Likelihood Ratio–Based Confidence Bands for Survival Functions

Myles HOLLANDER, Ian W. MCKEAGUE, and Jie YANG

Thomas and Grunkemeier introduced a nonparametric likelihood ratio approach to confidence interval estimation of survival probabilities based on right-censored data. We construct simultaneous confidence bands using this approach. The boundaries of the bands are contained within [0, 1]. A procedure essentially equivalent to a bias correction is developed. The resulting increase in coverage accuracy is illustrated by an example and a simulation study. We look at various versions of log-likelihood ratio–based confidence bands and compare them to the Hall–Wellner band and Nair's equal precision band. We also construct likelihood ratio–based bands for cumulative hazard functions.

KEY WORDS: Bias correction; Cumulative hazard rate; Kaplan–Meier estimator; Likelihood ratio statistic.

1. INTRODUCTION

Thomas and Grunkemeier (1975)—TG hereafter—introduced a nonparametric likelihood ratio (LR) method for obtaining confidence intervals for a survival function $S_0(t)$ at a given time point t. Our problem is to construct a *simultaneous confidence band* for $S_0(t)$ over the time span of interest. We show that TG's pointwise confidence intervals can be adapted for this purpose. Our approach is based on their nonparametric LR statistic

$$R(p,t) = \frac{\sup\{L(S)\colon\ S(t) = p, S \in \Theta\}}{L(S_n)}, \tag{1}$$

where Θ is the family of all discrete survival functions supported by the uncensored lifetimes, $0 < p < 1$, S_n is the Kaplan–Meier estimator of S_0, and L is the likelihood function

$$L(S) = \prod_u [S(T_i-) - S(T_i)] \prod_c S(T_i), \tag{2}$$

where the T_i are the possibly right-censored failure times. The first product in (2) is taken over uncensored times, and the second product is taken over censored times. Kaplan and Meier (1958) showed that S_n, given by (13), is a nonparametric maximum likelihood estimator in the sense that it maximizes L without constraint.

Li (1995a) showed that the LR defined by (1) is unchanged when Θ is replaced by the set of all survival functions on the interval $[0, \infty)$, so it represents a "full" LR. The TG asymptotic $100(1-\alpha)\%$ confidence interval for $S_0(t)$ is given by

$$\{p\colon\ -2\log R(p,t) \le \chi^2_{1,\alpha}\}, \tag{3}$$

where $0 < \alpha < 1$ and $\chi^2_{q,\alpha}$ is the upper α quantile of the chi-squared distribution with q degrees of freedom.

A simultaneous large-sample confidence band for S_0 can be obtained essentially by replacing p and $\chi^2_{1,\alpha}$ in (3) by a survival function and an appropriate threshold determined by the data. We obtain a LR-based confidence band of the form

$$\mathcal{B} = \{S(t)\colon\ -2\log R(S(t),t) \le C^2(t), \qquad t \in [0,\tau]\} \tag{4}$$

such that

$$P(S_0 \in \mathcal{B}) \to 1-\alpha, \quad \text{as} \quad n \to \infty, \tag{5}$$

where $C(t)$ is given by (9) and τ is the end of follow-up. The ratio $L(S)/L(S_n)$ may be regarded to be an "inverse distance" between S and S_n in the sense that the larger its value, the closer S is to S_n. Thus $\mathcal{B}$ may be interpreted as a "neighborhood" of the Kaplan–Meier estimator S_n.

Note that for a fixed p, $-2\log R(p,t)$ is the empirical LR for the upper p quantile and $\{t_0\colon\ -2\log R(p,t_0) \le \chi^2_{1,\alpha}\}$ is the corresponding empirical likelihood confidence set. Owen (1988, 1990) used the empirical LR approach to construct confidence intervals for the mean, for a class of M estimates that includes quantiles, and for other differentiable statistical functionals in noncensored iid settings. This method has recently been further extended to deal with problems arising in linear regression, generalized linear models, and other settings. It is advantageous to use empirical likelihood for several reasons. As noted by Hall and La Scala (1990), empirical confidence regions automatically reflect emphasis on the observed data set. This is seen in TG's confidence intervals for survival probabilities. Moreover, empirical likelihood regions are range preserving and transformation respecting. That is, a LR based confidence interval for $\phi(\theta)$, a function of a parameter θ, is obtained by applying ϕ to each value in the corresponding confidence interval for θ. This suggests that our LR confidence band for survival functions can be transformed to give a confidence band for the cumulative hazard function $A(t) = \phi(S)(t) = \int_0^t S(s-)^{-1}\,dF(s), t \le \tau$, where $F = 1 - S$. In Section 3 we show that this is a valid procedure.

It was shown by TG and by Li (1995a) that $\mathcal{L}(S,t) = -2\log R(S(t),t)$ converges in distribution to chi-squared with 1 df for each fixed $t \in [0,\tau]$. We consider the signed root-log-LR statistic

Myles Hollander is Distinguished Research Professor and Ian W. McKeague is Professor and Chairman, Department of Statistics and Statistical Consulting Center, Florida State University, Tallahassee, FL 32306. Jie Yang is Research Associate, John S. Stafford Trading, Chicago, IL 60605. Ian McKeague acknowledges research support from National Science Foundation grant ATM-9417528.

Journal of the American Statistical Association
March 1997, Vol. 92, No. 437, Theory and Methods

$$W(S,t) = \operatorname{sgn}(S_n(t) - S(t))\sqrt{\mathcal{L}(S,t)} \qquad (6)$$

and show that the process $\{\hat{\sigma}(t)W(S_0,t), t \in [0,\tau]\}$ converges weakly to a mean zero Gaussian martingale with variance that is consistently estimated by $\hat{\sigma}^2(t) = \hat{\sigma}_0^2(t)$, given by (17). The limiting process can be transformed to a Brownian bridge B^0, so that

$$\sup_{t\in[0,\tau]} \left|\frac{\hat{\sigma}(t)W(S_0,t)}{1+\hat{\sigma}^2(t)}\right| \xrightarrow{\mathcal{D}} \sup_{x\in[0,d]} |B^0(x)|, \qquad (7)$$

where d is consistently estimated by $\hat{d} = \hat{\sigma}^2(\tau)/\{1+\hat{\sigma}^2(\tau)\}$ (see Andersen, Borgan, Gill, & Keiding 1993, Chap. IV, Sec. 1.3). This result is used to construct our asymptotic $100(1-\alpha)\%$ confidence band for S_0:

$$\mathcal{B} = \left\{S(t) : \left|\frac{\hat{\sigma}(t)W(S,t)}{1+\hat{\sigma}^2(t)}\right| \le K_{q,\alpha}(d), t \in [0,\tau]\right\}$$

$$= \{S(t) : |W(S(t),t)| \le C(t), t \in [0,\tau]\}, \qquad (8)$$

where

$$C(t) = K_{q,\alpha}(\hat{d})\,\frac{1+\hat{\sigma}^2(t)}{\hat{\sigma}(t)}, \qquad t > 0, \qquad (9)$$

$C(0) = 0$, and $K_{q,\alpha}(d)$ is the upper α quantile of the distribution of $\sup_{x\in[0,d]} |B^0(x)|$. An explicit formula for the distribution was given by Hall and Wellner (1980, eq. 2.9). (Some critical values for different choices of α can also be found in that work.) Because $\hat{\sigma}^2(t)$ is a step function, $C(t)$ is also a step function, so the confidence band can be computed in finitely many steps (see Sec. 2). As we have seen in (4), the confidence band can be interpreted as a class of survival functions. The boundaries of the band are restricted by [0, 1], which is especially appealing for survival probabilities near 0 or 1.

Simultaneous confidence bands for the survival function based on the limiting distribution of the Kaplan–Meier estimator, as obtained by Efron (1967) and Breslow and Crowley (1974), have been available since the work of Gillespie and Fisher (1979). The most commonly used confidence band is that due to Hall and Wellner (1980), HW hereafter. They proposed the asymptotic $100(1-\alpha)\%$ confidence band

$$S_n(t) \pm n^{-1/2}S_n(t)K_{q,\alpha}(d)(1+\hat{\sigma}^2(t)), \qquad t \in [0,\tau]. \quad (10)$$

This band reduces to the well-known Kolmogorov band for uncensored survival data. Another commonly used confidence band is the equal precision band (Nair 1984)

$$S_n(t) \pm n^{-1/2}e_\alpha S_n(t)\hat{\sigma}(t), \quad \forall\, t\colon\ a \le \frac{\hat{\sigma}^2(t)}{1+\hat{\sigma}^2(t)} \le b. \quad (11)$$

Here $e_\alpha = e_\alpha(a,b)$ satisfies

$$P\left\{\sup_{u\in[a,b]} \frac{|B^0(u)|}{[u(1-u)]^{1/2}} \le e_\alpha\right\} = 1-\alpha.$$

To construct this band, one must choose values of a and b. Some critical values for different choices of α and $a = 1-b$ were given by Nair (1984, tab. 2). The equal precision (EP) band has the same form as the standard asymptotic pointwise confidence intervals for $S_0(t)$: $S_n(t) \pm z_{\alpha/2}n^{-1/2}S_n(t)\hat{\sigma}(t)$. When there is no censoring, the band reduces to

$$S_n(t) \pm e_\alpha\sqrt{[S_n(t)(1-S_n(t))]/n}. \qquad (12)$$

Nair's simulations indicate that the Rényi-type band described by Gill (1980) is generally inferior to the other bands, and thus we have excluded it from our numerical study.

