Simultaneous confidence bands for ratios of survival functions via empirical likelihood

Ian W. McKeague\textsuperscript{a,1}, Yichuan Zhao\textsuperscript{b,*,2}

\textsuperscript{a}Department of Statistics, Florida State University, Tallahassee, FL 32306, USA
\textsuperscript{b}Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, USA

Received March 2001; received in revised form July 2002

Abstract

We derive a simultaneous confidence band for the ratio of two survival functions based on independent right-censored data. Earlier authors have studied such bands for the difference of two survival functions, but the ratio provides a more appropriate comparison in some applications, e.g., in comparing two treatments in biomedical settings. Our approach is formulated in terms of empirical likelihood and allows us to avoid the use of simulation techniques that are often needed for Wald-type confidence bands. By the transformation-preserving property we also obtain confidence bands for the difference in the cumulative hazard functions. The approach is illustrated with a real data example.

\textcopyright 2002 Elsevier Science B.V. All rights reserved.

\textbf{Keywords:} Distribution-free; Kaplan–Meier estimator; Nelson–Aalen estimator; Right censoring

1. Introduction

In biomedical settings, it is frequently of interest to compare two survival functions $S_j(t)$, $j = 1, 2$, in terms of their Kaplan–Meier estimators. The purpose of this note is to derive a computationally simple simultaneous confidence band for the ratio $S_1(t)/S_2(t)$. This ratio furnishes a useful and meaningful comparison between two treatments in terms of relative survival probabilities, and is appropriate when the risks of treatment failure are moderate (otherwise the difference might be more appropriate). Our approach is based on the empirical likelihood method.
The use of empirical likelihood in survival analysis goes back to Thomas and Grunkemeier (1975) who derived pointwise confidence intervals for $S_j(t)$ (see also Li, 1995; Murphy, 1995). The method was extended by Owen (1988, 1990) to a variety of statistical problems. Hollander et al. (1997) used the method to obtain simultaneous confidence bands for $S_j(t)$. In the two-sample context, Einmahl and McKeague (1999) found confidence bands for Q-Q plots. For censored data, confidence intervals for differences of treatment means and quantiles were derived by Jing (1995), Qin (1994, 1997) and Qin and Zhao (1997, 2000).

Numerous Wald-type confidence procedures have been proposed for the comparison of two survival functions. For example, Dabrowska et al. (1989) introduced a relative change function defined in terms of cumulative hazards and found simultaneous bands for this function under the assumption of proportional hazards. Parzen et al. (1997) constructed simultaneous confidence bands for the difference $S_j(t) - S_j(t)$ using a simulation technique. A simple nonparametric confidence interval procedure for the difference or ratio of two median failure times was proposed by Su and Wei (1993), Lin and Ying (1993) extended their results to the case of dependent data.

In the present article we focus on the ratio, and find two tractable likelihood-ratio based bands, one of which does not need simulation. The proposed confidence bands and various asymptotic results are presented in Section 2. In Section 3 we give an illustrative example. 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 $(Z_{ji}, \delta_{ji})$, where $j = 1, 2$ indexes the sample, $i = 1, \ldots, n_j$ indexes the observations within each sample, and $Z_{ji} = X_{ji} \cup Y_{ji}, \delta_{ji} = 1\{Z_{ji} < T_{ji}\}$. The distribution functions of $X_{ji}$ and $Y_{ji}$ are denoted $F_j$ and $G_j$, respectively. The survival functions $S_j = 1 - F_j$ to be compared 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 empirical likelihood is given by

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

where $\hat{S}_j$ belongs to $\Gamma$, the space of all survival functions on $[0, \infty)$. The empirical likelihood ratio for $S_1(t)/S_2(t)$ at $\hat{\theta}(t) > 0$ for a given $t \geq 0$ is defined by

$$R(\hat{\theta}(t), t) = \frac{\sup\{L(\hat{S}_1, \hat{S}_2); \hat{S}_1(t)/\hat{S}_2(t) = \hat{\theta}(t), (\hat{S}_1, \hat{S}_2) \in \Gamma \times \Gamma\}}{\sup\{L(\hat{S}_1, \hat{S}_2); (\hat{S}_1, \hat{S}_2) \in \Gamma \times \Gamma\}}.$$

