I N S T I T U T D E

S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N

P A P E R

0129

LIKELIHOOD RATIO CONFIDENCE BANDS IN NONPARAMETRIC REGRESSION WITH CENSORED DATA

G. LI and I. VAN KEILEGOM

http://www.stat.ucl.ac.be

Likelihood ratio confidence bands in nonparametric regression with censored data

Gang Li*
University of California at Los Angeles
Department of Biostatistics

Ingrid Van Keilegom†
Université catholique de Louvain
Institute of Statistics

July 17, 2001

Abstract

Let (X, Y) be a random vector, where Y denotes the variable of interest possibly subject to random right censoring, and X is a covariate. We construct confidence intervals and bands for the conditional survival and quantile function of Y given X using a nonparametric likelihood ratio approach. This approach was introduced by Thomas and Grunkemeier (1975), who estimated confidence intervals of survival probabilities based on right censored data. The method is appealing for several reasons: it always produces intervals inside $[0, 1]$, it does not involve variance estimation, and can produce asymmetric intervals. Asymptotic results for the confidence intervals and bands are obtained, as well as simulation results, in which the performance of the likelihood ratio intervals and bands is compared with that of the normal approximation method. We also propose a bandwidth selection procedure based on the bootstrap and apply the technique on a real data set.

KEY WORDS: Beran estimator; Confidence band; Confidence interval; Empirical likelihood; Likelihood ratio; Nonparametric regression; Right censoring.

*Gang Li's research was supported by NIH grant CA78314-03

†Ingrid Van Keilegom's research was supported by 'Projet d'Actions de Recherche Concertées', No. 98/03–217, from the Belgian government

1 Introduction

Conditional survival and quantile functions are useful tools for assessing the influence of risk factors on survival. Let Y_i, C_i and X_i denote the survival time, censoring time and covariate, respectively, of the ith subject under study, $1 \leq i \leq n$. Assume that one observes n i.i.d. triples $(T_1, \Delta_1, X_1), \ldots, (T_n, \Delta_n, X_n)$ where, for each i, $T_i = \min(Y_i, C_i)$ and $\Delta_i = I(Y_i \leq C_i)$ and assume that Y_i is independent of C_i for given X_i. This paper derives nonparametric likelihood ratio-based confidence intervals and bands for the conditional survival function $S(t|x) = P(Y_i > t|X_i = x)$ and the conditional quantile function $S^{-1}(p|x)$, for fixed x, on the basis of right censored regression data.

Nonparametric estimation of $S(t|x)$ and related applications have been studied by a number of authors; see, among others, Beran (1981), Dabrowska (1987, 1992), McKeague and Utikal (1990), Akritas (1994), Li and Doss (1995), McKeague et al. (1995), McKeague and Sun (1996), and Li (1997) for random design regression, and González Manteiga and Cadarso Suarez (1994), and Van Keilegom and Veraverbeke (1997, 1998) for fixed design regression. Confidence intervals and bands for $S(t|x)$ or $S^{-1}(p|x)$ can be obtained based on normal approximations as discussed later in Section 2 (Remark 2.3). However, normal confidence intervals for $S(t|x)$ could include impossible values outside the range [0,1]. Their small sample properties are also not satisfactory as shown in Section 3. This is similar to the pitfalls of the Kaplan-Meier method for homogeneous data (with no covariate) as discussed by Hollander et al. (1997). Construction of normal confidence intervals for $S^{-1}(p|x)$ is also problematic because it requires density estimation in order to estimate the variance and the resulting variance estimate is not stable.

In the absence of covariates, Thomas and Grunkemeier (1975) introduced a nonparametric likelihood ratio method to construct confidence intervals for a survival probability from right censored data. The resulting confidence intervals never include values outside [0,1] and showed better small sample performance than the Kaplan-Meier normal approximation method in simulations.

In this paper, we extend the method of Thomas and Grunkemeier (1975) to the nonparametric regression setting to obtain confidence intervals and bands for $S(t|x)$ and $S^{-1}(p|x)$. To explain the idea, we begin with a nonparametric likelihood ratio test for the null hypothesis $H_0 : S(t|x) = p$ where $0 < p < 1$. Define a weighted likelihood ratio for

H_0 by

$$R(p, t|x) = \frac{\sup\{L(S(\cdot|x))|S(t|x) = p,\ S(\cdot|x) \in \Theta\}}{\sup\{L(S(\cdot|x))|S(\cdot|x) \in \Theta\}}, \tag{1.1}$$

where $L(S(\cdot|x)) = \prod\{[S(T_i - |x) - S(T_i|x)]^{\Delta_i}[S(T_i|x)]^{1-\Delta_i}\}^{W_i}$ is a weighted likelihood function, Θ is the space of all survival functions supported on $(0, \infty)$, and the weight functions $\{W_i\}$ are defined in the next section. Here we only mention that the contribution of individual i to the likelihood of $S(\cdot|x)$ is determined by the weight W_i and that the $\{W_i\}$ assign heavy weights to individuals whose covariate values are close to x and vice versa. Clearly large values of $R(t|x)$ provide evidence in favor of H_0. We propose to form a confidence interval for $S(t|x)$ for a given t as the set of all values p for which the H_0 is not rejected by a likelihood ratio test based on $R(p, t|x)$. Similarly, for a given p, a confidence interval for $S^{-1}(p|x)$ is defined as the set of all values t for which the H_0 is not rejected. Confidence bands can be obtained by pasting together an appropriate set of confidence intervals. We establish large sample properties of the nonparametric likelihood ratio $R(p, t|x)$ and show how to determine critical values for the likelihood ratio confidence intervals and bands. The nonparametric likelihood ratio approach is appealing for several reasons. For example, it respects the range of the parameter. It uses data to determine the shape of the confidence intervals and can, consequently, produce asymmetric intervals to reflect possible skewness in the distribution of the estimated parameter. It does not involve variance estimation which is especially important for quantile interval estimation. It also leads to narrower confidence intervals and bands than the normal approximation method as shown in our simulation study.

For simplicity, we only considered kernel weight functions for the weighted likelihood function $L(\cdot)$ in (1.1) although other types of weight functions could also be used. A related issue concerns the selection of the bandwidth in the kernel weight function. It is worth noting that, unlike in estimation problems where the bandwidth is often chosen as the value for which the mean (integrated) squared error is minimal, a different criterion should be used for the construction of confidence intervals and bands. In this paper we consider the bandwidth for which the coverage error of the interval or band is minimal. A similar idea has been used by Hall (1991, 1992) in the context of density estimation. Because the coverage error is generally unknown for a given bandwidth, a bootstrap method is used to obtain an approximate optimal bandwidth. More detailed discussion of bandwidth selection will be given in Remark 2.2 of Section 2.