A shortcoming of the usual HW and the EP confidence bands is that they may contain values outside [0, 1]. One way to overcome this problem is to use the log-log transformation, $g(x) = \log(-\log x)$, or the arcsine transformation, $g(x) = \arcsin\sqrt{x}$, suggested, respectively, by Kalbfleisch and Prentice (1980) and TG. The latter is variance stabilizing for the situation with no censoring. Our LR band does not require such an ad hoc procedure. Nevertheless, it is possible and advantageous to consider transformations to improve the approximation to the asymptotic distribution. With the transformation that Nair used to obtain his EP band, we can have an LR band $\mathcal{B} = \{S(t)\colon \mathcal{L}(S,t) \le e_\alpha^2(a,b)\}$. Essentially, this takes the same form as TG's confidence interval in (3), except that the threshold $C^2(t)$ is replaced by a fixed critical value larger than $\chi^2_{1,\alpha}$.

We recommend the LR confidence bands because, as we show herein (a) they have satisfactory coverage accuracy, (b) their boundaries are naturally contained in [0, 1], and (c) they are transformation preserving.

The LR approach is useful in other incomplete data problems in survival analysis. For example, Li, Hollander, McKeague, and Yang (1996) found confidence bands for the quantile function, Li (1995b) gave confidence intervals for survival probabilities based on truncated data, and Murphy (1995) gave pointwise confidence intervals for the survival function and the cumulative hazard function.

In Section 2 we introduce our two LR confidence bands and develop the requisite asymptotic theory. The second of these bands corrects for possible "small-sample" bias in the first band. We discuss LR confidence bands for cumulative hazard functions in Section 3. In Section 4 we illustrate our survival function bands in examples and report the results of a simulation study. We compare our LR bands to the Hall–Wellner band and to Nair's equal precision band, including their arcsin and log-log transformed bands. We present proofs in the Appendix.

2. CONSTRUCTION OF CONFIDENCE BANDS

Our basic LR confidence band for the survival function is studied in Section 2.1. The bias correction is discussed in Section 2.2.

2.1 Construction of Basic Likelihood Ratio Bands

Let $X_1, \ldots, X_n$ be iid survival times with survival function S, and let $C_1, \ldots, C_n$ be iid censoring times with survival function S_C, independent of the X_i's. We observe

Table 1. The 1994 JASA Time-to-First-Review Data (Time in Days)

T_i	δ_i	T_i	δ_i	T_i	δ_i	T_i	δ_i	T_i	δ_i	T_i	δ_i	T_i	δ_i	T_i	δ_i	T_i	δ_i	T_i	δ_i	T_i	δ_i
214	1	201	1	28	1	252	0	118	1	187	0	28	1	28	1	76	1	56	0	28	0
184	1	274	1	287	0	96	1	33	1	152	1	21	1	118	0	18	1	21	1	27	0
150	1	265	1	195	1	175	1	69	1	46	1	1	1	40	1	88	0	55	0	27	0
70	1	120	1	86	1	54	1	133	1	103	1	0	1	6	1	85	0	55	0	25	0
16	1	141	1	137	1	167	1	126	1	37	1	144	0	91	1	85	0	54	0	25	0
141	1	48	1	74	1	150	1	84	1	170	1	144	0	34	1	85	0	18	1	22	0
210	1	204	1	71	1	219	1	197	1	64	1	140	0	21	1	20	1	54	0	22	0
132	1	312	0	140	1	86	1	85	1	182	0	14	1	1	1	83	0	53	0	21	0
30	1	220	1	22	1	1	1	15	1	180	0	0	1	111	0	82	0	50	0	21	0
204	1	188	1	120	1	111	1	206	1	176	0	27	1	111	0	81	0	50	0	15	0
84	1	84	1	176	1	128	1	125	1	175	0	23	1	1	1	81	0	1	1	15	0
36	1	84	1	181	1	178	1	57	1	64	1	126	1	48	1	11	1	15	1	15	0
38	1	215	1	155	1	40	1	181	1	42	1	139	0	110	0	77	0	50	0	1	1
69	1	33	1	74	1	131	1	215	0	175	0	55	1	47	1	77	0	47	0	1	1
33	1	55	1	29	1	20	1	3	1	149	1	137	0	68	1	70	1	47	0	14	0
49	1	140	1	100	1	220	1	13	1	158	1	114	1	74	1	74	0	46	0	12	0
203	1	147	1	195	1	84	1	175	1	169	0	56	1	98	1	71	0	16	1	12	0
203	1	41	1	127	1	32	1	37	1	169	0	124	1	105	0	23	1	43	0	12	0
218	1	94	1	34	1	95	1	182	1	22	0	121	1	104	0	28	1	43	0	8	0
267	1	292	1	177	1	188	1	210	0	168	0	1	1	104	0	70	0	18	1	8	0
99	1	131	1	150	1	115	1	92	1	157	1	27	1	103	0	44	1	43	0	8	0
21	1	221	1	265	0	238	0	208	0	89	1	130	0	90	1	69	0	42	0	8	0
78	1	39	1	174	1	1	1	30	1	165	1	130	0	98	0	68	0	42	0	7	0
150	1	3	1	104	1	187	1	28	1	14	1	130	0	98	0	67	0	40	0	7	0
237	1	16	1	203	1	125	1	168	1	161	0	127	0	98	0	64	1	0	1	7	0
91	1	129	1	109	1	110	1	202	0	161	0	100	1	97	0	30	1	12	1	7	0
21	1	210	1	217	1	32	1	114	1	159	0	126	0	96	1	41	1	39	0	6	0
224	1	240	1	238	1	32	1	105	1	91	1	126	0	97	0	62	0	35	0	5	0
126	1	141	1	210	1	228	1	196	0	146	1	28	1	18	1	61	0	35	0	5	0
167	1	231	1	22	1	80	1	195	0	159	0	125	0	96	0	20	1	35	0	4	0
105	1	119	1	148	1	64	1	114	1	134	1	125	0	92	0	57	0	30	1	1	0
146	1	291	0	142	1	231	0	75	1	13	1	125	0	91	0	57	0	35	0	1	0
50	1	199	1	126	1	64	1	194	0	159	0	95	1	91	0	57	0	34	0		
28	1	67	1	220	1	228	0	143	1	18	1	95	1	91	0	57	0	34	0		
288	1	263	1	145	1	18	1	106	1	155	0	123	0	31	1	57	0	34	0		
37	1	155	1	21	1	55	1	128	1	154	0	123	0	27	1	57	0	33	0		
18	1	189	1	256	0	154	1	200	0	124	1	123	0	83	1	42	1	33	0		
113	1	0	1	253	0	139	1	129	1	73	1	109	1	33	1	57	0	29	1		
22	1	209	1	22	1	91	1	138	1	51	1	119	0	11	1	57	0	28	0		
234	1	223	1	80	1	196	1	152	1	21	1	6	1	88	0	57	1	27	0		

$(T_1, \delta_1), \ldots, (T_n, \delta_n)$, where $T_i = \min(X_i, C_i), \delta_i = I(X_i \leq C_i)$, and $I(E)$ is the indicator of the event E. We reserve the notation S_0 for the true underlying survival function.

We introduce some notation as follows. Let $Y(t) = \sum_{i=1}^{n} I(T_i \geq t)$ be the number of individuals at risk just before time t, let $N(t) = \sum_{i=1}^{n} I\ (T_i \leq t, \delta_i = 1)$ be the counting process that records the uncensored failures, and let $\Delta N(s) = N(s) - N(s-)$. The Kaplan–Meier estimator of S_0 is

$$S_n(t) = \prod_{s \leq t} \left(1 - \frac{\Delta N(s)}{Y(s)}\right), \qquad 0 \leq t < \infty, \qquad (13)$$

where $\Delta N(s)/Y(s)$ is defined to be 0 whenever $Y(s) = 0$. Due to tail instability of the Kaplan–Meier estimator, simultaneous confidence bands can be obtained only on an interval $[0, \tau]$, where $H(\tau) > 0$ and $H(t) = S_0(t) S_C(t)$. We fix such a τ from now on.

We show that $W(S_0, t)$ converges weakly to a Gaussian process; that is the basic result for constructing our confidence band. To that end, recall from TG that

$$\mathcal{L}(S, t) = -2 \sum_{s \leq t} \left\{ [Y(s) - \Delta N(s)] \times \log\left(1 + \frac{\lambda_n}{Y(s) - \Delta N(s)}\right) \right\} + 2 \sum_{s \leq t} \left\{ Y(s) \log\left(1 + \frac{\lambda_n}{Y(s)}\right) \right\}, \qquad (14)$$

where $\lambda_n = \lambda_n(S, t)$ satisfies

$$\prod_{s \leq t} \left(1 - \frac{\Delta N(s)}{Y(s) + \lambda_n}\right) = S(t). \qquad (15)$$

To obtain an explicit expression for W, it is necessary to solve (15) for λ_n. An explicit solution is not possible, but a suitable approximation for $\mathcal{L}(S, t)$ will suffice. Thus our first step is to find an approximation for λ_n. For future reference, we define

$$\sigma_r^2(t) = \int_0^t \frac{dA(s)}{H(s-)^{r+1}}, \qquad t \in [0, \tau], \qquad r = 0, 1, \ldots, 4, \qquad (16)$$