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 $r_{ji} = \sum_{k=1}^{n_j} 1\{Z_{jk} < T_{ji}\}$ denotes the size of the risk set at $T_{ji} - d_{ji} = \sum_{k=1}^{n_j} 1\{Z_{jk} - T_{ji} < d_{ji} - 1\}$ denotes the number of “deaths” occurring at time $T_{ji}$. Define $K_j(t) = \#\{i: T_{ji} \leq t\}$ and $D_j = \max_i r_{ji} \leq (d_{ji} - r_{ji})$. It can be shown using Lagrange's method (cf. Thomas and
Grunkemeier, 1975 or Li, 1995) that
\[ -2 \log R(\hat{\theta}(t), t) = -2 \sum_{i=1}^{K_i(t)} \left( (r_{1i} - d_{1i}) \log \left( 1 + \frac{\hat{\lambda}_n}{r_{1i} - d_{1i}} \right) - r_{1i} \log \left( 1 + \frac{\hat{\lambda}_n}{r_{1i}} \right) \right) \\
-2 \sum_{i=1}^{K_i(t)} \left( (r_{2i} - d_{2i}) \log \left( 1 - \frac{\hat{\lambda}_n}{r_{2i} - d_{2i}} \right) - r_{2i} \log \left( 1 - \frac{\hat{\lambda}_n}{r_{2i}} \right) \right), \] (2.3)
where the Lagrange multiplier \( D_1 < \lambda_n < -D_2 \) satisfies the equation
\[ \log \prod_{i=1}^{K_i(t)} \left( 1 - \frac{d_{1i}}{r_{1i} + \lambda_n} \right) - \log \prod_{i=1}^{K_i(t)} \left( 1 - \frac{d_{2i}}{r_{2i} - \lambda_n} \right) = \log(\hat{\theta}(t)). \] (2.4)

Eq. (2.4) has a unique solution \( \lambda_n \) provided \( D_1 < 0, \ j = 1, 2 \), because as a function of \( \lambda_n \) the l.h.s. of (2.4) is continuous, strictly increasing and tends to \( \pm \infty \) as \( \lambda_n \uparrow -D_2 \) or \( \lambda_n \downarrow D_1 \). A similar (but more restrictive) Lagrange multiplier equation appears in (2.3) of Einmahl and McKeague (1999).

Let \( H_j(s) = S_j(s)(1 - G_j(s)) \). We assume throughout that \( n_j/n \to p_j > 0 \) as \( n \to \infty \). Let \( \tau_1 \) 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
\[ \sigma_j^2(t) = \int_0^t \frac{dF_j(s)}{S_j(s)H_j(s)}, \quad t \in [\tau_1, \tau_2] \] (2.5)
and \( \sigma^2(t) = \sigma_1^2(t)/p_1 + \sigma_2^2(t)/p_2 \). It is easy to show that
\[ \hat{\sigma}_j^2(t) = n_j \sum_{i: \tau_j \leq t} \frac{d_{ji}}{r_{ji}(r_{ji} - d_{ji})}, \] (2.6)
is a uniformly consistent estimator of \( \sigma_j^2(t), t \in [\tau_1, \tau_2] \), see Andersen et al. (1993, IV.1.3). Thus
\[ \hat{\sigma}^2(t) = n \hat{\sigma}_1^2(t)/n_1 + \hat{\sigma}_2^2(t)/n_2 \] is a uniformly consistent estimator of \( \sigma^2(t), t \in [\tau_1, \tau_2] \). In the uncensored case, note that \( \sigma_j^2(t) = F_j(t)/(1 - F_j(t)) \).

2.2. Confidence bands

Now we state our main result and explain how it can be used to construct simultaneous confidence bands for \( \theta(t) = S_1(t)/S_2(t) \) over the time span of interest \([\tau_1, \tau_2] \).

**Theorem 2.1.** Under the above conditions, \(-2\hat{\sigma}^2(t) \log R(\hat{\theta}(t), t)\) converges in distribution to \( U^2(t) \) in \( D[\tau_1, \tau_2] \), where \( U(t) \) is a Gaussian martingale with mean zero and variance function \( \sigma^2(t) \).

Theorem 2.1 can be used to obtain the following two types of confidence bands for \( \theta(t) \). First note that
\[ -2 \log R(\hat{\theta}(t), t) \overset{D}{\to} \left( \frac{B(\sigma^2(t))}{\hat{\sigma}(t)} \right)^2, \]
in \( D[\tau_1, \tau_2] \), where \( B \) is a standard Wiener process. Thus
\[ \mathcal{B}_1 = \{(t, \hat{\theta}(t)) : -2 \log R(\hat{\theta}(t), t) \leq c_s[\hat{e}_1, \hat{e}_2], t \in [\tau_1, \tau_2]\}, \]
(2.7)
is an asymptotic 100(1 − z)% confidence band for \( \theta(t) \), where \( \hat{c}_i = \hat{\sigma}^2(t_i) \) and \( c_s[e_1, e_2] \) is the upper \( z \)-quantile of the distribution of \( \sup_{t \in [e_1, e_2]} |B(t)|^2 \). If the critical value above is replaced by a \( \chi^2 \)

critical value, then we obtain a pointwise band. Thus \( \mathcal{B}_1 \) has the equal precision property (cf. Nair, 1984). Simulation needs to be used to obtain \( c_s[e_1, e_2] \), because, as far as we are aware, tables for \( c_s[e_1, e_2] \) are not available.

