Available online at www.sciencedirect.com
SCIENCE DIRECT®
Statistics & Probability Letters 76 (2006) 327–339
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Width-scaled confidence bands for survival functions

Ian W. McKeague[a,1], Yichuan Zhao[b,*]

[a]Department of Biostatistics, Columbia University, New York, NY 10032, USA
[b]Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, USA

Received 11 August 2003; received in revised form 30 September 2004
Available online 24 August 2005

Abstract

A popular simultaneous confidence band for survival functions is the equal precision band of Nair [1984. Confidence bands for survival functions with censored data: a comparative study. Technometrics 26, 265–275] This band is found by adjusting the level of Wald-type pointwise confidence intervals centered on the Kaplan–Meier estimator. The present paper develops a complementary method of adjusting pointwise confidence intervals to produce a simultaneous band. Our approach is to scale the *width*, rather than the level, of the pointwise confidence intervals. The resulting adjustment of the pointwise band, called a width-scaled band, provides an attractive alternative to the equal precision band. Empirical likelihood-based width-scaled bands are studied in the one- and two-sample censored data settings. An example with real data is included.

Keywords: Empirical likelihood; Equal precision band; Simultaneous inference; Right censoring

1. Introduction

Simultaneous confidence bands for survival functions play an important role in clinical studies and medical decision making. For example, they are useful in quantifying a patient's prognosis at stages of a disease or under various prospective treatments, and provide simultaneous coverage

*Corresponding author. Tel.: +1 404 654 5877; fax: +1 404 651 2246.
E-mail address: yzhao@mathstat.gsu.edu (Y. Zhao).
[1]Acknowledges partial support under NSF Grant DMS-0204688.

0167-7152/$ - see front matter
doi:10.1016/j.spl.2005.08.006

throughout the follow-up period. Wald-type bands centered on the Kaplan–Meier estimator were originally developed by Hall and Wellner (1980). An alternative Wald-type band, known as the equal precision band, was formulated by Nair (1984). More recently, Hollander et al. (1997) derived a band based on the empirical likelihood (EL) approach.

In the present paper, we introduce a new type of simultaneous confidence band for survival functions and ratios of survival functions. Our approach is related to the equal precision band, which adjusts the level of pointwise confidence intervals to furnish a simultaneous band. Instead of adjusting the level, however, we propose to adjust the *width* of the pointwise confidence intervals. The resulting band, called a width-scaled band, is visually consistent with the pointwise band and is more stable in the tails than the band derived using the standard maximal deviation statistic approach.

We focus on EL width-scaled bands for right censored data. EL arose in the classical survival analysis setting in the work of Thomas and Grunkemeier (1975), and was generalized by Owen (1988, 1990) to a variety of statistical problems. Other background references on EL relevant for the present paper are Li (1995), Einmahl and McKeague (1999), Li and Van Keilegom (2002), and McKeague and Zhao (2002).

The general idea behind the construction of width-scaled EL bands for an arbitrary target function is presented in Section 2. The band is studied in detail in the two sample setting for ratios of survival functions and it is shown how the band can be adjusted for a covariate effect. In Section 3, we give an illustrative example. The proposed band is found to be effective for reducing instability in the tails, especially when estimating ratios of survival functions or ratios of cumulative hazard functions. The main step in the construction (placed in Appendix A) involves a weak limit of the EL statistic evaluated at a re-centered target function.

2. Proposed width-scaled bands

We introduce the main idea in terms of a generic EL $L(\eta)$, where η represents a model specific distribution that is identifiable from the available data (n i.i.d. observations) and from which some target function of interest, $\theta(t) = \theta(t|\eta)$, is specified. Here, t is restricted to a given interval $[\tau_1, \tau_2]$. The EL ratio for $\theta(t)$ is

$$R(\tilde{\theta}(t), t) = \frac{\sup\{L(\eta): \theta(t|\eta) = \tilde{\theta}(t)\}}{\sup\{L(\eta)\}},$$

where the suprema range over all distributions η. The pointwise EL confidence interval for $\theta(t)$ is based on a version of Wilks's theorem for the EL statistic:

$$-2\log R(\theta(t), t) \xrightarrow{\mathrm{D}} \chi_1^2 \tag{2.1}$$

as $n \to \infty$, for each fixed $t \in [\tau_1, \tau_2]$. We impose (2.1) as a condition and also assume that the EL statistic is convex as a function of $\theta(t)$; these assumptions hold in the models considered below. Let $\theta_{\mathrm{L}}(t)$ and $\theta_{\mathrm{U}}(t)$ denote the lower/upper bounds of the $100(1-\alpha)\%$ EL-based pointwise confidence interval for $\theta(t)$:

$$\{\theta(t) : -2\log R(\theta(t), t) \leqslant \chi^2_{1,\alpha}\} = [\theta_L(t), \theta_U(t)],$$

where $\chi^2_{1,\alpha}$ is the upper α-quantile of the chi-square distribution with 1 degree of freedom.

