
$$

where in the last equality we use (A.7) for $k = 3, 4$. Combining (A.5), (A.6) and the uniform consistency of $\hat{\sigma}^2_{\text{diff}}(t)$ shows that the above process has the limiting distribution indicated in the theorem. □

In order to prove Theorem 2.2 we need the following lemma.

Lemma A.2. *Under the assumptions of Theorem 2.2, the Lagrange multiplier solving (2.5) satisfies $\lambda_n = \lambda_n(t) = O_P(n^{1/2})$ uniformly over $[\tau_1, \tau_2]$.*

Proof. The proof follows a similar pattern to the proof of Lemma A.1. First assume $\lambda_n(t) < 0$. Then, as in Li (1995, p. 101–102),

$$
-\sum_{i=1}^{K_1(t)} \log\left(1-\frac{d_{1i}}{r_{1i}+\lambda_n(t)\hat{g}_1(T_{1i})}\right)\hat{g}_1(T_{1i}) \quad \geq \quad \sum_{i=1}^{K_1(t)} \frac{d_{1i}}{r_{1i}}\left(\frac{n_1}{n_1-|\lambda_n(t)|\hat{g}_1(T_{1i})}\right)\hat{g}_1(T_{1i})
$$

and

$$
\begin{aligned}
\sum_{i=1}^{K_2(t)} \log\left(1-\frac{d_{2i}}{r_{2i}-\beta(t)\lambda_n(t)\hat{g}_2(T_{2i})}\right)\hat{g}_2(T_{2i}) \quad &\geq \quad \sum_{i=1}^{K_2(t)}\left(\log\left(1-\frac{d_{2i}}{r_{2i}}\right)+\frac{d_{2i}}{r_{2i}}\right)\hat{g}_2(T_{2i})\\
&- \quad \sum_{i=1}^{K_2(t)} \frac{d_{2i}}{r_{2i}}\left(\frac{n_2\hat{g}_2(T_{2i})}{n_2+\beta(t)|\lambda_n(t)|\hat{g}_2(T_{2i})}\right).
\end{aligned}
$$

Combining the above two inequalities and (2.6), using $1/(1 + x) \leq 1$ for $x \geq 0$ and $1/(1 - x) \geq 1 + x$ for $0 \leq x < 1$, we obtain

$$
-\hat{\theta}_1(t)+\beta(t)\tilde{\theta}_2(t) \quad \geq \quad \sum_{i=1}^{K_1(t)} \frac{d_{1i}\hat{g}_1^2(T_{1i})}{r_{1i}}\frac{|\lambda_n|}{n_1}. \tag{A.12}
$$

Second, supposing $\lambda_n(t) \geq 0$, a similar argument leads to

$$\begin{aligned} -\beta(t) \sum_{i=1}^{K_2(t)} \frac{d_{2i}\hat{g}_2(T_{2i})}{r_{2i}} \left(\frac{n_2}{n_2 - \beta(t)|\lambda_n|\hat{g}_2(T_{2i})} \right) \quad &\geq \quad - \sum_{i=1}^{K_1(t)} \frac{d_{1i}\hat{g}_1(T_{1i})}{r_{1i}} \left(\frac{n_1}{n_1 + |\lambda_n|\hat{g}_1(T_{1i})} \right) \\ &+ \quad \sum_{i=1}^{K_1(t)} \frac{d_{1i}\hat{g}_1(T_{1i})}{r_{1i}} \\ &+ \quad \sum_{i=1}^{K_1(t)} \hat{g}_1(T_{1i}) \log \left(1 - \frac{d_{1i}}{r_{1i}} \right). \qquad \text{(A.13)} \end{aligned}$$

In a similar fashion to (A.12), from (A.13) we obtain

$$\tilde{\theta}_1(t) - \beta(t)\hat{\theta}_2(t) \quad \geq \quad \beta^2(t) \sum_{i=1}^{K_2(t)} \frac{d_{2i}\hat{g}_2^2(T_{2i})}{r_{2i}} \frac{|\lambda_n|}{n_2}. \qquad \text{(A.14)}$$