Although its use dates back at least to Thomas and Grunkemeier (1975), rigorous treatment of the nonparametric likelihood ratio method was first given in a different problem by Owen (1988, 1990), who studied the now well known empirical likelihood method to construct nonparametric confidence intervals for the mean of a random vector based on i.i.d. complete data. Since Owen's (1988, 1990) work, the likelihood ratio approach has been extended to a wide range of nonparametric and semiparametric problems, including linear models (Owen (1991) and Chen (1993, 1994)), generalized linear models (Kolaczyk (1994)), quantile estimation (Chen and Hall (1993)), biased sample models (Qin (1993)), generalized estimating equations (Qin and Lawless (1994)), confidence intervals and bands for unconditional survival and quantile functions (Li (1995), Murphy (1995), Owen (1995), Li, Hollander, McKeague and Yang (1996) and Hollander, McKeague and Yang (1997)), dependent process models (Kitamura (1997)), partial linear models (Wang and Jing (1999)), confidence tubes for multiple quantile plots (Einmahl and McKeague (1999)), and nonparametric regression (Chen and Qin (2000)).

In Section 2 we study the asymptotic properties of $R(p, t|x)$ and discuss how to compute the resulting likelihood ratio confidence intervals and bands for $S(t|x)$ and $S^{-1}(p|x)$. We also include a discussion on how to choose the bandwidth sequence. Section 3 presents a simulation study to study the performance of the developed likelihood ratio method compared with the normal approximation method. An illustration is given in Section 4 using data on cancer of the larynx. All proofs are collected in the Appendix.

2 Main results

Let $(X_i, Y_i, C_i, T_i, \Delta_i)$, $i = 1, \ldots, n$, be n independent random vectors as defined in Section 1 and let (X, Y, C, T, Δ) have the joint distribution of each $(X_i, Y_i, C_i, T_i, \Delta_i)$. Let $F(t|x) = P(Y \leq t|x)$, $S(t|x) = P(Y > t|x)$, $G(t|x) = P(C \leq t|x)$, $H(t|x) = P(T \leq t|x)$, $H^u(t|x) = P(T \leq t, \Delta = 1|x)$ and $F_X(x) = P(X \leq x)$. The assumed independence of Y and C for given X implies that $1-H(t|x) = (1-F(t|x))(1-G(t|x))$. The probability density functions of the distributions defined above will be denoted with lower case letters.

We define the local (or weighted) log likelihood by

$$\log L(S(\cdot|x)) \tag{2.1}$$

$$= \sum_{i=1}^{n} W_i(x;h_n)\{\Delta_i \log[S(T_i - |x) - S(T_i|x)] + (1-\Delta_i)\log S(T_i|x)\},$$

where

$$W_i(x;h_n) = \frac{K\left(\frac{x-X_i}{h_n}\right)}{\sum_{j=1}^{n} K\left(\frac{x-X_j}{h_n}\right)}$$

are Nadaraya-Watson weights, K is a known density function (kernel) and h_n is a sequence of positive constants tending to zero as $n \to \infty$ (bandwidth). It can be shown that, in the absence of ties, the local likelihood function is maximal at

$$\hat{S}(t|x) = \prod_{T_i \le t, \Delta_i = 1} \left\{1 - \frac{W_i(x;h_n)}{\sum_{j=1}^{n} W_j(x;h_n) I(T_j \ge T_i)}\right\},$$

which is the estimator proposed by Beran (1981). We refer to Section 1 for references on this estimator and related applications. The estimator reduces to the Kaplan-Meier (1958) estimator when no covariates are present (i.e. $W_i(x;h_n) = n^{-1}$) and to the estimator of Stone (1977) when there is no censoring (i.e. $T_i = Y_i, \Delta_i = 1$). The proof that $\hat{S}(t|x)$ maximizes the likelihood function is very similar to the proof given in Kaplan and Meier (1958).

The idea behind constructing likelihood ratio confidence intervals for an unknown probability $S(t|x)$ is to include all values p in the interval for which the test for $H_0 : S(t|x) = p$ based on the weighted likelihood ratio statistic $R(p,t|x)$ defined in (1.1) is not rejected. Since large values of $R(p,t|x)$ provide evidence in favor of H_0, define for $t > 0$ and $0 < r < 1$,

$$C(t,r|x) = \{p : R(p,t|x) \ge r\}.$$

We will show that for an appropriate choice of r, $C(t,r|x)$ will be a $100(1-\alpha)\%$ confidence interval for $S(t|x)$. Analogously, a confidence interval for $S^{-1}(p|x)$ $(0 < p < 1)$ will be of the form

$$Q(p,r|x) = \{t : R(p,t|x) \ge r\}$$

for some $0 < r < 1$. For the statement of the asymptotic results below, we need to introduce the following estimators :

$$\hat{f}_X(x) = (nh_n)^{-1} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)$$

$$\hat{H}(t|x) = \sum_{i=1}^{n} W_i(x;h_n) I(T_i \leq t)$$

$$\hat{H}^u(t|x) = \sum_{i=1}^{n} W_i(x;h_n) I(T_i \leq t, \Delta_i = 1).$$

Also, let $\tau_x = \inf\{t : H(t|x) = 1\}$ and define

$$\sigma^2(t|x) = \int K^2(u)\,du\, f_X^{-1}(x) \int_0^t \frac{dH^u(s|x)}{(1 - H(s|x))^2},$$

which is the asymptotic variance of the cumulative hazard function of Y given $X = x$, if $H(\cdot|x)$ is continuous (see e.g. Li and Doss (1995)). This function is estimated by

$$\hat{\sigma}^2(t|x) = \int K^2(u)\,du\, \hat{f}_X^{-1}(x) \int_0^t \frac{d\hat{H}^u(s|x)}{(1 - \hat{H}(s|x))(1 - \hat{H}(s - |x))}.$$

Theorem 2.1 (Confidence interval for $\boldsymbol{S(t|x)}$) *Assume $nh_n^5 \to 0$ and $nh_n \to \infty$. Assume further that the regularity conditions (R1)-(R3) of Li and Doss (1995, Section 4) hold and that K is a bounded density function, symmetric around zero. Let $0 < t < \tau_x$ be such that $S(t|x) < 1$.*

(a) Then,

$$-2nh_n \frac{\hat{f}_X(x)}{\int K^2(u)\,du} \log R(S(t|x), t|x) \xrightarrow{d} \chi_1^2.$$

(b) Let

$$r_\alpha = \exp\left\{ -\frac{\int K^2(u)\,du}{2nh_n \hat{f}_X(x)} \chi_1^2(\alpha) \right\},$$

where $\chi_1^2(\alpha)$ is the $100(1-\alpha)$th percentile of the χ_1^2 distribution. Then,

$$\lim_{n\to\infty} P\{S(t|x) \in C(t, r_\alpha|x)\} = 1 - \alpha.$$

(c) A $100(1-\alpha)\%$ confidence interval for $S(t|x)$ is given by

$$C(t, r_\alpha|x) = \{p : R(p, t|x) \geq r_\alpha\}.$$

The proof can be found in the Appendix.