The idea behind the new simultaneous confidence band is to re-scale the pointwise EL band about an estimator $\theta_n(t)$ of $\theta(t)$, the nonparametric maximum likelihood estimator being the usual choice for $\theta_n(t)$. The lower and upper bounds of the proposed band are given by

$$\theta_L^{WS}(t) = \theta_n(t) + \hat{\gamma}(\theta_L(t) - \theta_n(t)),$$
$$\theta_U^{WS}(t) = \theta_n(t) + \hat{\gamma}(\theta_U(t) - \theta_n(t)), \tag{2.2}$$

respectively, where $\hat{\gamma} \geqslant 1$ is a data-driven "inflation" factor that is specified below. Let $\tilde{\theta}(t) = (\theta(t) - (1-\gamma)\theta_n(t))/\gamma$, where $\gamma \geqslant 1$. In addition to the above conditions, assume that

$$-2\log R(\tilde{\theta}(t), t) \xrightarrow{D} U^2(t)/(\gamma^2\sigma^2(t)) \tag{2.3}$$

in $D[\tau_1, \tau_2]$, where $U(t)$ is a continuous Gaussian martingale with mean zero and variance function $\sigma^2(t)$ that can be uniformly consistently estimated by $\hat{\sigma}^2(t)$. Here, the process $U(t)$ can be replaced by $B(\sigma^2(t))$, where $B(t)$ is a standard Wiener process. This condition is seen to be stronger than (2.1) by setting $\gamma = 1$. Let $\hat{c}(\alpha)$ be the estimate of the upper α-quantile $c(\alpha)$ of

$$\sup_{t\in[\tau_1,\tau_2]} B^2(\sigma^2(t))/\sigma^2(t) = \sup_{s\in[\sigma^2(\tau_1),\sigma^2(\tau_2)]} B^2(s)/s \tag{2.4}$$

obtained by plugging-in $\hat{\sigma}^2(t)$ in place of $\sigma^2(t)$. These quantiles can be readily computed via simulation.

In order to specify a $100(1-\alpha)\%$ width-scaled confidence band we set $\hat{\gamma} = \sqrt{\hat{c}(\alpha)}/z_{\alpha/2}$, where $z_{\alpha/2}$ is the upper α-quantile of standard normal. To see that this choice of $\hat{\gamma}$ gives the desired simultaneous coverage probability, note that by (2.2)

$$\begin{aligned}
P(\theta_L^{WS}(t) \leqslant \theta(t) \leqslant \theta_U^{WS}(t),\ t \in [\tau_1, \tau_2]) &= P(\theta_n(t) + \hat{\gamma}(\theta_L(t) - \theta_n(t)) \leqslant \theta(t) \leqslant \theta_n(t) \\
&\qquad + \hat{\gamma}(\theta_U(t) - \theta_n(t)),\ t \in [\tau_1, \tau_2]) \\
&= P(\theta_L(t) \leqslant \tilde{\theta}(t) \leqslant \theta_U(t),\ t \in [\tau_1, \tau_2]) \\
&= P(-2\log R(\tilde{\theta}(t), t) \leqslant \chi^2_{1,\alpha},\ t \in [\tau_1, \tau_2]) \\
&= P\left(\sup_{t\in[\tau_1,\tau_2]} -2\log R(\tilde{\theta}(t), t) \leqslant \hat{c}(\alpha)/\hat{\gamma}^2\right) \\
&\longrightarrow P\left(\sup_{t\in[\tau_1,\tau_2]} U^2(t)/\sigma^2(t) \leqslant c(\alpha)\right) = 1 - \alpha,
\end{aligned}$$

where in the last three steps we used the specified form of $\hat{\gamma}$, the consistency of $\hat{c}(\alpha)$, and condition (2.3) with $\gamma = \sqrt{c(\alpha)}/z_{\alpha/2}$. This completes the construction of the proposed width-scaled simultaneous confidence band

$$\{(t, \theta(t)) : \theta_L^{WS}(t) \leqslant \theta(t) \leqslant \theta_U^{WS}(t), t \in [\tau_1, \tau_2]\}.$$

We now specialize to the standard one-sample framework with independent right censoring and consider the problem of finding a confidence band for the survival function $\theta(t) = S(t)$. The observations are i.i.d. of the form (Z_i, δ_i), $i = 1, \ldots, n$, where $Z_i = X_i \wedge Y_i$, $\delta_i = 1_{\{X_i \leqslant Y_i\}}$, and the X_i, Y_i are independent and nonnegative. The distribution functions of X_i and Y_i are denoted F and G, respectively. The survival function $S = 1 - F$ is assumed to be continuous. Let $H(s) = S(s)(1 - G(s))$, and assume $S(\tau_1) < 1$, $H(\tau_2) > 0$.

The EL for S is given by

$$L(S) = \prod_{i=1}^{n} \{S(Z_i-) - S(Z_i)\}^{\delta_i} S(Z_i)^{1-\delta_i}, \tag{2.5}$$

the nonparametric maximum likelihood estimator is the Kaplan–Meier estimator $S_n(t)$, and the EL ratio for $S(t)$ is

$$R(\tilde{S}(t), t) = \frac{\sup\{L(S) : \tilde{S}(t) = S(t)\}}{\sup\{L(S)\}}, \tag{2.6}$$

where the suprema range over all survival functions. Next,

$$\sigma^2(t) = \int_0^t \frac{\mathrm{d}F(s)}{S(s)H(s-)}, \quad t \in [\tau_1, \tau_2],$$

and, by Andersen et al. (1993, IV.1.3),

$$\hat{\sigma}^2(t) = n \sum_{i:T_i \leqslant t} \frac{d_i}{r_i(r_i - d_i)}$$

is a uniformly consistent estimator of $\sigma^2(t)$, $t \in [\tau_1, \tau_2]$, where d_i and r_i are defined in Appendix A. Results of Thomas and Grunkemeier (1975) and Li (1995) imply that condition (2.1) is satisfied. Condition (2.3) is checked in Appendix A.