By $n_j/n \to p_j > 0$, we have

$$\sqrt{n}\{[\hat{\theta}_1(t) - \theta_1(t)] - \beta(t)[\hat{\theta}_2(t) - \theta_2(t)]\} \quad \xrightarrow{\mathcal{D}} \quad \frac{S_1(t)U_1(t)}{\sqrt{p_1}} + \frac{\beta(t)S_2(t)U_2(t)}{\sqrt{p_2}},$$

or in terms of $\theta_1(t)/\theta_2(t) = \beta(t)$, we have

$$\sqrt{n}(\hat{\theta}_1(t) - \beta(t)\hat{\theta}_2(t)) \quad \xrightarrow{\mathcal{D}} \quad \frac{S_1(t)U_1(t)}{\sqrt{p_1}} + \frac{\beta(t)S_2(t)U_2(t)}{\sqrt{p_2}}. \qquad \text{(A.15)}$$

Combining (A.6) and (A.15) gives

$$\begin{aligned} \sqrt{n}(\hat{\theta}_1(t) - \beta(t)\tilde{\theta}_2(t)) \quad &\xrightarrow{\mathcal{D}} \quad \frac{S_1(t)U_1(t)}{\sqrt{p_1}} + \frac{\beta(t)S_2(t)U_2(t)}{\sqrt{p_2}}, \\ \sqrt{n}(\tilde{\theta}_1(t) - \beta(t)\hat{\theta}_2(t)) \quad &\xrightarrow{\mathcal{D}} \quad \frac{S_1(t)U_1(t)}{\sqrt{p_1}} + \frac{\beta(t)S_2(t)U_2(t)}{\sqrt{p_2}}. \end{aligned}$$

Thus the l.h.s. of (A.12) and (A.14) are $O_P(n^{-1/2})$. Using $n_j/n \to p_j > 0$, $\beta(t) \geq \theta_1(\tau_1)/\theta_2(\tau_2) = F_1(\tau_1)/F_2(\tau_2) > 0$, (A.7), (A.12), and (A.14), we find that $\lambda_n = O_P(n^{1/2})$ uniformly for $t \in \mathcal{T}$. □

Proof of Theorem 2.2. This proof is a variation of the proof of Theorem 2.1. Recall $f_j'(0) = \tilde{\gamma}_j(t)/n_j$ and note that $0 \leq \beta(t) \leq \theta_1(\tau_2)/\theta_2(\tau_1)$, $t \in \mathcal{T}$. For any $\lambda_n = O_P(n^{1/2})$, by a Taylor expansion we have

$$f_j((-\beta)^{j-1}(t)\lambda_n) \quad = \quad -\tilde{\theta}_j(t) + \frac{\tilde{\gamma}_j(t)(-\beta(t))^{j-1}\lambda_n}{n_j} + \frac{f_j''((-\beta(t))^{j-1}\xi_{jn})(\beta(t))^{2(j-1)}\lambda_n^2}{2}, \qquad \text{(A.16)}$$

where $|\xi_{jn}| \leq |\lambda_n|$. By (A.7) with $k = 3$, $f_j''(\xi_{jn})\lambda_n^2 = O_P(n_j^{-1})$, and using $n_j/n \to p_j > 0$, (2.6) and (A.16), we obtain

$$0 = -\tilde{\theta}_1(t) + \beta(t)\tilde{\theta}_2(t) + \frac{\tilde{\gamma}_1(t)\lambda_n}{n_1} + \frac{\tilde{\gamma}_2(t)\beta^2(t)\lambda_n}{n_2} + O_P(1). \qquad (A.17)$$

It follows from (A.17) that

$$\lambda_n = n\hat{\sigma}_{\text{ratio}}^{-2}(\tilde{\theta}_1(t) - \beta(t)\tilde{\theta}_2(t) + O_P(n^{-1})). \qquad (A.18)$$

Since $\lambda_n\hat{g}_j(T_{ji})/(r_{ji} - d_{ji}) = o_P(1)$ and $\lambda_n\hat{g}_j(T_{ji})/r_{ji} = o_P(1)$ uniformly in $i = 1, \ldots, K_j(\tau_2)$, using a Taylor expansion for (2.5) and (A.18) we have