Let $e_\alpha(u_1, u_2)$ $(u_1 \leq u \leq u_2)$ be the $100(1-\alpha)$th percentile of the distribution of

$$\sup_{u_1 \leq u \leq u_2} \left| \frac{B^0(u)}{\sqrt{u(1-u)}} \right|$$

where $B^0(u)$ is a Brownian bridge on $[0,1]$. Some specific percentiles of this distribution can be found in Nair (1984).

Theorem 2.2 (Confidence band for $S(\cdot|x)$) *Assume that the conditions of Theorem 2.1 hold. Let $0 < t_1 < t_2 < \tau_x$ be such that $S(t_1|x) < 1$.*

(a) The process

$$-2nh_n \frac{\hat{f}_X(x)}{\int K^2(u)\,du} \log R(S(t|x), t|x)$$

$(t_1 \le t \le t_2)$ converges weakly to the process $\left\{\frac{B^0(u)}{\sqrt{u(1-u)}}\right\}^2$, where $u = \sigma^2(t|x)/(1+\sigma^2(t|x))$.

(b) Let $\hat{u}_j = \hat{\sigma}^2(t_j|x)/(1+\hat{\sigma}^2(t_j|x))$ $(j = 1,2)$ and

$$r_\alpha(\hat{u}_1, \hat{u}_2) = \exp\left\{-\frac{\int K^2(u)\,du\, e_\alpha^2(\hat{u}_1, \hat{u}_2)}{2nh_n \hat{f}_X(x)}\right\}.$$

Then,

$$\lim_{n\to\infty} P\{S(t|x) \in C(t, r_\alpha(\hat{u}_1, \hat{u}_2)|x) \text{ for all } t_1 \le t \le t_2\} = 1-\alpha.$$

(c) A $100(1-\alpha)\%$ confidence band for $S(t|x)$ $(t_1 \le t \le t_2)$ is given by

$$C(t, r_\alpha(\hat{u}_1, \hat{u}_2)|x) \quad (t_1 \le t \le t_2).$$

The proof can be found in the Appendix.

Theorem 2.3 (Confidence interval for $S^{-1}(p|x)$) *Assume that the conditions of Theorem 2.1 hold. Let $0 < p < 1$ be such that $G(S^{-1}(p|x)|x) < 1$ and assume that $S(\cdot|x)$ is strictly decreasing in a neighborhood of $S^{-1}(p|x)$.*

(a) Then,

$$\lim_{n\to\infty} P\{S^{-1}(p|x) \in Q(p, r_\alpha|x)\} = 1-\alpha.$$

(b) A $100(1-\alpha)\%$ confidence interval for $S^{-1}(p|x)$ is given by

$$Q(p, r_\alpha|x) = \{t : R(p,t|x) \ge r_\alpha\}.$$

The result follows readily from part (a) of Theorem 2.1.

Theorem 2.4 (Confidence band for $S^{-1}(\cdot|x)$) *Assume that the conditions of Theorem 2.1 hold. Let $0 < p_1 < p_2 < 1$ be such that $G(S^{-1}(p_1|x)|x) < 1$ and assume that $S(\cdot|x)$ is strictly decreasing on $[S^{-1}(p_2|x), S^{-1}(p_1|x)]$.*

(a) Let $\hat{v}_j = \hat{\sigma}^2(\hat{S}^{-1}(p_j|x)|x)/(1+\hat{\sigma}^2(\hat{S}^{-1}(p_j|x)|x))$ $(j=1,2)$. Then,

$$\lim_{n\to\infty} P\{S^{-1}(p|x) \in Q(p, r_\alpha(\hat{v}_2, \hat{v}_1)|x) \text{ for all } p_1 \le p \le p_2\} = 1-\alpha.$$

(b) A $100(1-\alpha)\%$ confidence band for $S^{-1}(p|x)$ $(p_1 \le p \le p_2)$ is given by

$$Q(t, r_\alpha(\hat{v}_2, \hat{v}_1)|x) \quad (p_1 \le p \le p_2).$$

The proof is very similar to that of Theorem 2.2 and will therefore be omitted.

Remark 2.1. [**Computation of the confidence intervals and bands**] We first describe how to compute $C(t,r|x)$. It follows from Lemma A.2 that

$$R(p,t|x) = \phi(\lambda_n(p,t|x)),$$

where

$$\phi(u) = \prod_{T_i \le t, \Delta_i = 1} \left(\frac{R_i}{R_i+u}\right)^{R_i} \left(\frac{R_i - W_i + u}{R_i - W_i}\right)^{R_i - W_i},$$

and W_i, R_i and λ_n are defined by (A.1), (A.2) and (A.3) in the Apppendix. Let $D = \max_{T_i \le t, \Delta_i = 1, W_i > 0} (W_i - R_i)$. It can be verified that $\phi(u)$ is strictly increasing on $(D, 0]$, strictly decreasing on $[0,\infty)$, $\lim_{x \searrow D} \phi(x) = 0$, $\lim_{x\to+\infty} \phi(x) = 0$, and $\phi(0) = 1$. Furthermore, $\lambda_n(p,t|x)$ is continuous and strictly increasing in p. Therefore,

$$C(t,r|x) = \{p: \ R(p,t|x) \ge r\} = [p_L(t), p_U(t)],$$

where

$$p_L(t) = \prod_{T_i \le t, \Delta_i = 1} \left(1 - \frac{W_i}{R_i + \lambda_L(t)}\right),$$

$$p_U(t) = \prod_{T_i \le t, \Delta_i = 1} \left(1 - \frac{W_i}{R_i + \lambda_U(t)}\right),$$

and $\lambda_L(t)$ and $\lambda_U(t)$ are the unique solutions of the equation

$$\phi(\lambda) = r, \quad (0 < r < 1)$$

on the intervals $(D, 0)$ and $(0, \infty)$, respectively.

Now we discuss how to compute $Q(p, r|x)$. Let $t_1 < \ldots < t_k$ denote the distinct and ordered uncensored times for all individuals with $W_i > 0$. By mimicking the arguments of Li et al. (1996), one can show that $Q(p, r|x)$ is always an interval. The lower limit is the first t_j for which $R(p, t_j|x) \geq r$, and the upper boundary is the first subsequent t_j for which $R(p, t_j|x) < r$.