It is natural to ask what happens when we width-scale the *Wald-type* pointwise confidence interval $S_n(t) \pm z_{\alpha/2} S_n(t)\hat{\sigma}(t)n^{-1/2}$ about $S_n(t)$. In this case, the width-scale band reduces to Nair's (1984) equal precision band

$$\{(t, S(t)) : |S(t) - S_n(t)| \leqslant \hat{\gamma} z_{\alpha/2} S_n(t)\hat{\sigma}(t)n^{-1/2}, t \in [\tau_1, \tau_2]\}.$$

For two independent samples of right-censored data, the above assumptions are duplicated for the second sample, and the notation is supplied with a further subscript $j = 1, 2$ to identify sample membership. Assume that $n_j/n \to p_j > 0$ as $n = n_1 + n_2 \to \infty$. The EL $L(S_1, S_2)$ for the two survival functions is given by the product of the two (marginal) ELs, and the EL ratio of a functional $\theta(t) = \theta(t|S_1(t), S_2(t))$ for a given $t \geqslant 0$ is defined by

$$R(\tilde{\theta}(t), t) = \frac{\sup\{L(S_1, S_2) : \theta(t) = \tilde{\theta}(t)\}}{\sup\{L(S_1, S_2)\}}, \tag{2.7}$$

where the suprema range over all pairs of survival functions. A width-scaled EL confidence band for the ratio of the two survival functions $\theta(t) = S_1(t)/S_2(t)$ can be derived using results of McKeague and Zhao (2002). Let $S_{jn_j}(t)$ be the Kaplan–Meier estimator of $S_j(t)$, and $\theta_{\mathrm{L}}(t)$ and $\theta_{\mathrm{U}}(t)$ denote the lower/upper bounds of McKeague and Zhao's $100(1-\alpha)\%$ EL pointwise confidence band for $\theta(t)$. The simultaneous confidence band re-scales the pointwise band about $\theta_n(t) = S_{1n_1}(t)/S_{2n_2}(t)$, and has bounds given by (2.2) with $\hat{\gamma}$ specified as before, except $\hat{\sigma}^2(t) = n[\hat{\sigma}_1^2(t)/n_1 + \hat{\sigma}_2^2(t)/n_2]$ replaces the previous $\hat{\sigma}^2(t)$. Condition (2.3) can be checked in a similar way to the one-sample case.

We next show how the width-scaled confidence band for the ratio of the two survival functions can be adjusted for covariate effects. We adopt the approach of Li and Van Keilegom (2002), henceforth LV, who developed an EL confidence band for a single conditional survival function. Now define $\theta(t) = S_1(t|v)/S_2(t|v)$, where $S_j(t|v) = P(X_{ji} > t | V_{ji} = v)$ is the conditional survival function of X_{ji} given a fixed level v of a one-dimensional covariate V_{ji}. The level v can be allowed to depend on j, but for simplicity we take it to be the same for both samples. We assume that the observations $(Z_{ji}, \delta_{ji}, V_{ji})$ are i.i.d. for $i = 1, \ldots, n_j$, independent for $j = 1, 2$, and X_{ji} is conditionally independent of Y_{ji} given V_{ji}.

The EL ratio is defined as in (2.7), except that S_j is replaced by $S_j(\cdot|v)$ and the two marginal ELs are now replaced by LV's local EL for each sample. The Nadaraya–Watson weights needed for this local EL are defined (see the Appendix A) in terms of a (user-specified) bounded density function K, which is assumed to be symmetric about zero, and bandwidths $h_j > 0$. The covariate V_{ji} is assumed to have a density f_j, and we denote by

$$\hat{f}_j(v) = (n_j h_j)^{-1} \sum_{i=1}^{n_j} K\left(\frac{v - V_{ji}}{h_j}\right) \tag{2.8}$$

the corresponding kernel estimator.

The idea is to re-scale the pointwise EL interval about $\theta_n(t) = S_{1n_1}(t|v)/S_{2n_2}(t|v)$, where $S_{jn_j}(t|v)$ is Beran's estimator of $S_j(t|v)$ based on the same Nadaraya–Watson weights used in the local EL (see LV for the precise definition of Beran's estimator). Let the endpoints τ_1 and τ_2 of the time interval for simultaneous coverage satisfy the same conditions as before, except in terms of the *conditional* distribution functions of X_{ji} and Y_{ji} given $V_{ji} = v$. Also assume that each conditional survival function $S_j(t|v)$ is continuous in $t \in [\tau_1, \tau_2]$.

The following theorem justifies our procedure for constructing a width-scaled band for $\theta(t)$ by verifying that we have a limit of the same form as (2.3) for a *normalized* version of the local EL statistic; this is done through multiplication by

$$Q_n(t) = \frac{\hat{f}_1(v)\hat{\sigma}_1^2(t|v) + \hat{f}_2(v)\hat{\sigma}_2^2(t|v)}{\int K^2(u)\,\mathrm{d}u\,[\hat{\sigma}_1^2(t|v)/(n_1h_1) + \hat{\sigma}_2^2(t|v)/(n_2h_2)]},$$

where $\hat{\sigma}_j^2(t|v)$ is a uniformly consistent estimator for the asymptotic variance $\sigma_j^2(t|v)$ of Beran's estimator $S_{jn_j}(t|v)$ over $t \in [\tau_1, \tau_2]$, see Li and Doss (1995).

Theorem 2.1. *Assume $n_j h_j^5 \to 0$, $n_j h_j \to \infty$ and $n_j h_j/(n_1h_1 + n_2h_2) \to p_j > 0$ as $n_j \to \infty$, and that the regularity conditions* (R1)–(R3) *of Li and Doss* (1995, *Section* 4) *hold for each sample. Let $\tilde{\theta}(t) = (\theta(t) - (1-\gamma)\theta_n(t))/\gamma$ for some $\gamma \geqslant 1$. Then*

$$-2Q_n(t)\log R(\tilde{\theta}(t), t) \xrightarrow{\mathrm{D}} U^2(t|v)/(\gamma^2\sigma^2(t|v))$$

in $D[\tau_1, \tau_2]$, where $U(t|v)$ is a continuous Gaussian martingale (*in t*) *with mean zero and variance function $\sigma^2(t|v) = \sigma_1^2(t|v)/p_1 + \sigma_2^2(t|v)/p_2$.*

Note that

$$\hat{\sigma}^2(t|v) = (n_1h_1 + n_2h_2)[\hat{\sigma}_1^2(t|v)/(n_1h_1) + \hat{\sigma}_2^2(t|v)/(n_2h_2)]$$

consistently estimates $\sigma^2(t|v)$. Thus, we can set the inflation factor $\hat{\gamma} = \sqrt{\hat{c}(\alpha)}/z_{\alpha/2}$ as before, except that $\hat{\sigma}^2(t)$ is now replaced by $\hat{\sigma}^2(t|v)$ in the definition of $\hat{c}(\alpha)$.