A confidence band that does not need simulation can be obtained as follows. From Theorem 2.1,

\[
\sup_{t \in [\tau_1, \tau_2]} \frac{-2\hat{\sigma}^2(t) \log R(\theta(t), t)}{(1 + \hat{\sigma}^2(t))^2} \overset{\mathcal{D}}{\to} \sup_{x \in [d_1, d_2]} |B^0(x)|^2,
\]

by the continuous mapping theorem, where \( B^0 \) is a standard Brownian bridge process and

\[
d_i = \frac{\hat{\sigma}^2(t_i)}{1 + \hat{\sigma}^2(t_i)}, \quad i = 1, 2.
\]

Here we have used the fact that the processes \( B^0(\sigma^2(\cdot)/(1 + \sigma^2(\cdot))) \) and \( U(\cdot)/(1 + \sigma^2(\cdot)) \) have the same distribution (cf. Hall and Wellner, 1980). Since \( \hat{\sigma}^2(t) \) is a uniformly consistent estimator of \( \sigma^2(t) \), \( t \in [\tau_1, \tau_2] \), we know that \( d_i \) is consistently estimated by \( \hat{d}_i = \hat{\sigma}^2(t_i)/(1 + \hat{\sigma}^2(t_i)) \). This result yields the following asymptotic 100(1 − z)% confidence band for \( \theta(t) \):

\[
\mathcal{B}_2 = \{(t, \theta(t)) : -2 \log R(\theta(t), t) \leq C^2(t), t \in [\tau_1, \tau_2]\},
\]

where

\[
C(t) = K_s[\hat{d}_1, \hat{d}_2](1 + \hat{\sigma}^2(t))/\hat{\sigma}(t)
\]

and \( K_s[d_1, d_2] \) is the upper \( z \)-quantile of the distribution of \( \sup_{t \in [d_1, d_2]} |B^0(t)| \). Chung (1986) gave tables of \( K_s[d_1, d_2] \) for general \( d_1 \) and \( d_2 \); see also the computer program WIENER PACK by Chung (1987).

**Implementation:** We now explain how to compute the confidence bands \( \mathcal{B}_1 \) and \( \mathcal{B}_2 \). For fixed \( t \), let \( \psi(\lambda_n) \) denote the r.h.s. of (2.3). The function \( \psi \) satisfies the following properties:

1. \( \psi(\cdot) \) is strictly decreasing on the interval \( (D_1, 0] \) and increasing on \([0, -D_2)\), because its derivative

   \[
   \psi'(\lambda_n) = \sum_{i=1}^{K_s(t)} \frac{2\hat{\lambda}_n d_{1i}}{(r_{1i} - d_{1i} + \hat{\lambda}_n)(r_{1i} + \hat{\lambda}_n)} + \sum_{i=1}^{K_s(t)} \frac{2\hat{\lambda}_n d_{2i}}{(r_{2i} - d_{2i} - \hat{\lambda}_n)(r_{2i} - \hat{\lambda}_n)}
   \]

   is negative for \( \lambda_n \in (D_1, 0) \), zero for \( \lambda_n = 0 \) and positive for \( \lambda_n \in (0, -D_2) \).

2. \( \psi(\lambda_n) \to +\infty \) as \( \lambda_n \to D_1 \), or \( \lambda_n \to -D_2 \), and \( \psi(0) = 0 \).

These properties imply that there exist exactly two roots \( \lambda_L < 0 < \lambda_U \) for \( \psi(\lambda_L) = \psi(\lambda_U) = C^2(t) \) and that \( \{\lambda_n: \psi(\lambda_n) \leq C^2(t)\} = [\lambda_L, \lambda_U] \). Because \( \hat{\theta}(t) \) satisfying (2.4) is an increasing function of \( \lambda_n \), the confidence set for \( \theta(t) \) is a closed interval \( [\hat{\theta}_L, \hat{\theta}_U] \), where \( \hat{\theta}_L = \prod_{i=1}^{K_s(t)} (1 - d_{1i}/(r_{1i} + \hat{\lambda}_L))/\prod_{i=1}^{K_s(t)} (1 - d_{2i}/(r_{2i} - \hat{\lambda}_L)) \) and \( \hat{\theta}_U \) is same as \( \hat{\theta}_L \), but with \( \hat{\lambda}_L \) replaced by \( \hat{\lambda}_U \). The roots \( \hat{\lambda}_L \) and \( \hat{\lambda}_U \) are readily computed using the van Wijngaarden–Decker–Brent algorithm (see Press et al., 1992, p. 359).
2.3. Discussion