Remark 2.2. **[Bandwidth selection]** An important issue in the context of kernel regression is the selection of an appropriate bandwidth sequence h_n. Unlike in estimation problems, where the bandwidth is often chosen as the value for which the mean (integrated) squared error is minimal, a different criterion should be used for the construction of confidence intervals and bands. As was done in Hall (1991, 1992), among others, it is sensible to consider the bandwidth for which the coverage error of the interval or band is minimal. It is however not easy to find the optimal bandwidth under this criterion since the coverage error is unknown for a given bandwidth. In practice, we suggest to use an approximate optimal bandwidth that minimizes an estimated coverage error. The coverage error for a given bandwidth can be estimated using the bootstrap method for nonparametric censored regression proposed by Li and Datta (1999) for random design (and Van Keilegom and Veraverbeke (1997) for fixed design). More precisely, we generate B bootstrap samples. The resampling method consists of two stages. In the first stage, we resample with replacement from the set $\{X_1, \ldots, X_n\}$ to obtain the bootstrap sample $\{X_1^*, \ldots, X_n^*\}$. Next, for each i, we draw a pair (T_i^*, Δ_i^*) with replacement from $(T_1, \Delta_1), \ldots, (T_n, \Delta_n)$, giving probability $W_j(X_i^*; g_n)$ to (T_j, Δ_j) $(j = 1, \ldots, n)$, where $\{g_n\}$ is a second bandwidth sequence. Next, calculate from each of these B bootstrap samples a $100(1-\alpha)\%$ confidence band for $\hat{S}_{(g)}(\cdot|x)$ (where $\hat{S}_{(g)}(\cdot|x)$ denotes the Beran estimator $\hat{S}(\cdot|x)$ using the bandwidth g_n) and compute the coverage probability of these bands (i.e. the proportion of these B bands which include $\hat{S}_{(g)}(\cdot|x)$). Select the bandwidth h_n for which this coverage probability is closest to $1 - \alpha$ (i.e. for which the coverage error is minimal).

A further question is how to select the second bandwidth sequence g_n in the bootstrap.

In general, g_n is required to be slightly larger than h_n [cf. Li and Datta, 1999, Theorem 1 and Remark 2.1]. Our limited experience showed that $g_n = h_n^{0.8}$ performs well in the computation of the bootstrap coverage errors and that the result is not very sensitive to a small perturbation of g_n.

It would be interesting to find the optimal rate under the minimal coverage error criterion. This is, however, beyond the scope of the present paper. For the construction of confidence intervals for a density function based on completely observed data, Hall (1991, 1992) considered two methods for obtaining the optimal bandwidth: one method is based on undersmoothing and the other one requires the explicit estimation of the asymptotic bias. He found that undersmoothing leads to a smaller coverage error and is therefore preferable. He also showed that for both methods, the optimal bandwidth satisfies $nh_n^5 \to 0$, while in estimation problems this condition is $nh_n^5 \to C > 0$. Although we did not determine the rate of the optimal bandwidth in the present context theoretically, we carried out a simulation study to learn how the bands behave when $nh_n^5 = C > 0$ (in which case there is an extra bias term in the formula of the bands). The simulations strongly indicated that estimating the bias term often leads to coverage probabilities that are considerably too low. Further examination of the simulation results showed that the bias term is often badly estimated, and is the main reason of the large coverage errors. As will be seen in the next section, this is not the case when the asymptotic bias does not need to be estimated (i.e. when $nh_n^5 \to 0$). It is well known that undersmoothing results in a smaller bias. A trade-off is a larger variance and thus a wider confidence interval. This is a price one has to pay for a confidence interval with a small coverage error.

Remark 2.3. [Normal approximation methods for constructing confidence intervals and bands] It has been shown that (cf. Li and Doss (1995, Theorem 4 and the remark following the proof of Theorem 1)), if $nh_n \to \infty$ and $nh_n^5 \to 0$, then $\sqrt{nh_n}(\hat{S}(t|x) - S(t|x))$ converges weakly to $S(t|x)U(t|x)$, where $U(t|x)$ is a continuous Gaussian martingale with mean zero and variance function $\sigma^2(t|x)$. This result enables one to construct confidence intervals and bands for $S(t|x)$. For example, an approximate $1-\alpha$ confidence interval for $S(t|x)$ is given by

$$\hat{S}(t|x) \pm z_{\alpha/2}\hat{S}(t|x)\hat{\sigma}(t|x)/\sqrt{nh_n},$$

where $z_{\alpha/2}$ is the $\alpha/2$ upper quantile of a standard normal distribution. Confidence bands for $S(t|x)$ can be derived using the method of Hall and Wellner (1980) for an unconditional survival function based on homogeneous data. Furthermore, weak convergence of $\hat{S}^{-1}(p|x)$ to a Gaussian process follows directly from that of $\hat{S}(t|x)$ and the functional delta-method (cf. Andersen et al. (1993), Propositions II.8.1 and II.8.5). Normal confidence intervals for $S^{-1}(p|x)$ can then be obtained in the usual manner. Hall-Wellner type confidence bands for $S^{-1}(p|x)$ can be derived using the idea described in Li et al. (1996, p.634).

3 Simulations

In this section we will carry out a number of simulations to illustrate the small sample performance of the proposed confidence intervals and bands for $S(t|x)$. More precisely, we will compare the coverage probability and the width of the proposed intervals and bands with those of the normal method. All simulation results are based on 1000 simulation runs. The programs are written in the statistical package Gauss.

Assume that the covariate X is uniformly distributed on the interval $[0, 1]$. The conditional distribution of the response Y given $X = x$ is exponential with mean $a_0 + a_1x + a_2x^2$, while the censoring time C has an exponential distribution with mean function $b_0 + b_1x + b_2x^2$ for certain constants a_0, a_1, a_2, b_0, b_1 and b_2. As usual, we assume that Y and C are independent, conditionally on X. Hence, it is readily verified that $(T|X = x) \sim \text{Exp}(1/(a_0 + a_1x + a_2x^2) + 1/(b_0 + b_1x + b_2x^2))$ and that

$$P(\Delta = 0|x) = \frac{a_0 + a_1x + a_2x^2}{a_0 + b_0 + (a_1 + b_1)x + (a_2 + b_2)x^2}. \tag{3.1}$$

We carry out simulations for samples of size 25, 50 and 100, for $x = 0.5$ and $\alpha = 0.05$. At $x = 0.5$, the probability of censoring is either 0.25, 0.50 or 0.75. The kernel function K is a biquadratic kernel $K(x) = (15/16)(1 - x^2)^2 I(|x| \leq 1)$. For each method, we used a bandwidth that minimizes the coverage error of the interval or band over a grid of bandwidths ranging from 0.2 up to 0.5 (see Remark 2.2 for a detailed description and references on this bandwidth selection method). Table 1 shows the result for confidence intervals at $t = S^{-1}(0.5|x)$ and for different choices of the parameters a_0, a_1, b_0, b_1 and b_2. The value of a_2 is determined in such a way that the probability of censoring, given by (3.1), equals one of the three considered values (0.25, 0.50 or 0.75).