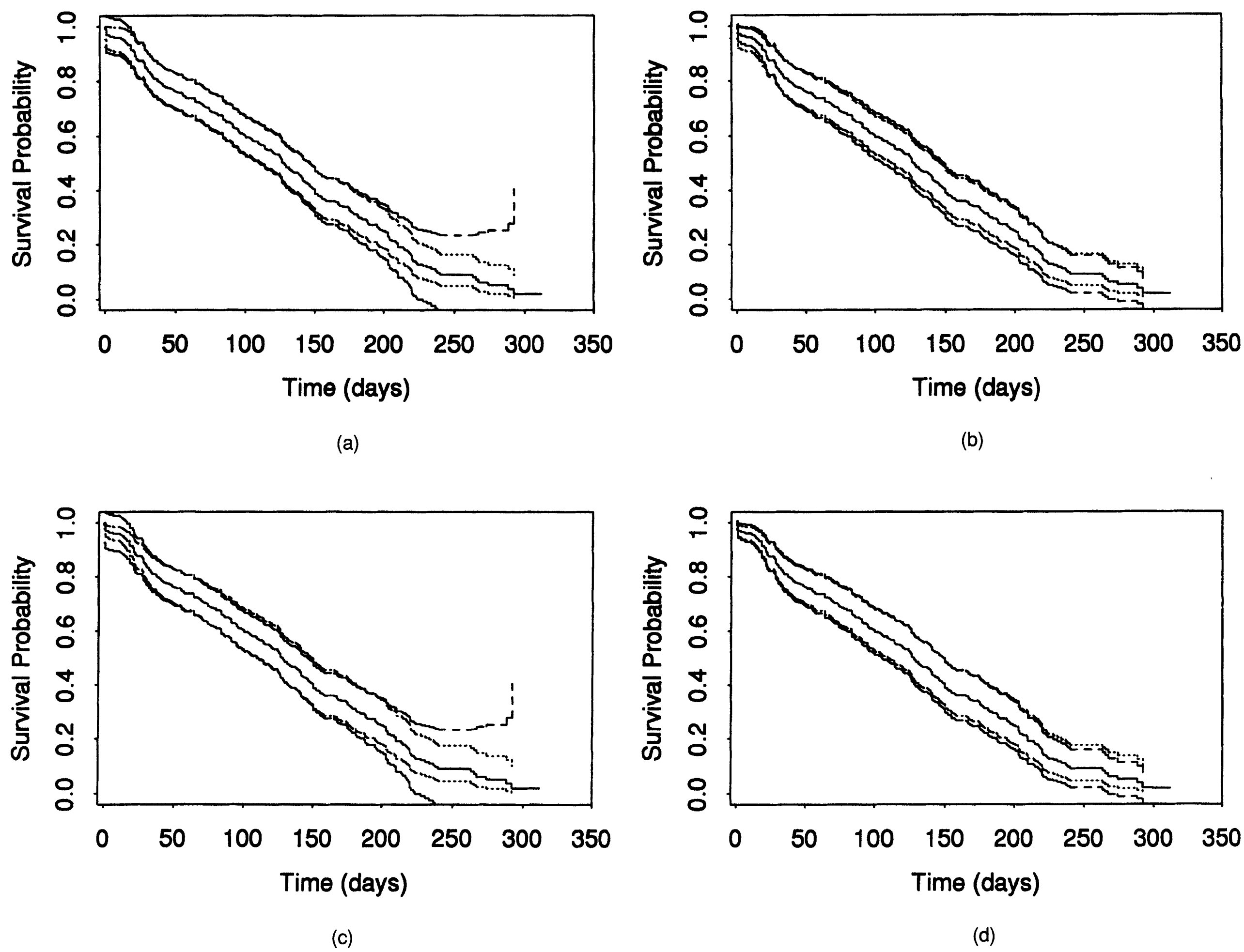

Figure 1. Estimated Probability (Solid Line) that a First Review of a JASA Manuscript Takes Longer Than t Days, With 95% LR Bands (Dotted Lines) and HW and EP Bands (Dashed Lines). (a) LR_1^c and HW; (b) LR_1^c and EP; (c) LR_2^c and HW; (d) LR_2^c and EP.

and $\sigma^2(t) = \sigma_0^2(t)$. The asymptotic variance of the Kaplan–Meier estimator at time t is $\sigma^2(t)S_0^2(t)$. It is easy to show that

$$\hat{\sigma}_r^2(t) = n^{r+1}\int_0^g \frac{I(Y>0)}{Y^{r+1}(Y-\Delta N)}\,dN \tag{17}$$

is a uniformly consistent estimator of $\sigma_r^2(t), t \in [0,\tau]$. Also, denote $\hat{\sigma}^2(t) = \hat{\sigma}_0^2(t)$ and

$$K(S,t) = \log S_n(t) - \log S(t). \tag{18}$$

Theorem 2.1. The signed root-log-LR of $\mathcal{L}(S,t)$ has an asymptotic expansion

$$W(S,t) = \frac{\sqrt{n}K(S,t)}{\hat{\sigma}} - \frac{\sqrt{n}}{3}\frac{\hat{\sigma}_1^2 K^2(S,t)}{\hat{\sigma}^5} + \frac{\sqrt{n}}{36\hat{\sigma}^9}(16\hat{\sigma}_1^4 - 9\hat{\sigma}^2\hat{\sigma}_2^2)K^3(S,t) + O_p(n^{-3/2}),$$

where the dependence of $\hat{\sigma}_r$ on t has been suppressed.

This and all subsequent expansions involving O_p terms hold uniformly in t over the interval $[0,\tau]$.

Remark 2.1. To find the asymptotic distribution of W, we need only the first two terms of the foregoing expansion. The two-term expansion of W is

$$W(S,t) = \frac{\sqrt{n}K(S,t)}{\hat{\sigma}} - \frac{\sqrt{n}}{3}\frac{\hat{\sigma}_1^2 K^2(S,t)}{\hat{\sigma}^5} + O_p(n^{-1}). \tag{19}$$

To derive the limiting distribution of W, we need a lemma.

Lemma 2.1. $\sqrt{n}K(S_0,t)$ converges weakly to $U(t)$ in $D[0,\tau]$ as $n \to \infty$, where $U(t)$ is a Gaussian martingale with mean zero and variance function $\sigma^2(t)$.

Theorem 2.2. $\hat{\sigma}(t)W(S_0,t)$ converges weakly to $U(t)$ in $D[0,\tau]$, where $U(t)$ is defined in Lemma 2.1.

It follows from Theorem 2.2 that

$$\frac{\hat{\sigma}(t)W(S_0,t)}{1+\hat{\sigma}^2(t)} \xrightarrow{\mathcal{D}} \frac{U(t)}{1+\sigma^2(t)} \tag{20}$$

in $D[0,\tau]$, so

$$\sup_{t\in[0,\tau]}\left|\frac{\hat{\sigma}(t)W(S_0,t)}{1+\hat{\sigma}^2(t)}\right| \xrightarrow{\mathcal{D}} \sup_{x\in[0,d]}|B^0(x)|, \tag{21}$$

where

$$d = \frac{\sigma^2(\tau)}{1+\sigma^2(\tau)}, \tag{22}$$

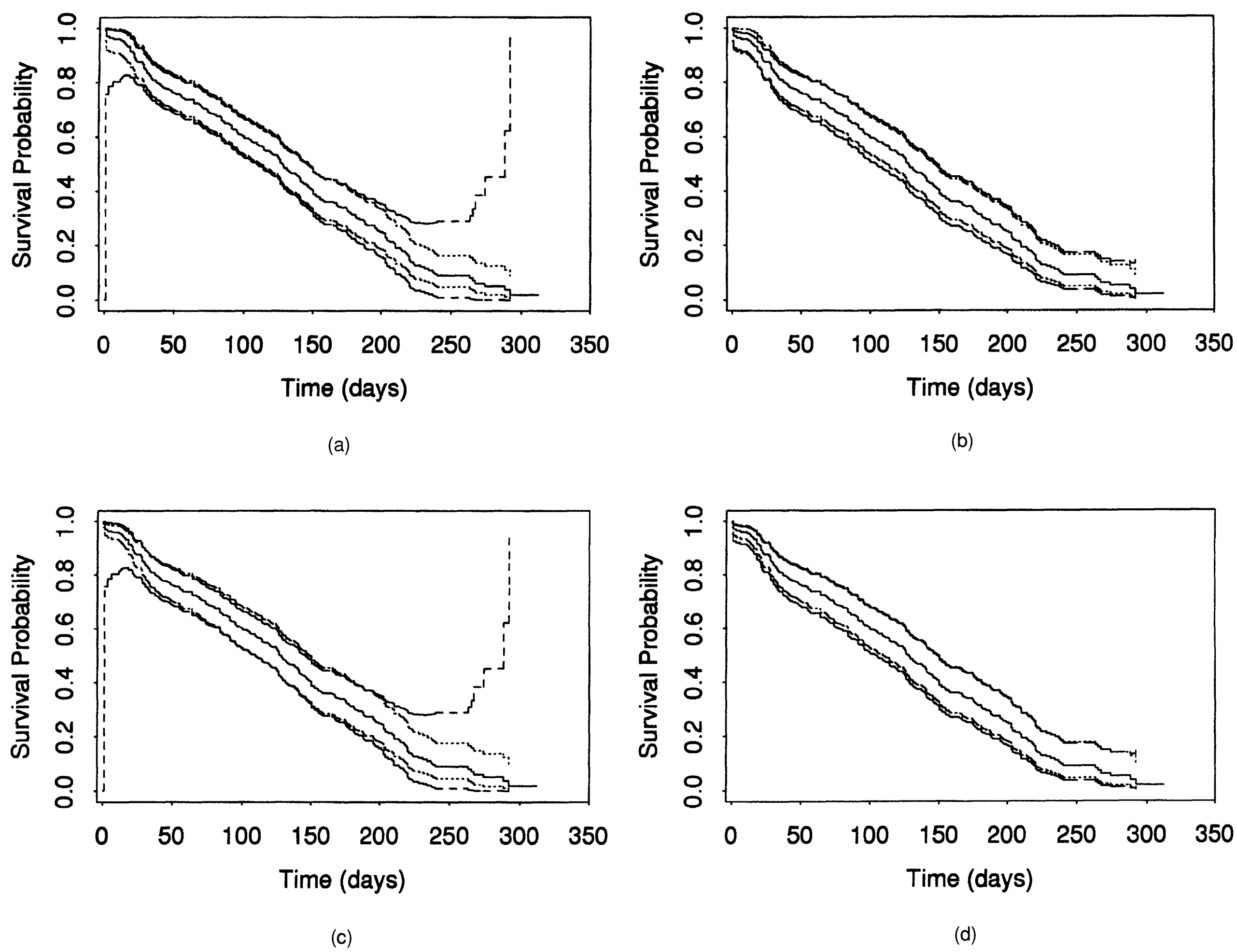

Figure 2. Comparison of LR-, EP-, and Log-Log-Transformed Bands for the JASA Data. See Figure 1 for key.

because the processes $B^0\{\sigma^2(\cdot)/[1+\sigma^2(\cdot)]\}$ and $U(\cdot)/[1+\sigma^2(\cdot)]$ have the same distribution.