Our approach can be generalized beyond the classical two-sample setting to a counting process framework in which the independence between the two samples can be relaxed. Instead of independence between \( X_{1i} \) and \( X_{2i} \), we only require \( P(X_{1i} = X_{2i}) = 0 \), which in turn implies that the counting processes corresponding to uncensored failures in the two samples have no simultaneous jumps. Such an extension includes the case of left-truncated data (see, e.g., Andersen et al., 1993, III.3–4).

The accuracy of the proposed confidence band for \( S_1(t)/S_2(t) \) may be improved in small samples by using a higher order approximation to the empirical likelihood ratio. That is, retain the second term in (A.20) along with the first to form a bias-corrected confidence band; Hollander et al. (1997) have studied such correction to confidence bands for \( S_j(t) \). Alternatively, it may be possible to use transformations to improve the accuracy of the proposed confidence band (cf. Andersen et al., 1993, IV.3.3). However, the best choice of transformation needs investigation. A similar issue arises in connection with the confidence band proposed by Parzen et al. (1997) for the difference \( S_1(t) - S_2(t) \).

Our approach can be easily adapted to obtain confidence bands for the ratio \( A_1(t)/A_2(t) \) of the cumulative hazard functions \( A_j(t) = -\log S_j(t) \), see Zhao (2002). A closely related problem to finding confidence bands for \( S_1(t)/S_2(t) \) is to construct confidence bands for the difference in the cumulative hazard functions \( A_1(t) - A_2(t) = -\log(S_1(t)/S_2(t)) \). Our approach immediately yields such confidence bands by the transformation-preserving property of the empirical likelihood method.

In some applications, it may be of greater interest to consider confidence bands for the ratio of cdf’s \( F_1(t)/F_2(t) \) rather than \( S_1(t)/S_2(t) \). Our approach does not readily extend to that case, because then \( \eta_j := \prod_{j=1}^{k_1}(1 - d_{ij}/(r_{ij} + (-1)^{-1} \lambda_n)) \) is replaced by \( 1 - \eta_j \) in (2.4), \( j = 1, 2 \), and the Taylor expansion w.r.t. \( \lambda_n \) becomes intractable.

Next consider left-censored data. That is, two independent samples of i.i.d. observations of the form \((Z_{ji}, \delta_{ji})\), where \( j = 1, 2 \) indexes the sample, \( i = 1, \ldots, n_j \) indexes the observations within each sample, and \( Z_{ji} = X_{ji} \vee Y_{ji}, \delta_{ji} = 1(X_{ji} \geq Y_{ji}) \). Following the same development as above, except changing the sign of \( Z_{ji} \) (time-reversal), we obtain a confidence band for \( F_1(t)/F_2(t) \).

For the ratio of survival functions based on \( k (\geq 3) \) independent samples it would be interesting to develop simultaneous confidence bands for all \( k - 1 \) comparisons. We have been unable to extend our present approach in that direction. However, our approach does extend when dealing with ratios of cumulative hazard functions.

3. An illustrative 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). Fig. 1 displays the proposed confidence bands for the ratio of the survival function for placebo over that for treatment. The corresponding ratio of the Kaplan–Meier curves is also displayed. Note that both simultaneous bands contain the horizontal line (ratio = 1), so there is no evidence of a difference between treatment and placebo on the basis of this analysis. The confidence band \( \mathcal{B}_1 \) is much narrower than \( \mathcal{B}_2 \) in
Fig. 1. Mayo Clinic trial, 95% simultaneous confidence bands for the ratio of survival functions (placebo/treatment), \( B_1 \) (left), \( B_2 \) (right).

the right tail, but wider over most of the follow-up period. We consider \( B_1 \) to be the more suitable band in view of its equal precision property.

Appendix A. Proof of Theorem 2.1

We need the following lemma describing the order of the Lagrange multiplier \( \lambda_n \).

**Lemma A.1.** Under the assumptions of Theorem 2.1, \( \lambda_n = \hat{\lambda}_n(t) = O_p(n^{1/2}) \) uniformly over \([\tau_1, \tau_2]\).