The table shows that the coverage probability of the likelihood method is very stable and everywhere close to 0.95 (by taking a finer grid of bandwidths it would be possible

Table 1: *Coverage probability and length of simulated confidence intervals for $S^{-1}(0.5|x)$ (obtained by selecting the bandwidth that minimizes the coverage error).*

Para-meters	$P(\Delta = 0\|x)$	n	coverage probab. normal	coverage probab. likel.	length normal	length likel.
a_0=1	0.25	25	0.927	0.952	0.488	0.504
a_1=2		50	0.933	0.949	0.349	0.339
b_0=4		100	0.930	0.943	0.278	0.312
b_1=5	0.50	25	0.910	0.952	0.542	0.633
b_2=50		50	0.938	0.950	0.391	0.377
		100	0.938	0.950	0.279	0.274
	0.75	25	0.824	0.944	0.668	0.748
		50	0.858	0.946	0.511	0.653
		100	0.886	0.948	0.443	0.661
a_0=1	0.25	25	0.921	0.951	0.488	0.459
a_1=-1		50	0.925	0.951	0.350	0.511
b_0=1		100	0.938	0.949	0.249	0.312
b_1=1	0.50	25	0.900	0.950	0.549	0.641
b_2=-1		50	0.936	0.952	0.397	0.382
		100	0.930	0.944	0.283	0.439
	0.75	25	0.818	0.960	0.675	0.762
		50	0.878	0.953	0.536	0.658
		100	0.909	0.959	0.409	0.670

to obtain coverage probabilities which are even closer to 0.95). On the other hand, the coverage probability of the normal method is everywhere below 0.95. Especially under heavy censoring, the behavior of the normal method is very poor. The table also shows that the normal method produces shorter intervals than the likelihood ratio method. However, this comparison is misleading, since the normal method has lower coverage probability. Table 2 shows that when the two intervals have about the same lengths, the likelihood interval has higher coverage probability than the normal method. Also note from Table 2 that the coverage probabilty for bandwidths other than the optimal one, is

Table 2: *Coverage probability and length of simulated confidence intervals for* $S^{-1}(0.5|x)$*, when* $a_0 = 1, a_1 = 2, b_0 = 4, b_1 = 5, b_2 = 50$ *and* $P(\Delta = 0|x) = 0.75$*. The numbers in italic are the ones for which the coverage error is minimal.*

n	h_n	coverage probab.		length	
		normal	likel.	normal	likel.
25	0.2	0.564	0.973	0.660	0.833
	0.3	0.710	*0.944*	0.704	0.748
	0.4	0.785	0.914	0.709	0.664
	0.5	*0.824*	0.891	0.668	0.595
50	0.2	0.731	0.973	0.689	0.763
	0.3	0.823	*0.946*	0.638	0.653
	0.4	0.849	0.933	0.575	0.565
	0.5	*0.858*	0.911	0.511	0.482
100	0.2	0.825	*0.948*	0.598	0.661
	0.3	0.856	0.926	0.521	0.540
	0.4	*0.886*	0.928	0.443	0.446
	0.5	0.883	0.916	0.382	0.373

quite stable for the likelihood method, but can be far from 0.95 for the normal method.

Next, we compare the behavior of the likelihood ratio confidence band and the Hall-Wellner band. The bandwidth is again determined such that the coverage error is minimal. The same distributions are considered as for the confidence intervals, except that $a_2 = b_2 = 0$, i.e. the survival and censoring times have a linear mean function. The value of b_0 is determined such that the probability of censoring at $x = 0.5$ equals one of the three considered values (0.25, 0.50 and 0.75). The bands are compared over the interval $[u_1, u_2] = [0.1, 0.9]$. The critical point $e_\alpha(\hat{u}_1, \hat{u}_2)$ equals 3.06 (see Nair (1984)). The result is shown in Table 3. Again we see that the likelihood method outperforms the normal method, both in terms of coverage probability and total area of the bands.

Table 3: *Coverage probability and total area of simulated confidence bands (obtained by selecting the bandwidth that minimizes the coverage error).*

Para-meters	$P(\Delta = 0\|x)$	n	coverage probab. normal	coverage probab. likel.	total area normal	total area likel.
a_0=0.5	0.25	25	0.969	0.981	1.574	1.298
a_1=1		50	0.967	0.973	1.109	0.974
b_1=1		100	0.956	0.959	0.771	0.701
	0.50	25	0.983	0.986	1.125	0.915
		50	0.968	0.974	1.020	0.864
		100	0.956	0.963	0.750	0.664
	0.75	25	0.993	0.986	0.677	0.308
		50	0.978	0.980	0.687	0.580
		100	0.966	0.983	0.633	0.472
a_0=1	0.25	25	0.956	0.966	0.715	0.596
a_1=-1		50	0.953	0.955	0.599	0.498
b_1=2		100	0.954	0.955	0.486	0.407
	0.50	25	0.971	0.963	0.469	0.372
		50	0.952	0.952	0.435	0.366
		100	0.950	0.966	0.371	0.347
	0.75	25	0.989	0.960	0.358	0.267
		50	0.957	0.957	0.347	0.295
		100	0.950	0.952	0.297	0.285

4 Data analysis

We will illustrate the proposed method on data on cancer of the larynx, discussed by Kardaun (1983). Of each of 90 male larynx cancer patients, diagnosed and treated during the period 1970-1978 in a hospital in the Netherlands, the age at diagnosis of the cancer (in years) and the time between the first treatment and the death of the patient (in years) are recorded (among other variables). At the end of the study in 1981, 40 out of the 90 patients were alive. Since their survival time could not be recorded, these individuals are

Table 4: *Bootstrapped coverage probabilities for the larynx cancer data.*

h_n	15	17.5	20	22.5	25
cover. prob. normal	0.945	0.939	0.949	0.945	0.939
cover. prob. likel.	0.974	0.973	0.962	0.959	0.961

right censored. We wish to construct a confidence band for the survival distribution at a given age.

To select an appropriate bandwidth h_n, we use the criterion based on minimizing the estimated coverage error using the bootstrap as described in Remark 2.2 with $g_n = h_n^{0.8}$. Let $K(x) = (15/16)(1 - x^2)^2 I(|x| \leq 1)$ be a biquadratic kernel function, let $\alpha = 0.05$ and $x = 65$ (so we construct a 95 % confidence band for a 65 years old patient). Table 4 shows the bootstrapped coverage probabilities for different values of the bandwidth h_n and based on $B = 1000$ bootstrap samples. For the normal method, the best result is obtained for $h_n = 20$, while the likelihood method performs the best when $h_n = 22.5$. Figure 1 shows the normal and likelihood ratio bands by using these bandwidth choices.