From (21) we can easily construct a confidence band for $S_0, \mathcal{B} = \{S(t)\colon \mathcal{L}(S,t) \le C^2(t), t \in [0,\tau]\}$; refer to (8). We call this band LR_1.

Because $\hat{\sigma}^2(t)$ is a step function, so is $C(t)$. Suppose that there are k observed distinct uncensored life times, $T_{(1)} < \cdots < T_{(k)}$. We can write $C(s) = \sum_{i=1}^{k} a_i I_{E_i}(s)$, where $E_i = [T_{(i-1)}, T_{(i)}), i = 2, \ldots, k+1$, and $T_{(k+1)} = \tau$. For each fixed $i \le k$, piecewise constant confidence limits, $[P_L, P_U]$, for $S_0(t), t \in E_i$ are given by

$$P_L = \prod_{s \le t} \left\{1 - \frac{\Delta N(s)}{Y(s) + \lambda_L}\right\}$$

and

$$P_U = \prod_{s \le t} \left\{1 - \frac{\Delta N(s)}{Y(s) + \lambda_U}\right\},$$

where $\lambda_L < 0 < \lambda_U$ are the two solutions of the equation

$$-2 \log R(\lambda, t) = C^2(t) = a_i^2, \qquad t \in E_i \qquad (23)$$

(see Li 1995a or Thomas and Grunkemeier 1975). We need to use a numerical algorithm to obtain the two roots of the foregoing nonlinear equation. The following two facts ensure the existence of exactly two roots, which are easy to find due to the monotonicity of $-2\log R(\cdot, t)$:

a. For each fixed $t, -2\log R(\cdot, t)$ is strictly decreasing on the interval $(D, 0]$ and increasing on $[0, \infty)$, where $D = N(t) - Y(t)$. To see this, note that the partial derivative of $-2\log R(\lambda, t)$ with respect to λ is

$$\int_0^t \frac{2\lambda\, dN(s)}{(Y(s) - \Delta N(s) + \lambda)(Y(s) + \lambda)}, \qquad (24)$$

which is negative for $\lambda \in (D, 0)$, zero for $\lambda = 0$, and positive for $\lambda > 0$ (cf. Thomas and Grunkenmeier 1975).

b. $-2\log R(\lambda, t) \to +\infty$ as $\lambda \to +\infty, -2\log R(\lambda, t) \to +\infty$ as $\lambda \to D$, and $-2\log R(0, t) = 0$.

2.2 Bias-Corrected Likelihood Ratio Bands

The function $C(t)$ defined in (9) is based on the asymptotic distribution of W. The departure of the actual distribution of W from its asymptotic distribution will directly

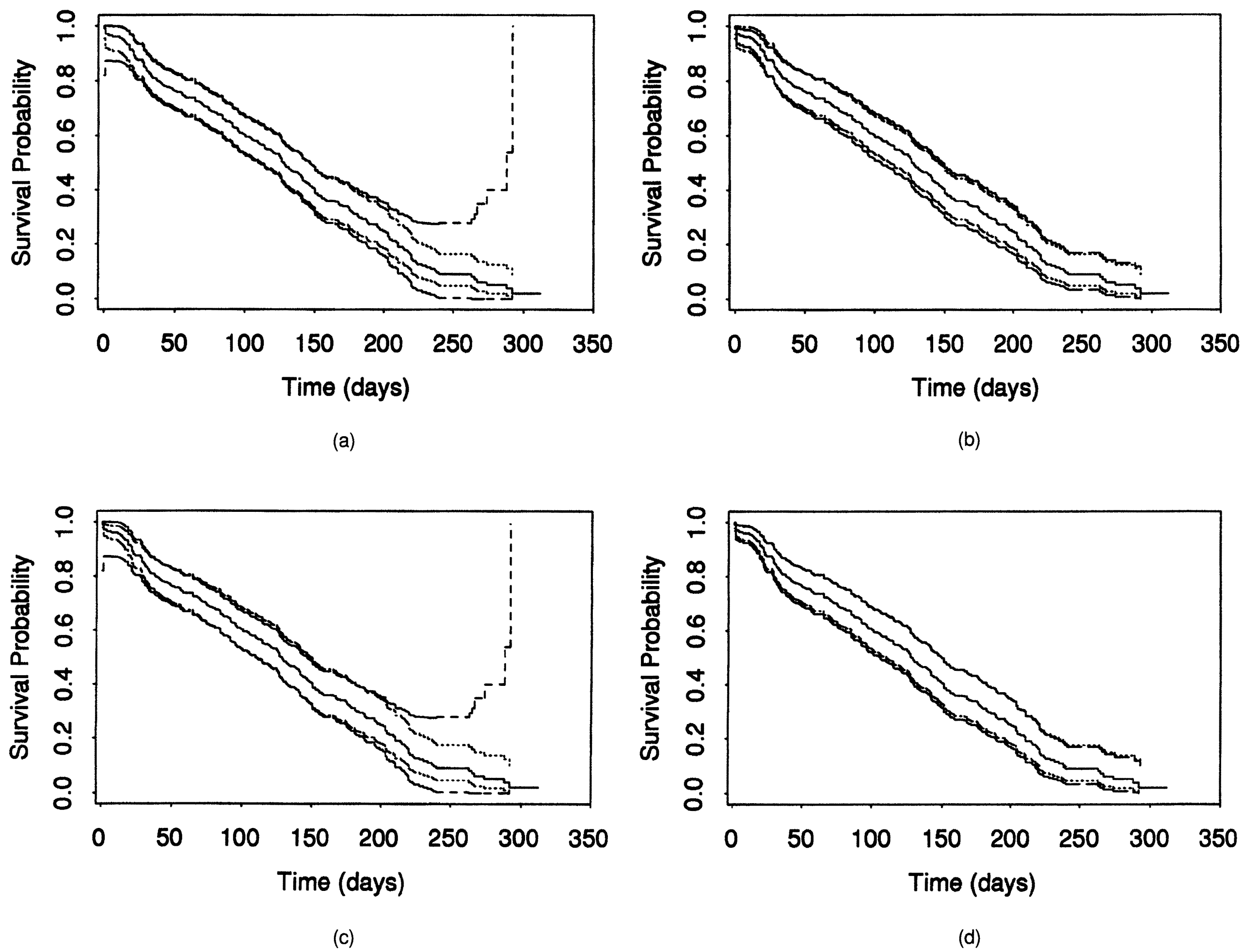

Figure 3. Comparison of LR-, EP-, and Arcsin-Transformed HW Bands for the JASA Data. See Figure 1 for key.

affect the width of our confidence band, most critically in small samples. One way to overcome this problem is to bootstrap the distribution of W. An alternative approach is to use a correction or adjustment that can improve the approximation to the asymptotic distribution. DiCiccio and Romano (1989) introduced a mean adjustment to the empirical LR. In this section we follow their idea to obtain a bias correction to our confidence bands by adjusting $\mathcal{L}(S,t)$ to achieve a similar effect. From now on, a tilde placed over an expression indicates that it has been bias corrected. Let

$$\begin{aligned} \tilde{W}(S,t) &= W(S,t) + \frac{\sqrt{n}}{3}\frac{\hat{\sigma}_1^2 K^2(S,t)}{\hat{\sigma}^5} \\ &= \frac{\sqrt{n}K(S,t)}{\hat{\sigma}} + O_p(n^{-1}), \end{aligned} \tag{25}$$

where we have used (19). Note that

$$W = \frac{\sqrt{n}K(S,t)}{\hat{\sigma}} + O_p(n^{-1/2}),$$

because the second term in $\tilde{W}$ is $O_p(n^{-1/2})$. They both have the same asymptotic distribution; that is, they satisfy (20) and (21). Ignoring terms of order $O_p(n^{-1})$ in $\tilde{W}^2$, let

$$\begin{aligned} \tilde{\mathcal{L}}(S,t) &= \mathcal{L}(S,t) + \frac{2}{3}\frac{n\hat{\sigma}_1^2 K^3(S,t)}{\hat{\sigma}^6} \\ &= \frac{nK^2(S,t)}{\hat{\sigma}^2} + O_p(n^{-1}). \end{aligned} \tag{26}$$

Our bias-corrected confidence band is

$$\tilde{\mathcal{B}} = \{S(t)\colon\ \tilde{\mathcal{L}}(S,t) \le C^2(t), t \in [0,\tau]\}.$$

DiCiccio and Romano (1989) showed that a mean adjustment improves the coverage accuracy of certain empirical LR–based confidence intervals from $O(n^{-1/2})$ to $O(n^{-1})$. We expect the same improvement in our bias-corrected confidence bands, but because censoring complicates the problem, a proof is beyond the scope of this article. In the case of pointwise confidence intervals for the survival probability, the adjusted confidence set is

$$\{p\colon\ -2\log R(p,t) \le \tilde{\chi}_{1,\alpha}\},$$

where

$$\tilde{\chi}_{1,\alpha} = \chi_{1,\alpha} - \frac{2}{3}\frac{n\hat{\sigma}_1^2 K^3(p,t)}{\hat{\sigma}^6}$$