**Proof.** Let \( \hat{A}_j(t) \) denote the Nelson–Aalen estimator of \( A_j(t) \). Then \( (\sqrt{n}(\hat{A}_j(t) - A_j(t)), j = 1, 2) \) \( \overset{p}{\rightarrow} (U_j(t), j = 1, 2) \), where the \( U_j(t) \) are independent Gaussian martingales with mean zero and \( \text{cov}(U_j(s_1), U_j(s_2)) = \sigma_j^2(s_1 \wedge s_2) \) (cf. Andersen et al., 1993, pp. 193–194). By \( n_j/n \to p_j > 0 \) it follows that

\[
\sqrt{n}[(\hat{A}_1(t) - A_1(t)) - (\hat{A}_2(t) - A_2(t))] \overset{p}{\rightarrow} \frac{U_1(t)}{\sqrt{p_1}} + \frac{U_2(t)}{\sqrt{p_2}}.
\]

From \( S_1(t)/S_2(t) = \theta(t) \) we have \( A_1(t) - A_2(t) = -\log(\theta(t)) \), so

\[
\hat{A}_1(t) - \hat{A}_2(t) + \log(\theta(t)) = O_p(n^{-1/2}) \quad \text{uniformly for } t \in [\tau_1, \tau_2]. \tag{A.1}
\]

First assume \( \lambda_n(t) < 0 \). Then by Li (1995, pp. 101–102)

\[
-\log \prod_{i=1}^{K_i(t)} \left( 1 - \frac{d_{1i}}{r_{1i} + \hat{\lambda}_n(t)} \right) \geq \hat{A}_1(t) \left( \frac{n_1}{n_1 - |\hat{\lambda}_n(t)|} \right),
\]
\[ \log \prod_{i=1}^{\kappa_2(t)} \left(1 - \frac{d_{2i}}{r_{2i} - \hat{\lambda}_2(t)}\right) \geq -\hat{A}_2(t) \left(\frac{n_2}{n_2 + |\hat{\lambda}_2(t)|}\right) + \log S_{2n_2(t)} + \hat{A}_2(t), \]

where \( S_{2n_2(t)} \) is the Kaplan–Meier estimator of \( S(t) \). Combining the above two inequalities and (2.4) we get

\[ -\log \theta(t) \geq \hat{A}_1(t) \left(\frac{n_1}{n_1 - |\hat{\lambda}_1(t)|}\right) - \hat{A}_2(t) \left(\frac{n_2}{n_2 + |\hat{\lambda}_2(t)|}\right) + \log S_{2n_2(t)} + \hat{A}_2(t). \quad (A.2) \]

If \( \hat{\lambda}_n(t) \geq 0 \), a similar argument leads to

\[ \log \theta(t) - \log S_{1n}(t) - \hat{A}_1(t) \geq -\hat{A}_1(t) \left(\frac{n_1}{n_1 + |\hat{\lambda}_1(t)|}\right) + \hat{A}_2(t) \left(\frac{n_2}{n_2 - |\hat{\lambda}_2(t)|}\right). \quad (A.3) \]

Form the partition \([\tau_1, \tau_2] = \mathcal{T}_\omega \cup \mathcal{T}_\omega \cup \mathcal{T}_\omega\), where \( \mathcal{T}_\omega = \{t \in [\tau_1, \tau_2]: S_1(t) = S_2(t)\} \), \( \mathcal{T}_\omega = \{t \in [\tau_1, \tau_2]: S_1(t) < S_2(t)\} \), \( \mathcal{T}_\omega = \{t \in [\tau_1, \tau_2]: S_1(t) > S_2(t)\} \).

Case 1. \( t \in \mathcal{T}_\omega \). First suppose \( \hat{\lambda}_n(t) < 0 \). Then by (A.2)

\[ |\hat{\lambda}_n|(n_1 \hat{A}_1(t) + n_2 \hat{A}_2(t) + (n_1 - n_2) (\log S_{2n_2(t)} + \hat{A}_2(t))) \]

\[ \leq -n_1n_2(\hat{A}_1(t) + \log S_{2n_2(t)}). \quad (A.4) \]