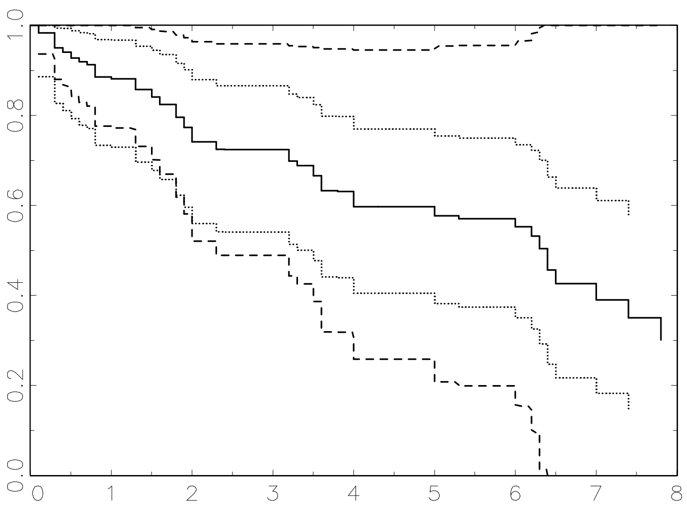

Figure 1: *Normal and likelihood ratio confidence bands for the larynx cancer data. The full line is the Beran estimator $\hat{S}(t|x)$, while the dashed respectively dotted lines are the upper and lower bound of the normal respectively likelihood ratio confidence band.*

The bands are at most time points narrower for the likelihood ratio method than for the normal approximation method. Note that this is in agreement with the results of the simulations in Section 3, which also indicated that the normal method leads to wider confidence bands than the likelihood ratio method.

Appendix

Throughout this appendix we use the following abbreviated notations:

$$W_i = W_i(x; h_n), \tag{A.1}$$

$$R_i = \sum_{j=1}^{n} W_j I(T_j \geq T_i). \tag{A.2}$$

We start with two lemmas which enable us to calculate the likelihood ratio statistic $R(p, t|x)$. The proof of the first one is very similar to part of the proof of Theorem 2.1 in Li (1995) and is therefore omitted.

Lemma A.1 *Assume that t is inside the convex hull of all uncensored survival times of individuals with $W_i > 0$ and assume that $S(\cdot|x)$ is continuous. Then,*

$$R(p, t|x) = \frac{\sup \left\{ \prod_{\Delta_i=1} h_i^{W_i} (1 - h_i)^{R_i - W_i} \right\}}{\prod_{\Delta_i=1} \left(\frac{W_i}{R_i} \right)^{W_i} \left(1 - \frac{W_i}{R_i} \right)^{R_i - W_i}},$$

where the supremum is taken over all $h_1, \ldots, h_n$ such that $\prod_{T_i \leq t, \Delta_i=1} (1 - h_i) = p$ and $0 \leq h_i \leq 1$ for $i = 1, \ldots, n, \Delta_i = 1$.

Let $\lambda_n(p, t|x)$ (or simply λ_n) be the unique solution of

$$\prod_{T_i \leq t, \Delta_i=1} \left\{ 1 - \frac{W_i}{R_i + \lambda} \right\} = p \tag{A.3}$$

on the interval $(D, +\infty)$, where

$$D = \max_{T_i \leq t, \Delta_i=1, W_i > 0} (W_i - R_i).$$

(Note that the left hand side of (A.3) is a strictly increasing function of λ on $(D, +\infty)$, which converges to 0 as λ tends to D and to 1 as λ tends to $+\infty$.)

Lemma A.2 *Under the assumption of Lemma A.1,*

$$R(p,t|x) = \prod_{T_i \le t, \Delta_i = 1} \left(\frac{R_i}{R_i + \lambda_n(p,t|x)} \right)^{R_i} \left(\frac{R_i - W_i + \lambda_n(p,t|x)}{R_i - W_i} \right)^{R_i - W_i}.$$

Proof. We will show that the supremum in the numerator of $R(p,t|x)$ is reached when $h_i = W_i/(R_i + \lambda_n)$ for $T_i \le t, \Delta_i = 1$ and $h_i = W_i/R_i$ for $T_i > t, \Delta_i = 1$. Simple calculations then lead to the result. Define the Lagrange function

$$L(h_1,\ldots,h_n,\alpha) = \log \left\{ \prod_{T_i \le t, \Delta_i = 1} h_i^{W_i} (1-h_i)^{R_i - W_i} \right\} + \alpha \left\{ \log p - \log \left(\prod_{T_i \le t, \Delta_i = 1} (1-h_i) \right) \right\}.$$

It is easily verified that the system of equations $\frac{\partial L}{\partial h_i} = 0$, $\frac{\partial L}{\partial \alpha} = 0$ $(i = 1,\ldots,n)$ leads to

$$h_i = \frac{W_i}{R_i - \alpha} \qquad \text{if } T_i \le t, \Delta_i = 1$$

$$h_i = \frac{W_i}{R_i} \qquad \text{if } T_i > t, \Delta_i = 1.$$

From the definition of λ_n it follows that $\alpha = -\lambda_n$ and hence the result follows.

Lemma A.3 *Assume that the conditions of Theorem 2.1 hold. Then, for* $0 < t_1 < t_2 < \tau_x$*, such that* $S(t_1|x) < 1$*,*

$$\lambda_n(S(t|x), t|x) = O_P((nh_n)^{-1/2}),$$

uniformly in $t_1 \le t \le t_2$*.*

Proof. Take $t_1 \le t \le t_2$ fixed. We consider two cases. If $\lambda_n < 0$, then, since $\lambda_n > D$, we have that $R_i + \lambda_n > W_i \ge 0$ and $0 \le \frac{W_i}{R_i + \lambda_n} < 1$ for $T_i \le t, \Delta_i = 1$. Hence, the inequality $-\log(1-x) \ge x$ for $0 \le x < 1$ leads to

$$-\log S(t|x) = -\sum_{T_i \le t, \Delta_i = 1} \log\left(1 - \frac{W_i}{R_i + \lambda_n}\right) \ge \sum_{T_i \le t, \Delta_i = 1} \frac{W_i}{R_i} \frac{R_i}{R_i + \lambda_n}$$

$$\ge \frac{1}{1+\lambda_n} \sum_{T_i \le t, \Delta_i = 1} \frac{W_i}{R_i} = \frac{1}{1+\lambda_n} \hat{\Lambda}(t|x),$$

where $\hat{\Lambda}(t|x) = \sum_{T_i \le t, \Delta_i = 1} \frac{W_i}{R_i}$ is an estimator of the cumulative hazard function $\Lambda(t|x) = -\int_0^t \frac{dS(s|x)}{S(s|x)} = -\int_0^t \frac{dH^u(s|x)}{1-H(s|x)}$. Hence,

$$|\lambda_n| \le \frac{\hat{\Lambda}(t|x) + \log S(t|x)}{\log S(t|x)} = O_P((nh_n)^{-1/2}),$$

uniformly in $t_1 \leq t \leq t_2$, which follows from the weak convergence of $\hat{\Lambda}(t|x)$ (see e.g. Li and Doss (1995)).