3. Example

In this section, we consider an example involving the comparison of a treatment group with a placebo group. The data come from a Mayo Clinic trial of a treatment for primary biliary cirrhosis of the liver (Fleming and Harrington, 1991). 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). For each patient, the date of randomization, the disease and survival status as of July 1986, and a large number of risk factors such as serum bilirubin and age were recorded. Fleming and Harrington's Cox model analysis of these data found no detectable difference between treatment and placebo. They also found that the covariate serum bilirubin is the strongest univariate predictor of survival.

First, we ignore all covariates. Confidence bands for the ratio $\theta(t) = S_{\text{placebo}}(t)/S_{\text{treatment}}(t)$ are plotted in Fig. 1. The width-scaled EL band (first panel) is seen to be narrower in the right tail than the standard (McKeague and Zhao, 2002) EL band (second panel). The data-driven "inflation" factor $\hat{\gamma} = 1.51$.

The first panel of Fig. 2 displays the 90% EL width-scaled confidence band for the ratio $\beta(t)$ of the cumulative hazard function for placebo over that for treatment. The corresponding ratio of the Nelson–Aalen curves is also displayed. The width-scaled band is narrower than the standard EL band over the first 1000 days. The data-driven "inflation" factor $\hat{\gamma} = 1.65$.

Next, we consider adjustment for the covariate effect of serum bilirubin, with its level set at 2 mg/dl. The kernel-weighted EL approach of Li and Van Keilegom (2002) is used with bandwidth

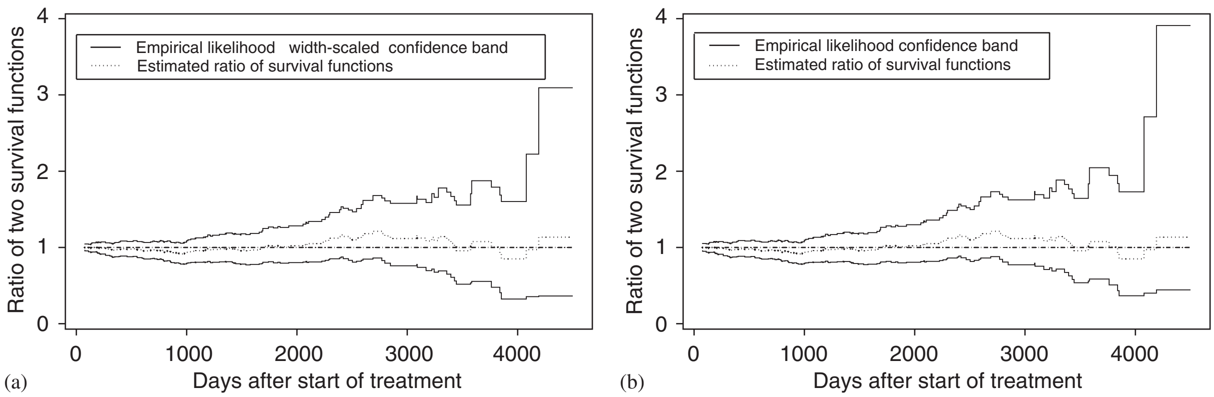

Fig. 1. Mayo Clinic trial, 95% confidence bands for the ratio of the survival functions (placebo/treatment): (a) width-scaled EL confidence band; (b) standard EL confidence band.

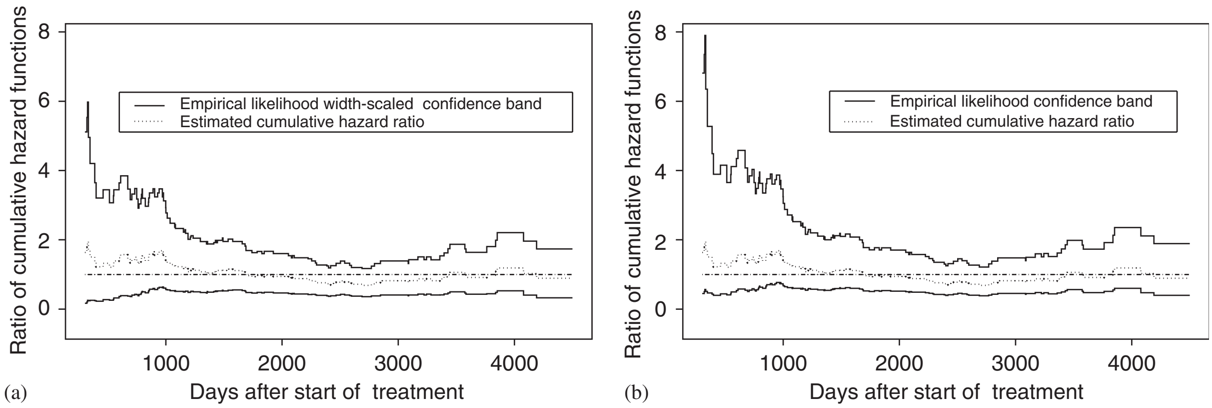

Fig. 2. Mayo Clinic trial, 90% confidence bands for the ratio $\beta(t)$ of the cumulative hazard functions: (a) width-scaled EL confidence band for $\beta(t)$; (b) standard EL confidence band for $\beta(t)$.