By \( \sqrt{n}(|\hat{A}_j(t) - A_j(t)|)^{1/2} U_j(t) \), for any \( \epsilon > 0 \) and \( n \) sufficiently large, we have \( \hat{A}_j(t) \geq \frac{1}{2} A_j(t) \geq -\frac{1}{2} \log S_j(t) \) for all \( t \in [\tau_1, \tau_2] \) with probability at least \( 1 - \epsilon \). It then follows by (A.4) that for any \( \epsilon > 0 \),

\[ 0 \leq \frac{|\hat{\lambda}_n| - n_1 \log S_1(t)}{4} \leq n_1n_2(\hat{A}_2(t) - \hat{A}_1(t)) \quad (A.5) \]

for all \( t \in \mathcal{T}_\omega \) with probability \( 1 - \epsilon \), for \( n \) sufficiently large. The same argument works for \( \hat{\lambda}_n(t) \geq 0 \) but with \( \hat{A}_2(t) \) and \( \hat{A}_1(t) \) switching places in (A.5), and using (A.3) instead of (A.2). In either case, using \( n_j/n \to p_j > 0 \), (A.1) with \( \theta(t) = 1 \) and (A.5), we find \( \hat{\lambda}_n = O_p(n^{1/2}) \) uniformly for \( t \in \mathcal{T}_\omega \).

Case 2. \( t \in \mathcal{T}_\omega \). First suppose \( \hat{\lambda}_n(t) < 0 \). Noting \( -\log S_{2n_2(t)} - \hat{A}_2(t) \leq 0 \), from (A.2) we obtain

\[ a_1|\hat{\lambda}_n|^2 + b_1|\hat{\lambda}_n| + c_1 \leq 0, \quad (A.6) \]

where \( a_1 = -\log \theta(t) > 0, b_1 = n_1 \hat{A}_1(t) + n_2 \hat{A}_2(t) + (n_1 - n_2) (\log \theta(t) + \log S_{2n_2(t)} + \hat{A}_2(t)), c_1 = n_1n_2(\hat{A}_1(t) - \hat{A}_2(t) + \log \theta(t) + \log S_{2n_2(t)} + \hat{A}_2(t)) \). Along the lines of the argument leading to (A.5), for any \( \epsilon > 0 \) and \( n \) sufficiently large

\[ b_1 = n_1(\hat{A}_1(t) + \log S_1(t)) + n_2 \hat{A}_2(t) - n_2 \log \theta(t) - n_1 \log S_2(t) \]

\[ + (n_1 - n_2) (\log S_{2n_2(t)} + A_2(t) + \hat{A}_2(t) - A_2(t)) \]

\[ \geq \sqrt{n_1} \sqrt{n_2}(\hat{A}_1(t) - A_1(t)) + n_2 \frac{A_2(t)}{2} - n_2 \log \theta(t) - \frac{|n_1 - n_2|M_1}{\sqrt{n_2}} \]

\[ \geq -\sqrt{n_1} \sqrt{n_2} M_1 - n_2 \frac{\log S_2(t)}{4} - n_2 \log \theta(t) \]