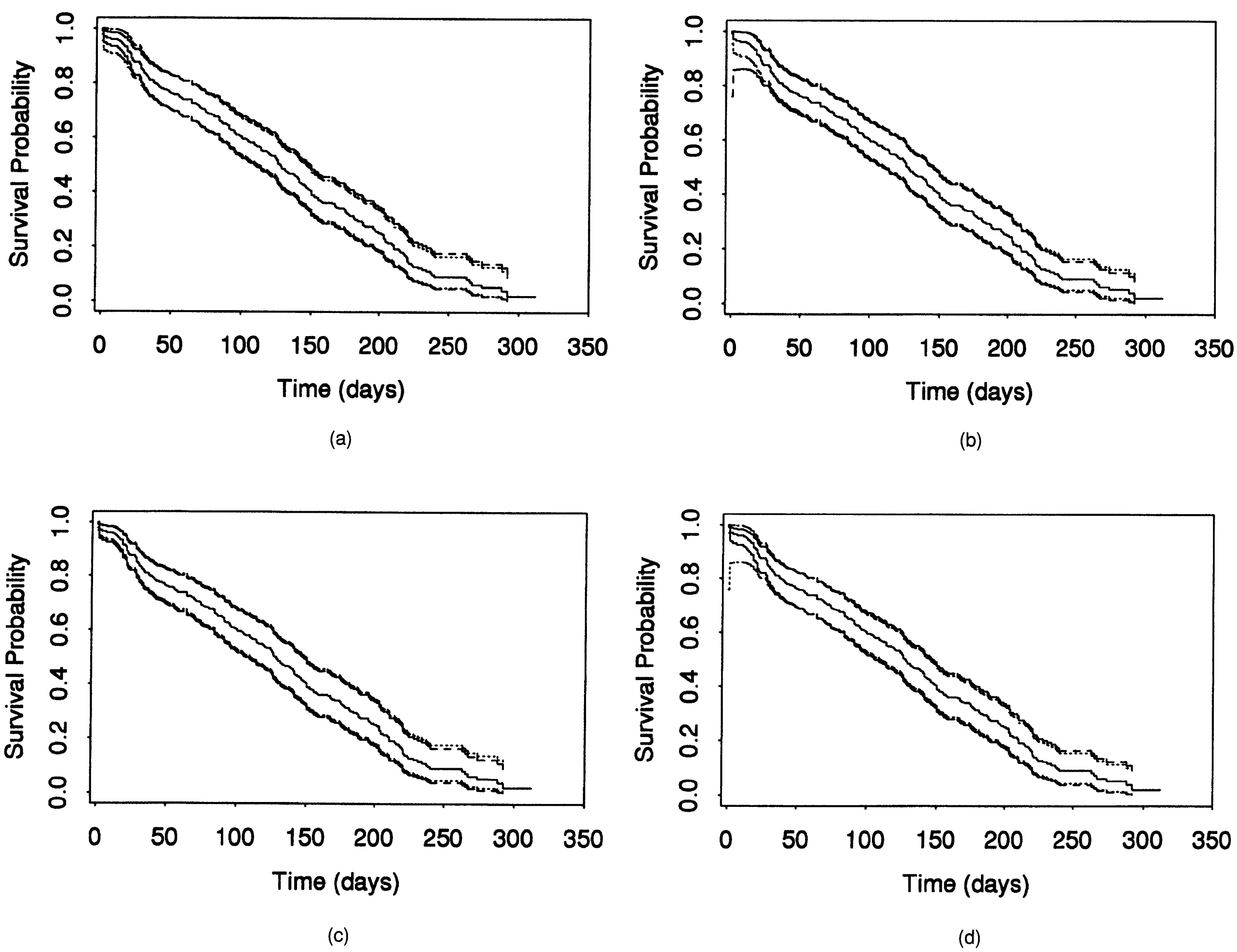

Figure 4. Effect of the Bias Correction on the LR Bands (JASA Data). (a) LR_1^c (dotted lines) and LR_2^c (dashed lines); (b) LR_1^c (dotted lines) and LR_1 (dashed lines); (c) LR_2^c (dotted lines) and LR_2 (dashed lines); (d) LR_1 (dotted lines) and LR_2 (dashed lines).

and

$$K(p,t) = \log S_n(t) - \log p.$$

In Section 4 we provide examples and simulation results to examine the effect of the bias correction.

3. LIKELIHOOD RATIO CONFIDENCE BANDS FOR CUMULATIVE HAZARD RATE

Another problem related to determining confidence bands for the survival function is to construct confidence bands for the cumulative hazard rate

$$A(t) = \phi(S)(t) = -\int_0^t \frac{dS(s)}{S(s-)}, \qquad t \in [0,\tau]. \quad (27)$$

Here we regard A as a functional ϕ of S. The true underlying cumulative hazard rate is denoted by A_0. Confidence bands for A_0 based on the limiting distribution of the Nelson–Aalen estimator $\hat{A}$ have been studied by many authors (see, e.g., Bie, Borgan, and Liestøl 1987). In this section we derive LR confidence bands for A_0 and show that they arise from applying ϕ to our LR confidence bands for S_0.

The log-LR with respect to A is

$$R(A,t) = \frac{\sup\{L(S)\colon\ \phi(S)(t) = A(t), S \in \Theta\}}{L(S_n)}, \quad (28)$$

where Θ is the same as in (1). We show that an asymptotic $100(1-\alpha)\%$ confidence band for A_0 is given by

$$\mathcal{A} = \{A(t)\colon\ -2\log R(A,t) \le C^2(t), t \in [0,\tau]\}, \quad (29)$$

where $C(t)$ is defined by (9).

It is well known that S may be expressed as the product integral $S(t) = \pi_{0<s\le t}(1 - dA(s))$ (see, e.g., Andersen et al. 1993). Without constraint, $\hat{A}$ and $S_n(t) = \pi_{0<s\le t}(1 - d\hat{A}(s))$ are nonparametric maximum likelihood estimators (cf. Andersen et al. 1993, Chap. IV, Sec. 1.5). Murphy (1995) showed that the constraint alters $\hat{A}$ and thus $\hat{S}$ by changing the size of risk set: $Y(t)$ is replaced by $Y(t) + \lambda_n(t)$, where λ_n satisfies

$$\sum_{s\le t} \frac{\Delta N(s)}{Y(s) + \lambda_n} = A(t). \quad (30)$$

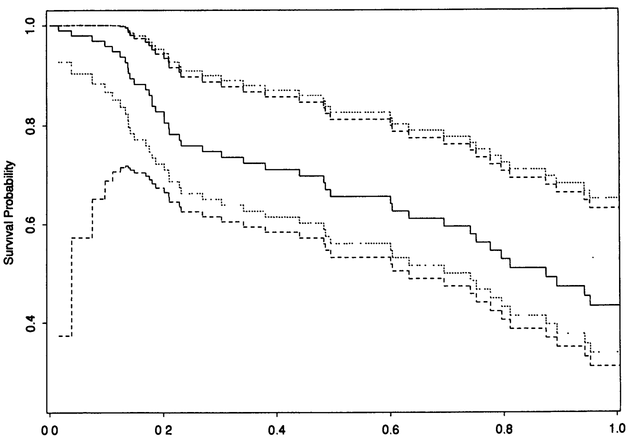

Figure 5. The Effect of the Bias Correction on LR_1; With 95% LR_1^c Band (Dotted Lines) and LR_1 Band (Dashed Lines).

By Taylor's expansion,

$$\frac{\Delta N(s)}{Y(s)+\lambda_n} = \frac{\Delta N(s)}{Y(s)}\left(1+\frac{\lambda_n}{Y(s)}+\cdots+\frac{\lambda_n^k}{Y^k(s)}+O_p(n^{-k/2})\right)$$

for $k \geq 1$, and it follows from (30) that

$$0 = K_A + \check{\sigma}^2\frac{\lambda_n}{n} - \check{\sigma}_1^2\frac{\lambda_n^2}{n^2}+\cdots+\check{\sigma}_4^2\frac{\lambda_n^5}{n^5}+O_p(n^{-3}), \quad (31)$$

where $K_A(A,t) = \hat{A}(t) - A(t)$ and

$$\check{\sigma}_r^2(t) = n^{r+1}\int_0^t \frac{I(Y>0)}{Y^{r+2}}\,dN, \qquad r \geq 0. \quad (32)$$

Here $\check{\sigma}_0^2 = \check{\sigma}^2$. Note that $\check{\sigma}_r^2$ is also a uniformly consistent estimator of σ_r^2 (cf. Andersen et al. 1993, Chap. IV, Sec. 1). Thus asymptotically, we can replace $\check{\sigma}_r$ by $\hat{\sigma}_r$. Furthermore, λ_n satisfies (A.1) with $K(S,t)$ replaced by K_A. As a result, the signed root of $-2\log R(A,t)$, say $W(A,t)$,

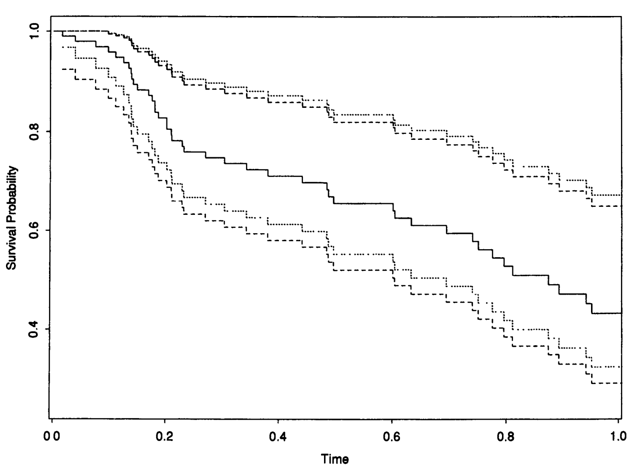

Figure 6. The Effect of the Bias Correction on LR_2; With 95% LR_2^c Band (Dotted Lines) and LR_2 Band (Dashed Lines).

has an asymptotic expansion (19), with $K(S,t)$ replaced by K_A. The asymptotic distributions of $\sqrt{n}K(S_0,t)$ and $\sqrt{n}K_A(A_0,t)$ agree (cf. Lem. 2.1), so the same is true of $W(A_0,t)$ and $W(S_0,t)$. It follows that

$$\mathcal{A} = \{A(t)\colon\ |W(A,t)| \leq C(t), t \in [0,\tau]\}$$

forms an asymptotic $100(1-\alpha)\%$ confidence band for A_0.