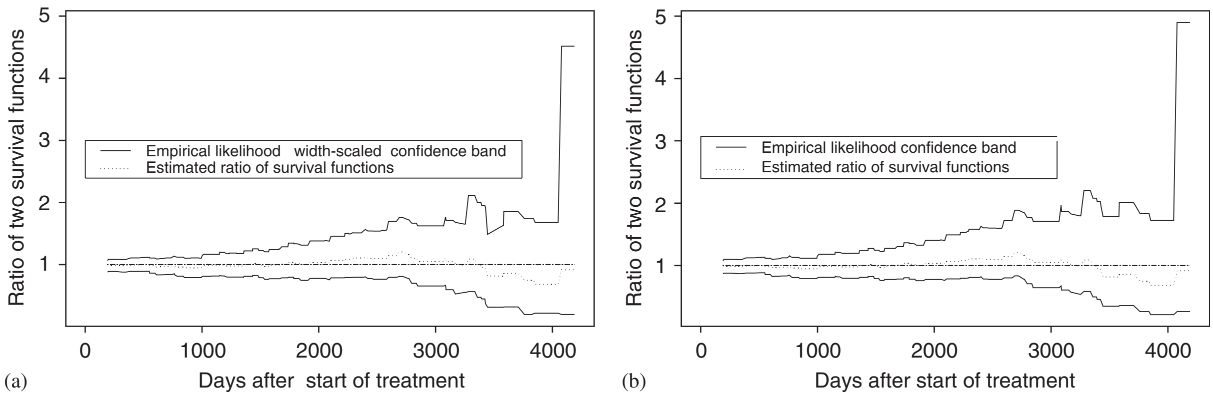

Fig. 3. Mayo Clinic trial, 95% confidence bands for the ratio of the conditional survival functions (placebo/treatment) at the 2 mg/dl serum bilirubin level: (a) width-scaled EL confidence band; (b) standard EL confidence band.

$h = 2$ and the biquadratic kernel function $K(x) = (\frac{15}{16})(1 - x^2)^2 I(|x| \leqslant 1)$. The first panel of Fig. 3 displays the 95% width-scaled EL confidence band for the ratio of the conditional survival function for placebo over that for treatment. The corresponding ratio of the Beran estimators is also displayed. Note that the variance blows up in the right tail. Comparing the two panels in Fig. 3, the width-scaled band is seen to be narrower than the standard EL band, especially in the right tail. The data-driven "inflation" factor $\hat{\gamma} = 1.46$.

The first panel of Fig. 4 displays the 90% width-scaled EL confidence band for the ratio of the conditional cumulative hazard function for placebo over that for treatment at the 2 mg/dl serum bilirubin level. The data-driven "inflation" factor $\hat{\gamma} = 1.60$. The width-scaled EL confidence band is narrower than the standard band (second panel) in the left tail.

In summary, the width-scaled bands we have examined provide greater stability in the tails compared with standard EL bands.

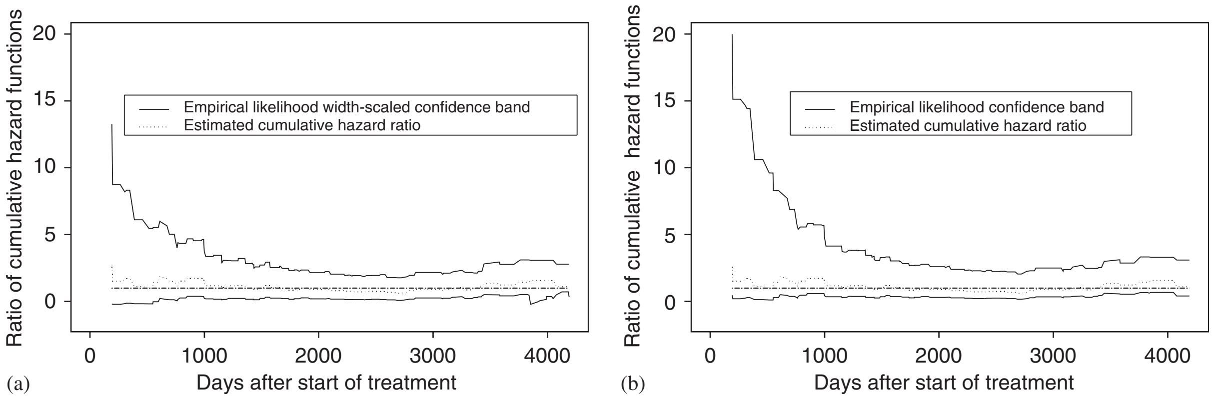

Fig. 4. Mayo Clinic trial, 90% confidence bands for the ratio of the conditional cumulative hazard functions (placebo/treatment) at 2 mg/dl serum bilirubin level: (a) width-scaled EL confidence band; (b) standard EL confidence band.

Acknowledgements

The authors are grateful for the helpful suggestions from a referee.

Appendix A

First, we show how condition (2.3) is checked in the one-sample case discussed in Section 2.

The ordered uncensored survival times, i.e., the X_i with corresponding $\delta_i = 1$, are written $0 \leqslant T_1 \leqslant \cdots \leqslant T_N < \infty$, and $r_i = \sum_{k=1}^{n} 1_{\{Z_k \geqslant T_i\}}$ denotes the size of the risk set at T_i-, $d_i = \sum_{k=1}^{n} 1_{\{Z_k = T_i, \delta_k = 1\}}$ denotes the number of "deaths" occurring at time T_i. Define $K(t) = \#\{i : T_i \leqslant t\}$ and $D = \max_{i:T_i \leqslant t}(d_i - r_i)$. It can be shown using Lagrange's method (cf. Thomas and Grunkemeier, 1975 or Li, 1995) that

$$-2\log R(\tilde{S}(t), t) = -2\sum_{i=1}^{K(t)}\left((r_i - d_i)\log\left(1 + \frac{\lambda_n(t)}{r_i - d_i}\right) - r_i \log\left(1 + \frac{\lambda_n(t)}{r_i}\right)\right), \tag{A.1}$$

where the Lagrange multiplier $D < \lambda_n(t) < \infty$ satisfies the equation

$$\prod_{i=1}^{K(t)}\left(1 - \frac{d_i}{r_i + \lambda_n(t)}\right) = \tilde{S}(t). \tag{A.2}$$