If $\lambda_n \geq 0$, then, since $\log(1-x)+x$ is non-increasing on $[0,1)$,

$$
\begin{aligned}
\log S(t|x) &= \sum_{T_i \leq t, \Delta_i=1} \log\left(1 - \frac{W_i}{R_i+\lambda_n}\right) \\
&\geq \sum_{T_i \leq t, \Delta_i=1} \left\{-\frac{W_i}{R_i+\lambda_n} + \log\left(1-\frac{W_i}{R_i}\right) + \frac{W_i}{R_i}\right\} \\
&= -\sum_{T_i \leq t, \Delta_i=1} \frac{W_i}{R_i}\frac{R_i}{R_i+\lambda_n} + \log \hat{S}(t|x) + \hat{\Lambda}(t|x) \\
&\geq -\frac{1}{1+\lambda_n}\hat{\Lambda}(t|x) + \log \hat{S}(t|x) + \hat{\Lambda}(t|x). \qquad \text{(A.4)}
\end{aligned}
$$

Next, since $\log(1-x) - \log(1-y) \leq \frac{y-x}{1-y}$ for $0 \leq x, y < 1$, it follows after some simple calculations that $\hat{\Lambda}(t|x) + \log \hat{S}(t|x) - \log S(t|x) \geq 0$. This together with (A.4) and the weak convergence of $\log \hat{S}(t|x)$ (see Li and Doss (1995)) yields

$$
|\lambda_n| \leq \frac{\log S(t|x) - \log \hat{S}(t|x)}{\hat{\Lambda}(t|x) + \log \hat{S}(t|x) - \log S(t|x)} = O_P((nh_n)^{-1/2}),
$$

uniformly in $t_1 \leq t \leq t_2$, from which the result follows.

Lemma A.4 *Assume that the conditions of Theorem 2.1 hold. Then, for* $0 < t_1 < t_2 < \tau_x$*, such that* $S(t_1|x) < 1$,

$$
\begin{aligned}
&-2nh_n \frac{\hat{f}_X(x)}{\int K^2(u)\,du} \log R(S(t|x), t|x) \\
&= nh_n \hat{\sigma}^{-2}(t|x)(\log \hat{S}(t|x) - \log S(t|x))^2 + o_P(1),
\end{aligned}
$$

uniformly in $t_1 \leq t \leq t_2$.

Proof. First note that for $T_i \leq t$, R_i converges to $1 - H(T_i|x) > 1 - H(t|x) > 0$ and hence R_i is bounded away from zero uniformly in all i for which $T_i \leq t$. Let

$$
A(\lambda) = \log\left\{\prod_{T_i \leq t, \Delta_i=1}\left(1 - \frac{W_i}{R_i+\lambda}\right)\right\}.
$$

Then, $A(0) = \log \hat{S}(t|x)$, $A(\lambda_n(S(t|x), t|x)) = \log S(t|x)$ and $A'(0) = \hat{f}_X(x)\hat{\sigma}^2(t|x)/\int K^2(u)\,du$, since

$$\begin{aligned}\hat{\sigma}^2(t|x) &= \int K^2(u)\,du\,\hat{f}_X^{-1}(x)\int_0^t \frac{d\hat{H}^u(s|x)}{(1-\hat{H}(s|x))(1-\hat{H}(s-|x))}\\ &= \int K^2(u)\,du\,\hat{f}_X^{-1}(x)\sum_{T_i\le t,\Delta_i=1}\frac{W_i}{R_i(R_i-W_i)}.\end{aligned}$$

Hence, a Taylor expansion yields that

$$\log S(t|x) = \log \hat{S}(t|x) + \lambda_n \frac{\hat{f}_X(x)\hat{\sigma}^2(t|x)}{\int K^2(u)\,du} + \frac{\lambda_n^2}{2}A''(\xi_n),$$

where $|\xi_n| < |\lambda_n|$. It follows that

$$\lambda_n = \int K^2(u)\,du\,\hat{f}_X^{-1}(x)\hat{\sigma}^{-2}(t|x)(\log S(t|x) - \log \hat{S}(t|x)) + O_P((nh_n)^{-1}),$$

since Lemma A.3 implies that

$$\begin{aligned}\lambda_n^2 A''(\xi_n) &= -\lambda_n^2 \sum_{T_i\le t,\Delta_i=1}\frac{(2(R_j+\xi_n)-W_j)W_j}{(R_j-W_j+\xi_n)^2(R_j+\xi_n)^2}\\ &= O_P((nh_n)^{-1}).\end{aligned}$$

Using again a Taylor expansion and Lemma A.2, we have

$$\begin{aligned}&-2nh_n\frac{\hat{f}_X(x)}{\int K^2(u)\,du}\log R(S(t|x),t|x)\\ &= \frac{-2nh_n\hat{f}_X(x)}{\int K^2(u)\,du}\sum_{T_i\le t,\Delta_i=1}\left\{(R_i-W_i)\log\left(1+\frac{\lambda_n}{R_i-W_i}\right) - R_i\log\left(1+\frac{\lambda_n}{R_i}\right)\right\}\\ &= \frac{nh_n\hat{f}_X(x)}{\int K^2(u)\,du}\left\{\lambda_n^2\sum_{T_i\le t,\Delta_i=1}\frac{W_i}{R_i(R_i-W_i)} - \frac{2\lambda_n^3}{3}\sum_{T_i\le t,\Delta_i=1}\left[\left(\frac{1}{R_i-W_i}\right)^2 - \left(\frac{1}{R_i}\right)^2\right]\right.\\ &\quad\left. + \frac{\lambda_n^4}{2}\sum_{T_i\le t,\Delta_i=1}\left[\left(\frac{1}{R_i-W_i}\right)^3 - \left(\frac{1}{R_i}\right)^3\right]\right\} + O_P((nh_n^3)^{-1/2})\\ &= T_1 + T_2 + T_3 + o_P(1).\end{aligned}$$

We start with T_1 :

$$\begin{aligned}T_1 &= nh_n\hat{f}_X^2(x)\left(\int K^2(u)\,du\right)^2\lambda_n^2\hat{\sigma}^2(t|x)\\ &= nh_n\hat{\sigma}^{-2}(t|x)(\log S(t|x) - \log\hat{S}(t|x))^2 + O_P((nh_n)^{-1}).\end{aligned}$$

Simple calculations show that $T_2 = O_P((nh_n)^{-1/2})$ and that $T_3 = O_P((nh_n)^{-1})$. This finishes the proof.

Proof of Theorem 2.1. (a) Use Lemma A.4 together with the asymptotic normality of $\log \hat{S}(t|x)$ given in Li and Doss (1995).
(b) Since $S(t|x) \in C(t, r_\alpha|x)$ if and only if $R(S(t|x), t|x) \geq r_\alpha$, this follows readily from part (a).