From the definition of $R(A,t)$ and (29),

$$\mathcal{A} = \{\phi(S)\colon\ S \in \mathcal{B}\} \quad (33)$$

where $\mathcal{B} = \{S\colon\ \mathcal{L}(S,t) \leq C^2(t), t \in [0,\tau]\}$. Of course, in (33) we can replace $\mathcal{B}$ by $\tilde{\mathcal{B}}$ to get a bias-corrected LR band for A_0. Note that $\mathcal{B} = \{S(t) = \pi_{s\leq t}(1-dA(s))\colon\ A \in \mathcal{A}\}$. As a result of (33), we can transform an LR confidence for S_0 into an LR band for A_0 or vice versa. In general, this transformation-preserving property does not hold for other confidence bands such as HW and EP bands. It holds for the LR bands because $W(S_0,t)$ and $W(A_0,t)$ are asymptotically equivalent and we use the LR approach instead of, say, inverting a goodness-of-fit test (see Nair 1984, sec. 3).

There is a connection between the LR bands for S_0 and general bands for A_0. To see this, assume that $S_0(t)$ is absolutely continuous, in which case $A_0(t) = -\log S_0(t)$. Using the Taylor expansion of $\log(1+x)$, it is easily shown that $\log S_n(t) = -\hat{A}(t) + O_p(n^{-1})$, where $\hat{A}(t) = \int_0^t Y(s)^{-1}\,dN(s)$ is the Nelson–Aalen estimator. It then follows from Theorem 2.1 that

$$\hat{\sigma}(t)W(S_0,t) = -\sqrt{n}(\hat{A}(t)-A_0(t)) + O_p(n^{-1/2}). \quad (34)$$

Bie et al. (1987) showed that for any continuous nonnegative function ψ on [0, τ],

$$\sup_{0\leq x\leq\tau}\sqrt{n}|\hat{A}(x)-A_0(x)|V(x) \xrightarrow{\mathcal{D}} \sup_{x\in[0,d]}|B^0(x)\psi(x)|, \quad (35)$$

where d is defined by (22) and $V(t) = \psi\{\hat{\sigma}^2(t)/[1+\hat{\sigma}^2(t)]\}\{1+\hat{\sigma}^2(t)\}^{-1}$. They pointed out that using (35) with different choices of $\psi(x)$ yields various types of confidence bands for A_0: A HW-type band is obtained by taking $\psi(x) \equiv 1$, and an EP-type band corresponds to $\psi(u) = [u(1-u)]^{-1/2}$ for $a \leq u \leq b$, and 0 otherwise.

Combining (34) and (35), we obtain

$$\sup_{0\leq t\leq\tau}|\hat{\sigma}(t)W(S_0,t)|V(t) \xrightarrow{\mathcal{D}} \sup_{x\in[0,d]}|B^0(x)\psi(x)|. \quad (36)$$

This allows us to construct various types of LR bands for A_0 and S_0 by replacing $C(t)$ in $\mathcal{A}$ and $\mathcal{B}$ by

$$C^*(t) = \frac{K'_{q,\alpha}(d)}{V(t)\hat{\sigma}(t)},$$

where $K'_{q,\alpha}(d)$ is the upper α quantile of the distribution of $\sup_{x\in[0,d]}|B^0(x)\psi(x)|$. The band LR_1 for S_0 given in Section 2.1 is based on $\psi(x) \equiv 1$. In the sequel we shall also consider the EP-type LR band for S_0 that uses $C^*(t) = e_\alpha(a,b)$; this band will be called LR_2. The LR_2 band has equal precision at each point of $S_0(t)$ in the sense that each interval forming the band has asymptotically equal probability that S_0 passes through it. We restrict attention to these two LR bands plus their corrected versions, LR_1^c and LR_2^c, as they are the easiest to implement.

Table 2. Observed Percentage Errors of the Nominal 95% LR, HW, and EP Bands Under Model I

		LR_1 bands		HW bands			LR_2 bands		EP bands		
θ	n	LR_1^c	LR_1	HW	HW_1	HW_a	LR_2^c	LR_2	EP	EP_1	EP_a
	25	5.52	4.58	4.04	6.16	3.24	4.04	3.48	12.58	7.34	3.8
∞	50	6.05	5.35	4.65	5.35	4.05	4.75	4.05	9.55	7.15	4.9
	100	6.25	6.0	5.3	5.85	4.8	4.8	4.4	7.85	6.45	5.7
	200	6.5	6.2	5.9	5.8	5.5	4.9	4.6	6.9	6.3	5.4
	25	4.7	3.88	4.18	5.76	3.76	4.1	3.44	11.46	7.26	4.12
3.72	50	5.44	4.82	4.78	5.82	4.82	4.54	4.28	9.32	6.98	4.48
	100	5.46	5.12	4.76	5.36	4.96	3.52	3.42	7.34	6.8	4.64
	200	5.6	5.7	6.1	6.4	5.9	4.1	3.85	5.7	5.50	4.80
	25	3.7	3.18	4.72	5.4	3.92	3.98	3.4	11.42	7.08	4.08
1.595	50	4.48	3.92	5.26	5.66	4.66	4.32	3.78	9.38	6.6	4.2
	100	4.5	4.4	4.5	5.05	4.7	4.45	4.45	6.65	6.6	3.9
	200	5.65	5.25	5.65	6.5	5.75	4.25	3.95	6.5	6.3	5.75
	25	1.5	1.46	3.92	4.88	1.74	3.06	2.56	11.08	6.62	3.3
.605	50	2.42	2.16	4.92	4.24	3.32	3.76	3.4	9.76	6.06	4.14
	100	3.1	2.84	4.8	4.76	4.38	4.18	4.08	6.9	6.18	3.94
	200	3.7	3.55	5.3	5.2	5	4.45	4.25	5.9	5.85	4.7

NOTE: $\theta = \infty$ corresponds to no censoring; θ = 3.72, to 25% censoring; θ = 1.595, to 50% censoring; and θ = .605, to 75% censoring.

4. EXAMPLES AND A SIMULATION STUDY

In this section we investigate the performance of our various LR confidence bands and apply them to a real data set. We compare LR_1, LR_1^c, LR_2, and LR_2^c to their six competitors: HW, HW_1, HW_a, EP, EP_1, and EP_a, where HW_1 denotes the log-log–transformed HW band and HW_a denotes the arcsin-transformed HW band. Each of the bands has an asymptotic coverage level of 95%.

4.1 Example 1

We analyzed data on 432 papers submitted to the Theory and Methods Section of JASA between January 1, 1994 and December 13, 1994; see Table 1. Each observation consists of the number of days between a manuscript's submission and its first review or the cut-off date, along with a censoring indicator (1 if a paper received its first review by the cut-off date; 0 otherwise). The data are tabulated columnwise in the order of the submission dates. There are 275 uncensored times and 157 censored times.

The bands in Figures 1–4 were produced in less than 5 seconds on a SUN Sparc 1. The HW band tends to be too wide in the lower and upper tails; its upper boundary may have values greater than 1, whereas its lower boundary may have values smaller than 0. Nair's EP band cannot avoid this shortcoming either. In his comparison of EP and HW, Nair (1984) reported that the former is narrower in the upper and lower tails, whereas the latter performs better in the middle of the distribution; this is also shown in Figure 1.

LR_1^c almost totally coincides with HW in the middle, between 50 and 150 days, but is narrower in the two tails (cf. Fig. 1a); it is strictly contained in the HW band after 150 days. Also, LR_1^c is contained within EP between 50 and 300 days, though the improvement is modest (cf. Fig. 1b). The LR_2^c band is narrower than HW in the tails, especially in the lower tail after 180 days, but is slightly wider than HW at times (cf. Fig. 1c). Finally, LR_2^c is very close to EP overall, but displays modest improvement in some regions (cf. Fig. 1d).

Due to the relatively large sample size in this example, the bias correction (cf. Figs. 4b and 4c) has almost no ef-

Table 3. Observed Percentage Errors of the Nominal 95% LR, HW, and EP Bands Under Model II

		LR_1 bands		HW bands			LR_2 bands		EP bands		
θ	n	LR_1^c	LR_1	HW	HW_1	HW_a	LR_2^c	LR_2	EP	EP_1	EP_a
	25	3.82	3.18	4.46	5.42	3.76	4.06	3.3	11.16	6.96	4.24
.5	50	4.5	3.98	4.88	5.72	4.58	4.12	4.04	9.06	6.76	4.18
	100	4.68	4.2	4.46	5.4	4.52	4.42	4.42	6.78	6.6	4.16
	200	5.9	5.7	5.5	6.2	5.55	4.5	4.5	5.9	6.2	5.6
	25	2.78	2.44	4.76	4.82	3.22	3.7	3.22	11.62	6.7	4.22
1.0	50	3.8	3.28	5.58	5.04	4.62	4.08	4.06	9.52	6.4	4.48
	100	4.32	3.98	4.88	5.12	4.78	4.44	4.42	7.02	6.08	4.26
	200	4.45	4.1	5.5	5.8	5.2	4.65	4.54	6.1	6.15	5.25
	25	2.2	2.18	2.74	4.84	1.22	2.42	2.12	9.92	6.46	2.78
2.0	50	2.82	2.52	4.12	4.02	2.48	3.32	3.02	9.4	6.14	3.74
	100	2.85	2.8	4.2	4.15	3.35	3.9	3.8	6.88	5.98	3.78
	200	3.75	3.65	4.85	4.5	4.45	4.45	4.1	6.1	6.15	5.25

NOTE: θ= .5 corresponds to 33% censoring; θ= 1.0, to 50% censoring; and θ= 2.0, to 66% censoring.