Setting $\tilde{S}(t) = (S(t) - (1-\gamma)S_n(t))/\gamma$ in what follows, the corresponding Lagrange multiplier satisfies

$$|\lambda_n(t)| \leqslant \max\left\{\frac{n(\log\tilde{S}(t) - \log S_n(t))}{A_n(t) + \log S_n(t) - \log\tilde{S}(t)}, \frac{n(-\log\tilde{S}(t) - A_n(t))}{-\log\tilde{S}(t)}\right\},$$

by Li (1995, (2.12) and (2.14)), where $A_n(t) = \sum_{i=1}^{K(t)} d_i/r_i$ is the Nelson–Aalen estimator of the cumulative hazard function of S. Noting that $\sqrt{n}(\log\tilde{S}(t) - \log S_n(t))$ has a martingale approximation, cf. (A.9), it can be shown that $\lambda_n(t) = O_P(n^{1/2})$ uniformly over $t \in [\tau_1, \tau_2]$. Then, again along the lines of Li (1995),

$$\lambda_n(t) = \frac{n(\log \tilde{S}(t) - \log S_n(t))}{\hat{\sigma}^2(t)} + O_P(1) \tag{A.3}$$

uniformly over $t \in [\tau_1, \tau_2]$, leading to

$$\begin{aligned} -2\log R(\tilde{S}(t), t) &= \frac{\hat{\sigma}^2(t)\lambda_n^2(t)}{n} + o_P(1) \\ &= \frac{n(\log \tilde{S}(t) - \log S_n(t))^2}{\hat{\sigma}^2(t)} + o_P(1) \end{aligned} \tag{A.4}$$

uniformly over $t \in [\tau_1, \tau_2]$.

Recall that $\sqrt{n}(S_n - S) \xrightarrow{D} -SU$ in $D[\tau_1, \tau_2]$, where U is the Gaussian martingale defined in Section 2, see Andersen et al. (1993, p. 262). Define the functional $h \mapsto \phi_1(h)$ from $D[\tau_1, \tau_2]$ into itself by

$$\phi_1(h) = \log((S - (1-\gamma)h)/\gamma) - \log h,$$

where S is the (true) survival function of X_i. Using the chain rule, ϕ_1 can be shown to be compactly differentiable with derivative at S given by

$$d\phi_1(S)h = -\frac{h}{\gamma S},$$

for $h \in D[\tau_1, \tau_2]$. Applying the functional delta method (Andersen et al., 1993, p. 111), we then obtain

$$\sqrt{n}(\phi_1(S_n) - \phi_1(S)) \xrightarrow{D} U/\gamma$$

in $D[\tau_1, \tau_2]$, which can be re-expressed as

$$\sqrt{n}(\log \tilde{S}(t) - \log S_n(t)) \xrightarrow{D} U(t)/\gamma. \tag{A.5}$$

We conclude from (A.9), (A.4), and the uniform consistency of $\hat{\sigma}^2(t)$, that

$$-2\log R(\tilde{S}(t), t) \xrightarrow{D} \frac{U^2(t)}{\gamma^2\sigma^2(t)}$$

in $D[\tau_1, \tau_2]$, as required. □

Proof of Theorem 2.1. The ordered uncensored survival times, i.e., the X_{ji} with corresponding $\delta_{ji} = 1$, are written $0 \leqslant T_{j1} \leqslant \cdots \leqslant T_{jN_j} < \infty$. Also, let $R_{ji} = \sum_{k=1}^{n_j} W_{jk} 1_{\{Z_{jk} \geqslant T_{ji}\}}$ and $d_{ji} = \sum_{k=1}^{n_j} W_{jk} 1_{\{Z_{jk} = T_{ji}, \delta_{jk}=1\}}$, where

$$W_{ji} = W_{ji}(v, h_j) = \frac{K(v - V_{ji}/h_j)}{\sum_{k=1}^{n_j} K(v - V_{jk}/h_j)}$$

are the Nadaraya–Watson weights. Define $K_j(t) = \#\{i : T_{ji} \leqslant t\}$ and $D_j = \max_{i:T_{ji}\leqslant t}(d_{ji} - R_{ji})$. As in McKeague and Zhao (2002) we have

$$-2\log R(\tilde{\theta}(t), t) = -2\sum_{i=1}^{K_1(t)}\left((R_{1i} - d_{1i})\log\left(1 + \frac{\lambda_n}{R_{1i} - d_{1i}}\right) - R_{1i}\log\left(1 + \frac{\lambda_n}{R_{1i}}\right)\right)$$
$$-2\sum_{i=1}^{K_2(t)}\left((R_{2i} - d_{2i})\log\left(1 - \frac{\lambda_n}{R_{2i} - d_{2i}}\right) - R_{2i}\log\left(1 - \frac{\lambda_n}{R_{2i}}\right)\right),$$

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

$$\log\prod_{i=1}^{K_1(t)}\left(1 - \frac{d_{1i}}{R_{1i} + \lambda_n}\right) - \log\prod_{i=1}^{K_2(t)}\left(1 - \frac{d_{2i}}{R_{2i} - \lambda_n}\right) = \log(\tilde{\theta}(t)). \tag{A.6}$$

Eq. (A.6) has a unique solution λ_n provided $D_j < 0$, $j = 1, 2$.