Proof of Theorem 2.2. (a) It follows from Lemma A.4 and from Li and Doss (1995) that, as $n \to \infty$,

$$\begin{aligned}
&-2nh_n \frac{\hat{f}_X(x)}{\int K^2(u)\,du} \log R(S(t|x), t|x) \\
&= nh_n \hat{\sigma}^{-2}(t|x)(\log \hat{S}(t|x) - \log S(t|x))^2 + o_P(1) \\
&\xrightarrow{d} \left\{ \frac{Z(t|x)}{\sigma(t|x)} \right\}^2
\end{aligned}$$

in $D[t_1, t_2]$, where $Z(t|x)$ is a continuous Gaussian process with mean 0 and covariance $\mathrm{Cov}(Z(s|x), Z(t|x)) = \sigma^2(\min(s,t)|x)$. Next, note that the processes $\frac{Z(t|x)}{\sigma(t|x)}$ and $(\beta B^0)(\frac{\sigma^2(t|x)}{1+\sigma^2(t|x)})$ are equal in distribution since they are Gaussian processes with the same mean and covariance function, where $\beta(u) = (u(1-u))^{-1/2}$ $(0 < u < 1)$. Hence, the result follows.
(b) It follows from (a) that

$$\begin{aligned}
&\lim_{n\to\infty} P\{S(t|x) \in C(t, r|x) \text{ for all } t_1 \leq t \leq t_2\} \\
&= \lim_{n\to\infty} P\{R(S(t|x), t|x) \geq r \text{ for all } t_1 \leq t \leq t_2\} \\
&= \lim_{n\to\infty} P\left\{ \sup_{t_1 \leq t \leq t_2} \left(-2nh_n \frac{\hat{f}_X(x)}{\int K^2(u)\,du} \log R(S(t|x), t|x) \right) \hat{f}_X^{-1}(x) \leq -\frac{2nh_n \log r}{\int K^2(u)\,du} \right\} \\
&= \lim_{n\to\infty} P\left\{ \sup_{\hat{u}_1 \leq u \leq \hat{u}_2} \left| \frac{B^0(u)}{\sqrt{u(1-u)}} \right| \leq \sqrt{-\frac{2nh_n \hat{f}_X(x) \log r}{\int K^2(u)\,du}} \right\}.
\end{aligned}$$

For $r = r_\alpha(\hat{u}_1, \hat{u}_2)$, this probability equals $1 - \alpha$.

References

Akritas, M.G. (1994). Nearest neighbor estimation of a bivariate distribution under random censoring. *Ann. Statist.*, **22**, 1299-1327.

Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993). *Statistical models based on counting processes.* Springer-Verlag, New York.

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical Report, Dept. Statist., Univ. California, Berkeley.

Chen, S.X. (1993). On the accuracy of empirical likelihood confidence regions for linear regression model. *Ann. Inst. Statist. Math.*, **45**, 621-637.

Chen, S.X. (1994). Empirical likelihood confidence intervals for linear regression coefficients. *J. Multiv. Anal.*, **49**, 24-40.

Chen, S.X. and Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Ann. Statist.*, **21**, 1166-1181.

Chen, S.X. and Qin, Y.S. (2000) Empirical likelihood confidence intervals for local linear smoothers. *Biometrika*, **87**, 946-953.

Dabrowska, D.M. (1987). Non-parametric regression with censored survival time data. *Scand. J. Statist.*, **14**, 181-197.

Dabrowska, D.M. (1992). Variable bandwidth conditional Kaplan-Meier estimate. *Scand. J. Statist.*, **19**, 351-361.

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

González Manteiga, W. and Cadarso Suarez, C. (1994). Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. *J. Nonpar. Statist.*, **4**, 65-78.

Hall, P. (1991). Edgeworth expansions for nonparametric density estimators, with applications. *Statistics*, **22**, 215-232.

Hall, P. (1992). *The bootstrap and edgeworth expansion.* Springer-Verlag, New York.

Hall, W. J. and Wellner, J. A. (1980). Confidence bands for a survival curve from censored data. *Biometrika*, **67**, 133-143.

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

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457-481.

Kardaun, O. (1983). Statistical survival analysis of male larynx-cancer patients - a case study. *Statist. Neerland.*, **37**, 103-125.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Ann. Statist.*, **25**, 2084-2102.

Kolaczyk, E.D. (1994). Empirical likelihood for generalized linear models. *Statist. Sinica*, **4**, 199-218.

Li, G. (1995). On nonparametric likelihood ratio estimation of survival probabilities for censored data. *Statist. Prob. Letters*, **25**, 95-104.

Li, G. (1997). Optimal rate local smoothing in a multiplicative intensity counting process model. *Math. Methods Statist.*, **6**, 224-244.

Li, G. and Datta, S. (1999). Bootstrapped confidence bands for conditional survival functions from right censored data. *Manuscript.*

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

Li, G., Hollander, M., McKeague, I.W. and Yang, J. (1996). Nonparametric likelihood ratio confidence bands for quantile functions from incomplete survival data. *Ann. Statist.*, **24**, 628-640.

McKeague, I.W. and Nikabadze, A.M. and Sun, Y. (1995). On omnibus test for independence of a survival time from a covariate. *Ann. Statist.*, **23**, 450-475.

McKeague, I.W. and Sun, Y. (1996). Towards an omnibus distribution-free goodness-of-fit test for the Cox model. *Statist. Sinica*, **6**, 579-588.

McKeague, I.W. and Utikal, K.J. (1990). Inference for a nonlinear counting process regression model. *Ann. Statist.*, **18**, 1172-1187.

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

Nair, V.N. (1984). Confidence bands for survival functions with censored data : a comparative study. *Technom.*, **26**, 265-275.

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

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

Owen, A. (1991). Empirical likelihood for linear models. *Ann. Statist.*, **19**, 1725-1747.

Owen, A. (1995). Nonparametric likelihood confidence bands for a distribution function. *J. Amer. Statist. Assoc.*, **90**, 516-521.

Qin, J. (1993). Empirical likelihood in biased sample problems. *Ann. Statist.*, **21**, 1182-1196.

Qin, J. and Lawless, J.F. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.*, **22**, 300-325.

Stone, C.J. (1977). Consistent nonparametric regression. *Ann. Statist.*, **5**, 595-645.

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

Van Keilegom, I. and Veraverbeke, N. (1997). Estimation and bootstrap with censored data in fixed design nonparametric regression. *Ann. Inst. Statist. Math.*, **49**, 467-491.

Van Keilegom, I. and Veraverbeke, N. (1998). Bootstrapping quantiles in a fixed design regression model with censored data. *J. Statist. Planning Inf.*, **69**, 115-131.

Wang, Q.H. and Jing, B.Y. (1999). Empirical likelihood for partial linear model with fixed design. *Statist. Prob. Letters*, **41**, 425-433.