\[ \geq -\frac{n_2 \log S_2(t)}{8} - n_2 \log \theta(t) > 0, \quad (A.7) \]
by \( n_j/n \to p_j > 0 \) uniformly in \( t \in \mathcal{F}_\prec \) with probability at least \( 1 - \varepsilon \), where \( M_1 > 0 \) is a constant. Since
\[
a_1 \lesssim -\log S_1(t) + \log S_2(t) \lesssim -\log S_1(\tau_2)
\]
and \( n_j/n \to p_j > 0 \), by (A.1) and (A.8) we obtain
\[
b_1^2 - 4a_1c_1 \geq \left[ \frac{-n_2 \log S_2(\tau_1)}{8} - n_2 \log \theta(t) \right]^2 + n_1 n_2 n^{-1/2} M_2 \log S_1(\tau_2)
\]
\[
\geq \left[ \frac{-n_2 \log S_2(\tau_1)}{16} - n_2 \log \theta(t) \right]^2 > 0,
\]
uniformly in \( t \in \mathcal{F}_\prec \) with probability at least \( 1 - \varepsilon \) for \( n \) sufficiently large, where \( M_2 > 0 \) is a constant. Thus the quadratic equation \( a_1 x^2 + b_1 x + c_1 = 0 \) has two distinct real roots \( x_1 < x_2 \) for all \( t \in \mathcal{F}_\prec \) with probability at least \( 1 - \varepsilon \) for \( n \) sufficiently large. Hence from (A.6) we have
\[
|\lambda_n| \leq x_2 = \frac{-b_1 + \sqrt{b_1^2 - 4a_1c_1}}{2a_1} = \frac{-2c_1}{b_1 + \sqrt{b_1^2 - 4a_1c_1}}
\]
for all \( t \in \mathcal{F}_\prec \) with probability at least \( 1 - \varepsilon \) for \( n \) sufficiently large. Second suppose \( \lambda_n(t) \geq 0 \). Then by (A.3)
\[
a_2|\lambda_n|^2 - b_2|\lambda_n| + c_2 \geq 0,
\]
where \( a_2 = a_1 = -\log \theta(t) > 0, b_2 = n_1 \hat{A}_1(t) + n_2 \hat{A}_2(t) + (n_1 - n_2)(\log \theta(t) - \log S_{1n}(t) - \hat{A}_1(t)), c_2 = n_1 n_2 (\hat{A}_1(t) - \hat{A}_2(t) + \log \theta(t) - \log S_{1n}(t) - \hat{A}_1(t)).
\]
The quadratic equation \( a_2 x^2 - b_2 x + c_2 = 0 \) has two distinct real roots \( x_3 < x_4 \) for all \( t \in \mathcal{F}_\prec \) with probability at least \( 1 - \varepsilon \) for \( n \) sufficiently large. Similarly
\[
b_2 \geq -\frac{n_2 \log S_2(\tau_1)}{8} - n_2 \log \theta(t) > 0,
\]
\[
b_2^2 - 4a_2c_2 \geq \left[ \frac{-n_2 \log S_2(\tau_1)}{16} - n_2 \log \theta(t) \right]^2 > 0
\]
with probability at least \( 1 - \varepsilon \). From (A.11) we know \( |\lambda_n| \leq x_3 \) or \( |\lambda_n| \geq x_4 \) in \( t \in \mathcal{F}_\prec \) with probability at least \( 1 - \varepsilon \) for \( n \) sufficiently large. If \( |\lambda_n| \geq x_4 = (b_2 + \sqrt{b_2^2 - 4a_2c_2})/(2a_2) > n_2 \) uniformly in \( t \in \mathcal{F}_\prec \) with probability at least \( 1 - \varepsilon \) for \( n \) sufficiently large, which contradicts \( |\lambda_n| \leq -D_2 = \min_{t_\varepsilon \leq t}(r_{2i} - d_{2i}) \leq n_2 - 1 \) (\( d_{2i} \geq 1 \)). Therefore
\[
|\lambda_n| \leq x_3 = \frac{b_2 - \sqrt{b_2^2 - 4a_2c_2}}{2a_2} = \frac{2c_2}{b_2 + \sqrt{b_2^2 - 4a_2c_2}}
\]
Using \( n_j/n \to p_j > 0 \), (A.1), (A.7), (A.9), (A.10), (A.12)–(A.14) we get
\[
P\{|\lambda_n(t)| \leq \sqrt{n} \max(M_3, M_4)\}
\geq P\{\text{for } \lambda_n(t) < 0, |\lambda_n(t)| \leq \sqrt{n}M_3, \text{for } \lambda_n(t) \geq 0, |\lambda_n(t)| \leq \sqrt{n}M_4\}
\[ P\{ \text{for } \hat{\lambda}_a(t) < 0, |\hat{\lambda}_a(t)| \leq \sqrt{n} M_2, \text{for } \hat{\lambda}_a(t) \geq 0, |\hat{\lambda}_a(t)| \leq \sqrt{n} M_4 \text{ or } \hat{\lambda}_a(t) > n_2 \} > 1 - e. \]

We find that \( \hat{\lambda}_a = \mathcal{O}(n^{1/2}) \) uniformly for \( t \in \mathcal{T}_< \).

**Case 3.** \( t \in \mathcal{T}_> \). This is similar to Case 2. First suppose \( \hat{\lambda}_a(t) < 0 \). Then from (A.2)

\[-a_1|\hat{\lambda}_a|^2 - b_1|\hat{\lambda}_a| - c_1 \geq 0,\]

where \( -a_1 = \log \theta(t) > 0 \). Second suppose \( \hat{\lambda}_a(t) \geq 0 \). Then by (A.3)

\[-a_2|\hat{\lambda}_a|^2 + b_2|\hat{\lambda}_a| - c_2 \leq 0,\]

where \( -a_2 = \log \theta(t) > 0 \). Following the same argument as in Case 2 we find that \( \hat{\lambda}_a = \mathcal{O}(n^{1/2}) \) uniformly for \( t \in \mathcal{T}_> \). This completes the proof. \( \square \)