$$\begin{aligned}
-2\log R(\beta(t), t, \hat{g}_1, \hat{g}_2) &= \lambda_n^2 \left(\sum_{i=1}^{K_1(t)} \frac{\hat{g}_1^2(T_{1i})d_{1i}}{r_{1i}(r_{1i} - d_{1i})} + \sum_{i=1}^{K_2(t)} \frac{\beta^2(t)\hat{g}_2^2(T_{2i})d_{2i}}{r_{2i}(r_{2i} - d_{2i})} \right) \\
&- \frac{2\lambda_n^3}{3} \sum_{i=1}^{K_1(t)} \hat{g}_1^3(T_{1i}) \left(\frac{1}{(r_{1i} - d_{1i})^2} - \frac{1}{r_{1i}^2} \right) \\
&+ \frac{2\beta^3(t)\lambda_n^3}{3} \sum_{i=1}^{K_2(t)} \hat{g}_2^3(T_{2i}) \left(\frac{1}{(r_{2i} - d_{2i})^2} - \frac{1}{r_{2i}^2} \right) \\
&+ \frac{\lambda_n^4}{2} \sum_{i=1}^{K_1(t)} \hat{g}_1^4(T_{1i}) \left(\frac{1}{(r_{1i} - d_{1i})^3} - \frac{1}{r_{1i}^3} \right) \\
&+ \frac{\beta^4(t)\lambda_n^4}{2} \sum_{i=1}^{K_2(t)} \hat{g}_2^4(T_{2i}) \left(\frac{1}{(r_{2i} - d_{2i})^3} - \frac{1}{r_{2i}^3} \right) + o_P(1) \\
&= n\hat{\sigma}_{\text{ratio}}^{-2}(\tilde{\theta}_1(t) - \beta(t)\tilde{\theta}_2(t) + O_P(n^{-1}))^2 + o_P(1),
\end{aligned}$$

where in the last equality we use (A.7) for $k = 3, 4$. Combining (A.6), (A.15) and the uniform consistency of $\hat{\sigma}_{\text{ratio}}^2(t)$ completes the proof. □

Lemma A.3. *The estimators $\hat{\gamma}_j$ and $\tilde{\gamma}_j$ defined in (2.8) and (A.8), respectively, converge uniformly in probability to γ_j over $[\tau_1, \tau_2]$.*

Proof. Note that $\phi(F_j, G_j)(t) = \int_0^t dF_j(s)/(1 - G_j(s-))$, $t \in [0, \tau_2]$, is a continuous functional of cdfs F_j and G_j in supremum norm (Andersen et al., 1993, Proposition

II.8.6., p. 113), so the result for $\hat{\gamma}_j$ follows from the uniform consistency of the Kaplan–Meier estimators of F_j and G_j on $[0, \tau_2]$. The result for $\tilde{\gamma}_j$ can be proved by adapting the argument on pages 191–192 of Andersen et al. (1993). □

References

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

Brenner, H., and Hakulinen, T. (2005) Substantial overestimation of standard errors of relative survival rates of cancer patients. *Amer. J. Epidemiol.*, **161**, 781–786.

Claeskens, G., Jing, B.-Y., Peng, L., and Zhou, W. (2003) Empirical likelihood confidence regions for comparison distributions and ROC curves. *Canadian J. Statist.*, **31**, 173–190.

Einmahl, J.H., and McKeague, I.W. (1999) Confidence tubes for multiple quantile plots via empirical likelihood. *Ann. Statist.*, **27**, 1348–1367.

Fleming, T.R., and Harrington, D.P. (1991) *Counting Processes and Survival Analysis*, New York: Wiley.

Hjort, N. L., McKeague, I. W., and Van Keilegom, I. (2005) Extending the scope of empirical likelihood. Unpublished manuscript.

Hollander, M., McKeague, I. W., and Yang, J. (1997) Likelihood ratio based confidence bands for survival functions. *J. Amer. Statist. Assoc.*, **92**, 215–226.

Li, G. (1995) On nonparametric likelihood ratio estimation of survival probabilities for censored data. *Statist. Probab. Lett.*, **25**, 95–104.

Lin, D.Y., Wei, L.J., and Ying, Z. (1993) Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika*, **80**, 557–572.

McKeague, I.W., and Zhao, Y. (2002) Simultaneous confidence bands for ratios of survival functions via empirical likelihood. *Statist. Probab. Lett.*, **60**, 405–415.

Murphy, S. (1995) Likelihood ratio-based confidence intervals in survival analysis. *J. Amer. Statist. Assoc.*, **90**, 1399–1405.

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

Owen, A. (1990) Empirical likelihood and confidence regions. *Ann. Statist.*, **18**, 90–120.

Owen, A. (2001) *Empirical Likelihood*, Chapman & Hall: New York.

Pan, X. R., and Zhou, M. (2002) Empirical likelihood ratio in terms of cumulative hazard function for censored data. *J. Multi. Anal.*, **80**, 166–188.

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

Thomas, D.R., and Grunkemeier, G.L. (1975) Confidence interval estimation for survival probabilities for censored data. *J. Amer. Statist. Assoc.*, **70**, 865–871.