Table 4. Observed Percentage Errors of the Nominal 95% LR, HW, and EP Bands Under Model III

		LR_1 bands		HW bands			LR_2 bands		EP bands		
(θ_1, θ_2)	n	LR_1^c	LR_1	HW	HW_1	HW_a	LR_2^c	LR_2	EP	EP_1	EP_a
	25	1.92	1.46	4.46	5.1	2.42	3.16	2.38	10.44	6.18	3.26
(1.35, 2)	50	3.38	2.84	4.98	5.14	3.9	4.18	3.7	10.62	6.42	4.46
	100	3.8	3.2	4.3	4.8	3.95	4.5	4.2	7.45	5.65	4.2
	200	4.35	3.95	5.0	4.9	4.9	4.75	4.65	10.6	8.0	7.3
	25	2.42	2.04	3.88	5.18	2.98	3.66	2.84	10.58	6.68	3.74
$(\sqrt{2}, \frac{1}{2})$	50	3.56	3.24	4.62	5.52	4.2	3.82	3.66	8.58	6.72	3.98
	100	4.14	3.88	4.88	5.3	4.92	3.84	3.72	6.52	6.2	3.88
	200	4.85	4.55	5.5	5.9	5.45	4.3	4.1	6.05	5.65	4.95

NOTE: $(\theta_1, \theta_2) = (1.35, 2)$ corresponds to 50% censoring; $(\theta_1, \theta_2) = (\sqrt{2}, \frac{1}{2})$, to 35% censoring.

fect. Nor do the log-log and arcsin transformations, except to force the boundaries of the HW and EP bands to be contained within [0, 1] (cf. Figs. 2 and 3). However, the transformations worsen the nonmonotonicity of the HW band in its lower and upper tails (cf. Figs. 2a, 2c, 3a, and 3c). For this reason, and because very little improvement in coverage probability accuracy is achieved using the transformations (see Sec. 4.3), we would not recommend their application to the HW band in this example. The upper boundaries of the LR bands are closer to the HW and EP bands than are the lower boundaries (cf. Figs. 1–3). There is little difference between LR_1 and LR_2, although LR_2 is slightly narrower in the upper tail. The same can be said of the difference between LR_1^c and LR_2^c (cf. Figs. 4a and 4d).

4.2 Example 2

To illustrate the effect of the bias correction, we generated 100 observations using S_0 = standard exponential and S_C = uniform (0, .5), amounting to a 50% censoring rate.

The bias corrections for the LR_1 and LR_2 bands over the range [0, 1] are shown in Figures 5 and 6.

The bias correction significantly reduces the widths of both LR_1 and LR_2, especially in the lower tail of LR_1. Also note that the lower boundaries are adjusted more than the upper boundaries.

4.3 Simulations

We now report the results of a simulation study to determine the error rates of our LR bands and their competitors. The data were generated using the following models:

I. S_0 = standard exponential, S_C = uniform $(0, \theta)$

II. S_0 = standard exponential, S_C = exponential with parameter θ

III. $S_0(t) = e^{-\theta_1 t^{\theta_2}}$ (Weibull), S_C = standard exponential.

The terminal point τ was adjusted so that the effective sample size, $n_r = \#\{i\colon T_i \geq \tau\}$ at τ, was at least 10% of the total sample size; this was done to avoid instability in the bands at large τ. Each observed error rate was based on 5,000 samples, and each of the confidence bands had nominal 95% coverage. From tables of Hall and Wellner (1980), the asymptotic 95% critical level needed for the LR_1, LR_1^c, HW, HW_1, and HW_a bands is 1.358, and that for the LR_2, LR_2^c, EP, EP_1, and EP_a bands is 3.31. Note that LR_2, LR_2^c, and Nair's EP band are valid only for $\{t \in [0,\tau] : a \leq \{\hat{\sigma}^2(t)/[1+\hat{\sigma}^2(t)]\} \leq b\}$. We chose $a = 1 - b = .05$. The results are reported in Tables 2–4.

The simulation results in Tables 2–4 clearly indicate that $LR_i^c, i = 1, 2$, outperform EP to a great extent. This is not unexpected, because TG originally showed that the LR confidence intervals are more accurate than those based on the simple normal approximation to the distribution of the Kaplan–Meier estimator. Nair's EP band is a natural extension of the intervals based on the normal approximation, whereas our LR bands are based on the LR method. In terms of coverage accuracy, LR_i, LR_i^c $(i = 1, 2)$, HW, HW_1, and HW_a perform quite well. As shown by Nair (1984), the nontransformed EP band gives excessive error rates at small sample sizes. EP's large coverage error can be substantially reduced by taking the arcsin or the log-log transformation, as shown by Borgan and Liestøl (1990). These authors also showed, however, that these two transformations bring very little improvement to the HW bands. This is again confirmed by our simulation results.

Note that under no censoring,

$$C(t) = K_{q,\alpha}(d)\,\frac{1+\hat{\sigma}^2(t)}{\hat{\sigma}(t)} = \frac{K_{q,\alpha}(d)}{(S_n F_n)^{1/2}},$$

which converges to $K_{q,\alpha}(d)(SF)^{-1/2}$. Therefore, LR_1^c and LR_1 are valid only for $a \leq F_n(t) \leq b$, a similar condition required by the Nair's EP band. According to Nair's simulation results, a and $1 - b$ should be chosen to be .05 or .1 for the bands to perform well. We already imposed this condition while choosing $a = 1 - b = .05$. If we choose $a = 1 - b = .1$, then the coverage error will be further reduced.

The LR bands perform consistently well in Models I–III. Yet note that LR_2 and LR_2^c are less affected by the model than LR_1 and LR_1^c. The LR and HW bands are less affected by the censoring than the EP bands.

The simulation results indicate that the LR bands are slightly conservative. This is not always the case, however—when τ is large, the LR bands tend to have smaller than nominal coverage probabilities. To see why this can happen, recall that the LR bands have relatively narrow tails, so, unlike the HW bands, their coverage accuracy is sensitive to changes in the terminal point τ; they become less reliable as τ increases. In the simulation study, the LR bands tended to be slightly more conservative than the HW bands due to the strong truncation of the tails and because the LR bands are a little wider in the middle of the distribution. (See Fig. 1c for an instance of this.)

5. CONCLUSIONS

The LR method provides an appealing approach to constructing confidence bands for survival functions. The LR bands perform reasonably well in most cases. We recommend their use in view of (a) their adequate coverage accuracy, (b) the desirable feature that their boundaries are naturally contained in [0, 1], and (c) their transformation-preserving property. In particular, we recommend LR_2^c for small sample sizes and LR_2 for medium and large sample sizes.

APPENDIX A: PROOFS

Lemma A.1 provides an approximation for the solution of (15), which is a first step in finding a suitable approximation for $\mathcal{L}(S,t)$.

Lemma A.1. The solution of Equation (15) has the asymptotic expansion

$$\lambda_n(S,t) = -\frac{n}{\hat\sigma^2(t)}\,K(S,t) + \frac{n\hat\sigma_1^2(t)}{\hat\sigma^6(t)}\,K^2(S,t) + \frac{n}{\hat\sigma^{10}(t)}\,\{3\hat\sigma_2^2(t)\hat\sigma^2(t) - 2\hat\sigma_1^4(t)\}K^3(S,t) + O_p(n^{-1}), \tag{A.1}$$

where $K(S,t)$ is defined by (18).

Proof of Lemma A.1

By (15), λ_n satisfies

$$\sum_{s\le t}\log\left\{1-\frac{\Delta N(s)}{Y(s)+\lambda_n}\right\} = \log S(t). \tag{A.2}$$

Li (1995a) has shown that $\lambda_n = O_p(n^{1/2})$. In fact, this order of λ_n holds uniformly for $t\in[0,\tau]$ due to (A.7), (A.8), and the uniform consistency of $\hat\sigma_\tau^2(t)$. For the same reason, the various other remainder terms in this article have their specified orders uniformly in $t\in[0,\tau]$. Using the Taylor expansions of $\log(1+x)$ and $1/(1+x)$, we can then develop an expansion of

$$\log\left\{1-\frac{\Delta N(s)}{Y(s)+\lambda_n}\right\},$$

which we rewrite as

$$\log\left\{1-\frac{\Delta N(s)}{Y(s)}\left(1+\frac{\lambda_n}{Y(s)}\right)^{-1}\right\}$$
$$= \log\left\{1-\frac{\Delta N(s)}{Y(s)}\left(1-\frac{\lambda_n}{Y(s)}+\frac{\lambda_n^2}{Y^2(s)}-O_p(n^{-3/2})\right)\right\}$$
$$= \log\left(1-\frac{\Delta N(s)}{Y(s)}\right) + \log\left\{1+\left(1-\frac{\Delta N(s)}{Y(s)}\right)^{-1} \times\left(\frac{\lambda_n\Delta N(s)}{Y^2(s)}-\frac{\lambda_n^2\Delta N(s)}{Y^3(s)}+O_p(n^{-5/2})\right)\right\}.$$