**Proof of Theorem 2.1.** The key step in the proof is to show that the solution of Eq. (2.4) has the asymptotic expansion

\[ \lambda_a = -\frac{n}{\sigma^2(t)} L(\theta, t) + \frac{n \hat{v}(t)}{\sigma^2(t)} L_2(\theta, t) + \mathcal{O}(n^{-1/2}), \tag{A.15} \]

uniformly in \( t \) over the interval \( [\tau_1, \tau_2] \), where \( L(\theta, t) = \log (S_{1n}(t)/S_{2n}(t)) - \log \theta(t) \) and \( \hat{v}(t) = n^2 [\hat{\sigma}^2_1(t)/\sigma^2_1(t) - \hat{\sigma}^2_2(t)/\sigma^2_2(t)] \), where \( \hat{\sigma}^2_j(t) \) is defined as \( \hat{\sigma}^2_j(t) \) in (2.6), but with \( r_{ji} \) replaced by \( r_j^2 \). Note that \( \hat{v}(t) \) converges uniformly in probability to \( v(t) = \sigma^2_1(t)/\sigma^2_1(t) - \sigma^2_2(t)/\sigma^2_2(t) \) over \( [\tau_1, \tau_2] \), where \( \hat{\sigma}^2_j(t) \) is defined as \( \hat{\sigma}^2_j(t) \) in (2.5), but with \( H_j(s-) \) replaced by \( H_j(s-)^2 \). We establish (A.15) as follows. By an argument of Hollander et al. (1997, p. 225), for any \( \lambda_a = \mathcal{O}(n^{1/2}) \) the \( j \)th term \((j = 1, 2)\) on the l.h.s of (2.4) has an expansion

\[ \log S_{nj}(t) + \hat{\sigma}^2 j_0(t) \frac{(-1)^{j-1} \lambda_a}{n_j} - \frac{\hat{\sigma}^2_j(t) \lambda_a^2}{n_j^2} + \mathcal{O}(n^{-3/2}). \tag{A.16} \]

Using \( n_j/n \to p_j > 0 \), (2.4) and (A.16) we obtain

\[ 0 = L(\theta, t) + \frac{\hat{\sigma}^2(t) \hat{\lambda}_a}{n} - \hat{v}(t) \frac{\lambda_a^2}{n^2} + \mathcal{O}(n^{-3/2}). \]

Solving the above equation for \( \hat{\lambda}_a \), cf. Hollander et al. (1997), gives (A.15). By the functional delta method

\[ (\sqrt{n} j \log S_{nj} - \log S_j, j = 1, 2) \xrightarrow{\mathcal{D}} (U_{j'}, j = 1, 2), \tag{A.17} \]

so by \( n_j/n \to p_j > 0 \) we get

\[ \sqrt{n} L(\theta, t) \xrightarrow{\mathcal{D}} \frac{U_1(t)}{\sqrt{p_1}} + \frac{U_2(t)}{\sqrt{p_2}} = U(t). \tag{A.18} \]

A Taylor expansion, \( n_j/n \to p_j > 0 \) and (2.3) imply

\[ -2 \log R(\theta(t), t) = \frac{\hat{\sigma}^2(t) \lambda_a^2}{n} - \frac{4}{3} \hat{v}(t) \lambda_a^3 n^2 + \mathcal{O}(n^{-1}). \tag{A.19} \]

Substituting (A.15) into (A.19), we obtain

\[ -2 \log R(\theta(t), t) = \frac{nL^2(\theta(t), t)}{\hat{\sigma}^2(t)} - \frac{2}{3} \frac{n \hat{v}(t) \lambda_a^2}{\hat{\sigma}^2(t)} + \mathcal{O}(n^{-1}). \tag{A.20} \]
It follows from (A.17) that
\[ \sup_{s \in [t_1, t_2]} |\log S_{j_n}(s) - \log S_j(s)| \overset{P}{\to} 0, \quad j = 1, 2. \tag{A.21} \]
We have the inequality
\[
\left| n \left( \log \frac{S_{1n}(t)}{S_{2n}(t)} - \log \frac{S_1(t)}{S_2(t)} \right)^3 \right| \leq nL^2(\theta, t) \left( \sup_{s \in [t_1, t_2]} |\log S_{1n}(s) - \log S_1(s)| \right) \\
+ nL^2(\theta, t) \left( \sup_{s \in [t_1, t_2]} |\log S_{2n}(s) - \log S_2(s)| \right). \tag{A.22} \]
Combining (A.18), (A.21) and (A.22) we have
\[
\sup_{s \in [t_1, t_2]} \left| n \left( \log \frac{S_{1n}(s)}{S_{2n}(s)} - \log \frac{S_1(s)}{S_2(s)} \right)^3 \right| \overset{P}{\to} 0. \tag{A.23} \]
Combining (A.18), (A.20), (A.23) and the uniform consistency of \( \hat{\sigma}^2(t) \) and \( \hat{v}(t) \) completes the proof. \( \Box \)

References