First, assume $\lambda_n(t) < 0$. Then by LV,

$$-\log\prod_{i=1}^{K_1(t)}\left(1 - \frac{d_{1i}}{R_{1i} + \lambda_n(t)}\right) \geqslant \hat{A}_1(t|v)\left(\frac{1}{1 - |\lambda_n(t)|}\right),$$
$$\log\prod_{i=1}^{K_2(t)}\left(1 - \frac{d_{2i}}{R_{2i} - \lambda_n(t)}\right) \geqslant -\hat{A}_2(t|v)\left(\frac{1}{1 + |\lambda_n(t)|}\right) + \log S_{2n_2}(t|v) + \hat{A}_2(t|v),$$

where $\hat{A}_j(t|v) = \sum_{i=1}^{K_j(t)} d_{ji}/R_{ji}$ is Beran's estimator of the cumulative hazard function $A_j(t|v)$. Combining the above two inequalities and (A.6), we get

$$-\log\tilde{\theta}(t) \geqslant \hat{A}_1(t|v)\left(\frac{1}{1 - |\lambda_n|}\right) - \hat{A}_2(t|v)\left(\frac{1}{1 + |\lambda_n|}\right) + \log S_{2n_2}(t|v) + \hat{A}_2(t|v).$$

Then, using $1/(1 - x) > 1 + x$ for $0 \leqslant x < 1$,

$$-\log\tilde{\theta}(t) - \hat{A}_1(t|v) - \log S_{2n_2}(t|v) \geqslant \hat{A}_1(t|v)|\lambda_n|. \tag{A.7}$$

If $\lambda_n(t) \geqslant 0$, a similar argument leads to

$$\log\tilde{\theta}(t) - \log S_{1n_1}(t|v) - \hat{A}_2(t|v) \geqslant \hat{A}_2(t|v)|\lambda_n|. \tag{A.8}$$

Next, since

$$\sqrt{n_j h_j}(S_{jn_j} - S_j) \xrightarrow{\mathrm{D}} -S_j U_j,$$

where $U_j(t|v)$ is a Gaussian martingale with mean zero and $\mathrm{var}(U_j(s|v)) = \sigma_j^2(s|v)$ (cf. Li and Doss, 1995, p. 796), and the two samples are independent, by $n_j h_j/(n_1h_1 + n_2h_2) \to p_j > 0$, it follows that

$$\sqrt{n_1h_1 + n_2h_2}\left(\begin{pmatrix} S_{1n_1} \\ S_{2n_2} \end{pmatrix} - \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}\right) \xrightarrow{\mathrm{D}} U^*,$$

where

$$U^* = \begin{pmatrix} \frac{U_1 S_1}{\sqrt{p_1}} \\ \frac{U_2 S_2}{\sqrt{p_2}} \end{pmatrix}.$$

Define the functional $h \mapsto \phi_2(h)$ from $D[\tau_1, \tau_2] \times D[\tau_1, \tau_2]$ into $D[\tau_1, \tau_2]$ by $\phi_2(S_1, S_2)(t) = S_1(t)/S_2(t)$. By the functional delta method, as before, we have

$$\sqrt{n}(\phi_2(S_{1n_1}, S_{2n_2}) - \phi_2(S_1, S_2)) \xrightarrow{\mathrm{D}} d\phi_2(S_1, S_2)U^*,$$

where

$$d\phi_2(S_1, S_2)U^* = \left(\frac{U_1}{\sqrt{p_1}} - \frac{U_2}{\sqrt{p_2}}\right)\frac{S_1}{S_2} = U(\cdot|v)\theta.$$

Also,

$$\sqrt{n_1 h_1 + n_2 h_2}(\phi_1(\theta_n) - \phi_1(\theta)) \xrightarrow{\mathrm{D}} \frac{U(\cdot|v)}{\gamma}$$

in $D[\tau_1, \tau_2]$, which can be re-expressed as

$$\sqrt{n_1 h_1 + n_2 h_2}(\log \tilde{\theta}(t) - \log \theta_n(t)) \xrightarrow{\mathrm{D}} \frac{U(t|v)}{\gamma}. \tag{A.9}$$

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

$$\begin{aligned}
\sqrt{n_j h_j}\left|\sum_{i=1}^{K_j(t)} \log\left(1 - \frac{d_{ji}}{R_{ji}}\right) + \sum_{i=1}^{K_j(t)} \frac{d_{ji}}{R_{ji}}\right| &\leqslant \left(\max_{i \leqslant K_j(t)} \frac{\sqrt{n_j h_j} d_{ji}}{R_{ji}}\right) \sum_{i=1}^{K_j(t)} \frac{d_{ji}}{R_{ji}} \\
&= \left(\max_{i \leqslant K_j(t)} \frac{\sqrt{n_j h_j} d_{ji}}{R_{ji}}\right) \hat{A}_j(t|v) \\
&\leqslant \left(\max_{i \leqslant K_j(t)} \frac{1}{R_{ji}}\right) \frac{M \hat{A}_j(t|v)}{\sqrt{n_j h_j} \hat{f}_j(v)} \\
&= \frac{1}{\min_{i \leqslant K_j(t)} R_{ji}} \frac{M \hat{A}_j(t|v)}{\sqrt{n_j h_j} \hat{f}_j(v)},
\end{aligned} \tag{A.10}$$

where in the third line we use (2.8) and $d_{ji} = W_{ji}$ a.s., since a.s. there are no ties among the $X_{jn_1}, \ldots, X_{jn_j}$ due to the continuity of the survival function of X_{jn_1}. Here, M is an upper bound on the kernel K.