Here we have used $n/Y(s) = O_p(1)$ uniformly for $s\in[0,\tau]$, which is implied by the weak law of large numbers and the monotonicity of $Y(s)$. It then follows from (A.2) that

$$\log S(t) = \log S_n(t) + \lambda_n\sum_{s\le t}\frac{\Delta N(s)}{Y(s)(Y(s)-\Delta N(s))} - \lambda_n^2\sum_{s\le t}\frac{\Delta N(s)}{Y^2(s)(Y(s)-\Delta N(s))} + \cdots + \lambda_n^5\sum_{s\le t}\frac{\Delta N(s)}{Y^5(s)(Y(s)-\Delta N(s))} + O_p(n^{-3}).$$

Equivalently,

$$0 = K + \hat\sigma^2\,\frac{\lambda_n}{n} - \hat\sigma_1^2\,\frac{\lambda_n^2}{n^2} + \cdots + \hat\sigma_4^2\,\frac{\lambda_n^5}{n^5} + O_p(n^{-3}), \tag{A.3}$$

where we have suppressed t in $\hat\sigma_\tau(t)$. It is difficult to obtain a closed-form expression for λ_n, the solution of this whole equation. Instead, we find an expansion of λ_n term by term. We begin by isolating the lowest order of λ_n, noting that the coefficients have decreasing orders as the powers of λ_n increase. From the equation

$$K + \hat\sigma^2\,\frac{\lambda_n}{n} + O_p(n^{-1}) = 0,$$

we easily obtain

$$\lambda_n = -\frac{nK}{\hat\sigma^2} + o_p(n^{1/2}).$$

Because $nK/\hat\sigma^2$ is of order $O_p(n^{1/2})$, we let $\lambda_n' = -n(K/\hat\sigma^2) + G$, where G is $o_p(n^{1/2})$. Then we substitute this λ_n' into (A.3). Combining higher-order terms, we obtain

$$\left[\frac{\hat\sigma^2}{n} + O_p(n^{-3/2})\right]G = \frac{\hat\sigma_1^2}{\hat\sigma^4}\,K^2 + O_p(n^{-3/2}),$$

which implies $G = (n\hat\sigma_1^2/\hat\sigma^6)K^2 + O_p(n^{-1/2})$. This enables us to obtain the second term in the expansion of λ_n, so that

$$\lambda_n = -\frac{nK}{\hat\sigma^2} + \frac{n\hat\sigma_1^2}{\hat\sigma^6}\,K^2 + O_p(n^{-1/2}).$$

Repeating this iteration will give (A.1).

Proof of Lemma 2.1

Let $\varphi(S)(t) = \log S(t)$. We apply the functional delta method, which essentially allows us to approximate $\sqrt{n}(\varphi(S_n)-\varphi(S_0))$ by $d\varphi(S_0)(\sqrt{n}(S_n-S_0))$ (cf. Andersen et al. 1993, thm. II.8.1), where $d\varphi$, the Hadamard derivative of φ, is pointwise multiplication by $-1/S$. The well-known weak convergence result

$$\sqrt{n}(S_n-S_0) \xrightarrow{\mathcal{D}} -S_0U \tag{A.4}$$

(see, e.g., Andersen et al. 1993, thm. IV.3.2) then gives $\sqrt{n}K(S_0,t) \xrightarrow{\mathcal{D}} U(t)$, as required.

Before proceeding to prove Theorem 2.1, we record the following expansion of $\mathcal{L}(S,t)$ (cf. proof of Lem. A.1):

$$\mathcal{L}(S,t) = \lambda_n^2 \sum_{s\le t} \frac{\Delta N(s)}{Y(s)(Y(s)-\Delta N(s))}$$
$$-\frac{2\lambda_n^3}{3}\sum_{s\le t}\left\{\left(\frac{1}{Y(s)-\Delta N(s)}\right)^2-\left(\frac{1}{Y(s)}\right)^2\right\}$$
$$+\frac{\lambda_n^4}{2}\sum_{s\le t}\left\{\left(\frac{1}{Y(s)-\Delta N(s)}\right)^3-\left(\frac{1}{Y(s)}\right)^3\right\}$$
$$+O_p(n^{-3/2})$$
$$= \hat\sigma^2\frac{\lambda_n^2}{n}-\frac{4}{3}\hat\sigma_1^2\frac{\lambda_n^3}{n^2}+\frac{3}{2}\hat\sigma_2^2\frac{\lambda_n^4}{n^3}+O_p(n^{-3/2}) \quad \text{(A.5)}$$

Substituting (A.1) into (A.5), we obtain, after some lengthy calculations,

$$\mathcal{L}(S,t) = \frac{nK^2}{\hat\sigma^2}-\frac{2}{3}\frac{n\hat\sigma_1^2K^3}{\hat\sigma^6}$$
$$+\frac{1}{6}\frac{n(6\hat\sigma_1^4-3\hat\sigma^2\hat\sigma_2^2)K^4}{\hat\sigma^{10}}$$
$$+\frac{1}{6}\frac{n(6\hat\sigma_1^2\hat\sigma_2^2-3\hat\sigma^2\hat\sigma_3^2)K^5}{\hat\sigma^{12}}+O_p(n^{-2}). \quad \text{(A.6)}$$

Proof of Theorem 2.1

We expand $W = \mathrm{sgn}(S_n(t)-S(t))\sqrt{\mathcal{L}(S,t)}$ term by term using (A.6). We begin by isolating the first term of $\mathcal{L}(S,t), nK^2/\hat\sigma^2$, and let $W = \sqrt{n}K/\hat\sigma$. Then let

$$W = \frac{\sqrt{n}K}{\hat\sigma}\,(1+G)$$

where G is small, and form the equation $W^2 = \mathcal{L}(S,t)$; that is,

$$n\,\frac{K^2}{\hat\sigma^2}\,(1+2G+G^2) = \mathcal{L}(S,t).$$

The unknown of lowest order on the left side is $2n(K^2/\hat\sigma^2)G$, which we equate to the term of lowest order on the right side of (A.6) (except for $nK^2/\hat\sigma^2$ itself). This enables us to obtain $-\frac{1}{3}(\hat\sigma_1^2K/\hat\sigma^4)$, a first-order approximation to G. Then repeating this process, replacing W by $(\sqrt{n}K/\hat\sigma)-(\sqrt{n}/3)\,(\hat\sigma_1^2K^2/\hat\sigma^5)$, to get the term of next highest order, yields the expansion of W.

Proof of Theorem 2.2

It follows from the uniform consistency of the Kaplan–Meier estimator (cf. Andersen et al. 1993, thm. IV.3.1) that

$$\sup_{s\in[0,\tau]} |\log S_n(s)-\log S_0(s)| \xrightarrow{P} 0. \quad \text{(A.7)}$$

The inequality

$$|\sqrt{n}\{\log S_n(s)-\log S_0(s)\}^2| \le \sup_{s\in[0,\tau]} |\log S_n(s)-\log S_0(s)||\sqrt{n}[\log S_n(s)-\log S_0(s)]|$$

then gives

$$\sup_{s\in[0,\tau]} |\sqrt{n}\{\log S_n(s)-\log S_0(s)\}^2| \xrightarrow{P} 0, \quad \text{(A.8)}$$

by Lemma 2.1. Combining (A.8), (20), Lemma 2.1, and the uniform consistency of $\hat\sigma^2$ and $\hat\sigma_1^2$ establishes the result.

[Received April 1995. Revised January 1996.]

REFERENCES

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

Bie, O., Borgan, Ø., and Liestøl, K. (1987), "Confidence Intervals and Confidence Bands for the Cumulative Hazard Rate Function and Their Small Sample Properties," *Scandinavian Journal of Statistics*, 14, 221–233.

Borgan, Ø., and Liestøl, K. (1990), "A Note on Confidence Intervals and Bands for the Survival Function Based on Transformations," *Scandinavian Journal of Statistics*, 17, 35–42.

DiCiccio, T. J., and Romano, J. P. (1989), "On Adjustments Based on the Signed Root of the Empirical Likelihood Ratio Statistic," *Biometrika*, 76, 447–456.

Efron, B. (1981), "Censored Data and the Bootstrap," *Journal of the American Statistical Association*, 76, 312–319.

Gill, R. D. (1980), *Censoring and Stochastic Integrals*, Mathematical Center Tracts 124. Amsterdam: Mathematisch Centrum.

Hall, P., and La Scala, B. (1990), "Methodology and Algorithms of Empirical Likelihood," *International Statistical Review*, 58, 109–127.

Hall, W. J., and Wellner, J. A. (1980), "Confidence Bands for a Survival Curve From Censored Data," *Biometrika*, 67, 133–143.

Kalbfleisch, J. D., and Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data*, New York: Wiley.

Li, G. (1995a), "On Nonparametric Likelihood Ratio Estimation of Survival Probabilities for Censored Data," *Statistics and Probability Letters*, 25, 95–104.

——— (1995b), "Nonparametric Likelihood Ratio Estimation of Probabilities for Truncated Data," *Journal of the American Statistical Association*, 90, 997–1003.

Li, G., Hollander, M., McKeague, I. W., and Yang, J. (1996), "Nonparametric Likelihood Ratio Confidence Bands for Quantile Functions From Incomplete Survival Data," *The Annals of Statistics*, 24, 628–640.

Murphy, S. (1995), "Likelihood Ratio–Based Confidence Intervals in Survival Analysis," Journal of the American Statistical Association, 90, 1399–1405.

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.

——— (1990), "Empirical Likelihood Ratio Confidence Regions," *The Annals of Statistics*, 18, 90–120.

Thomas, D. R., and Grunkemeier, G. L. (1975), "Confidence Interval Estimation of Survival Probabilities for Censored Data," *Journal of the American Statistical Association*, 70, 865–871.