From Beran (1981) (see also Dabrowska, 1987, p. 188),

$$\hat{H}_j(t|v) = \sum_{i=1}^{n_j} W_{ji}(v, h_j) 1_{\{Z_{ji} \geqslant t\}}$$

converges uniformly in $t \in [\tau_1, \tau_2]$ to $H_j(t|v) = P(Z_{j1} \geqslant t|v)$ a.s. for almost every v. Thus, for any $\varepsilon > 0$ and n sufficiently large, $\inf_{t\in[\tau_1,\tau_2]} \hat{H}_j(t|v) > H_j(\tau_2|v)/2 > 0$ with probability at least $1-\varepsilon$. Recalling that $R_{ji} = \sum_{k=1}^{n_j} W_{jk} 1_{\{Z_{jk} \geqslant T_{ji}\}}$, for n sufficiently large $\min_{i \leqslant K_j(t)} R_{ji} \geqslant \inf_{t\in[\tau_1,\tau_2]} \hat{H}_j(t|v) > H_j(\tau_2|v)/2$ for all $t \in [\tau_1, \tau_2]$ with probability at least $1-\varepsilon$. Thus, combining the uniform consistency (in t) of $\hat{A}_j(t|v)$ (cf. Proposition 2.2 of Dabrowska (1987)), and the weak consistency of $\hat{f}_j(v)$ (cf. Theorem 3.1.2 of Prakasa Rao (1983)), we conclude that (A.10) converges uniformly (over $t \in [\tau_1, \tau_2]$) in probability to zero.

Combining (A.9) and (A.10), we obtain

$$\sqrt{n_1h_1 + n_2h_2}(\log \tilde{\theta}(t) + \hat{A}_1(t|v) + \log S_{2n_2}(t|v)) \xrightarrow{\mathrm{D}} -\frac{U(t|v)}{\gamma}, \tag{A.11}$$

$$\sqrt{n_1h_1 + n_2h_2}(\log \tilde{\theta}(t) - \log S_{1n_1}(t|v) - \hat{A}_2(t|v)) \xrightarrow{\mathrm{D}} -\frac{U(t|v)}{\gamma}. \tag{A.12}$$

Thus, the l.h.s. of (A.7) and (A.8) are $O_P((n_1h_{1n_1} + n_2h_{2n_2})^{-1/2})$ uniformly in $t \in [\tau_1, \tau_2]$ from (A.11) and (A.12). Combining (A.7), (A.8), and the uniform consistency (in t) of $\hat{A}_j(t|v)$, we have $\lambda_n = O_P(1/(n_1h_{1n_1} + n_2h_{2n_2})^{1/2})$ uniformly for $t \in [\tau_1, \tau_2]$.

Along the same lines as LV,

$$\log \prod_{i=1}^{K_1(t)} \left(1 - \frac{d_{1i}}{R_{1i} + \lambda_n(t)}\right) = \log S_{1n_1}(t|v) + \frac{\lambda_n \hat{f}_1(v)\hat{\sigma}_1^2(t|v)}{\int K^2(u)\,\mathrm{d}u} + O_P((n_1h_1)^{-1}),$$

$$\log \prod_{i=1}^{K_2(t)} \left(1 - \frac{d_{2i}}{R_{2i} - \lambda_n(t)}\right) = \log S_{2n_2}(t|v) - \frac{\lambda_n \hat{f}_2(v)\hat{\sigma}_2^2(t|v)}{\int K^2(u)\,\mathrm{d}u} + O_P((n_2h_2)^{-1}).$$

Combining this with (A.6),

$$\lambda_n = \frac{\int K^2\,\mathrm{d}u(\log \tilde{\theta}(t) - \log \theta_n(t))}{\hat{f}_1(v)\hat{\sigma}_1^2(t|v) + \hat{f}_2(v)\hat{\sigma}_2^2(t|v)} + O_P((n_1h_1)^{-1} + (n_2h_2)^{-1}).$$

A Taylor expansion then gives

$$\begin{aligned} -2Q_n(t)\log R(\tilde{\theta}(t), t) &= Q_n(t)\lambda_n^2 \left(\frac{\hat{f}_1(v)\hat{\sigma}_1^2(t|v_1)}{\int K^2(u)\,\mathrm{d}u} + \frac{\hat{f}_2(v)\hat{\sigma}_2^2(t|v)}{\int K^2(u)\,\mathrm{d}u}\right) + O_P((n_1h_1 + n_2h_2)^{-1}) \\ &= \frac{(n_1h_1 + n_2h_2)(\log \tilde{\theta}(t) - \log \theta_n(t))^2}{\hat{\sigma}^2(t|v)} + O_P((n_1h_1 + n_2h_2)^{-1}), \end{aligned} \tag{A.13}$$

and we conclude from (A.9), (A.13), and the uniform consistency of $\hat{\sigma}^2(t|v)$, that

$$-2Q_n(t)\log R(\tilde{\theta}(t), t) \xrightarrow{\mathrm{D}} \left(\frac{U(t|v)}{\gamma\sigma(t|v)}\right)^2$$

in $D[\tau_1, \tau_2]$. □

References

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

Beran, R., 1981. Nonparametric regression with randomly censored survival data. Technical Report, Department of Statistics, University of California, Berkeley.

Dabrowska, D.M., 1987. Nonparametric regression with censored survival time data. Scand. J. Statist. 14, 181–197.

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

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

Hall, W.J., Wellner, J.A., 1980. Confidence bands for a survival curve from censored data. Biometrika 67, 133–143.

Hollander, M., McKeague, I.W., 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.

Li, G., Doss, H., 1995. An approach to nonparametric regression for life history data using local linear fitting. Ann. Statist. 23, 787–823.

Li, G., Van Keilegom, I., 2002. Likelihood ratio confidence bands in non-parametric regression with censored data. Scand. J. Statist. 29, 547–562.

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

Nair, V.N., 1984. Confidence bands for survival functions with censored data: a comparative study. Technometrics 26, 265–275.

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.

Prakasa Rao, B.L., 1983. Nonparametric Functional Estimation. Academic Press, New York.

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