Lifetime Data Anal (2006) 12:407–419
DOI 10.1007/s10985-006-9024-y

Confidence intervals for survival quantiles in the Cox regression model

Tze Leung Lai · Zheng Su

Received: 22 December 2005 / Accepted: 6 September 2006 /
Published online: 20 October 2006

Abstract Median survival times and their associated confidence intervals are often used to summarize the survival outcome of a group of patients in clinical trials with failure-time endpoints. Although there is an extensive literature on this topic for the case in which the patients come from a homogeneous population, few papers have dealt with the case in which covariates are present as in the proportional hazards model. In this paper we propose a new approach to this problem and demonstrate its advantages over existing methods, not only for the proportional hazards model but also for the widely studied cases where covariates are absent and where there is no censoring. As an illustration, we apply it to the Stanford Heart Transplant data. Asymptotic theory and simulation studies show that the proposed method indeed yields confidence intervals and bands with accurate coverage errors.

Keywords Bootstrap · Median survival · Proportional hazards model · Test-based confidence intervals and bands

1 Introduction

The proportional hazards model of Cox (1972) is a log-linear regression model that relates the cumulative hazard function $\Lambda(t|x)$ associated with a covariate vector x to a baseline hazard function $\Lambda(t)$ via

$$\Lambda(t|x) = \Lambda(t)\exp(\beta^T x). \tag{1.1}$$

T. L. Lai (✉)
Department of Statistics, Stanford University, Stanford, CA 94305, USA
e-mail: lait@stat.stanford.edu

Z. Su
Department of Applied Mathematics and Statistics, SUNY Stony Brook, NY 11733, USA
e-mail: zhengsu@ams.sunysb.edu

Based on a sample consisting of n observations $(\widetilde{t}_i, \delta_i, x_i)$, where $\widetilde{t}_i = \min(t_i, c_i)$ and $\delta_i = I_{\{t_i \le c_i\}}$ is the indicator of whether the actual failure time t_i is observed or is censored by c_i, the estimate $\widehat{\beta}$ of β is the maximizer of the partial likelihood function

$$\ell(\beta) = \sum_{i=1}^{n} \delta_i \left\{ \beta^T x_i - \log\left(\sum_{j:\tilde{t}_j \ge \tilde{t}_i} \exp(\beta^T x_j) \right) \right\}. \tag{1.2}$$

Confidence regions for β can be constructed by using the asymptotic normality of $(-\ddot{\ell}(\widehat{\beta}))^{-1/2}(\widehat{\beta} - \beta)$ or the limiting χ^2 distribution of $2\{ell(\widehat{\beta}) - \ell(\beta)\}$. In many applications, it is useful to estimate also the median survival time given a subject's covariate vector. In particular, by combining $\widehat{\beta}$ with Breslow's (1974) estimate $\widehat{\Lambda}$ of the baseline cumulative hazard function, Miller and Halpern (1982) used the median of the distribution function $1 - \exp\{-\widehat{\Lambda}(\cdot)e^{\hat{\beta}^T x}\}$ to estimate median survival, given a subject's age that forms the covariate vector x = (age, age^2), from the Stanford Heart Transplant data. Dabrowska and Doksum (1987) and Burr and Doss (1993) subsequently studied the problem of constructing confidence intervals and bands for the median survival time given a subject's covariates. Letting ξ_p (x) denote the pth quantile of the failure time distribution for a given covariate vector x (so that $p = \frac{1}{2}$ corresponds to the median) and $\widehat{\xi}_p(x)$ be the pth quantile of the preceding estimated distribution function, their approach is based on the approximate normality of $\{\widehat{\xi}_p(x) - \xi_p(x)\}/\widehat{\text{se}}_p(x)$, or its limiting Gaussian process indexed by x, where $\widehat{\text{se}}_p(x)$ denotes the estimated standard error of $\widehat{\xi}_p(x)$.

A major difficulty with this approach for sample sizes commonly encountered in practice lies in $\widehat{\text{se}}_p(x)$. The variance of the limiting normal distribution of $\sqrt{n}\{\widehat{\xi}_p(x) - \xi_p(x)\}$ involves the baseline hazard function $\lambda(t) = (d/dt)\Lambda(t)$. Although Dabrowska and Doksum (1987, p. 802) cite Tsiatis (1981) and Andersen and Gill (1982) in claiming consistency of their proposed estimator of the limiting variance, Tsiatis, Anderson and Gill have only established consistency for Breslow's estimate of Λ but not of the derivative λ. Burr and Doss (1993) make use of kernel smoothing of $\widehat{\Lambda}$ to estimate λ, and instead of applying the large-sample theory of $\{\widehat{\xi}_p(x) - \xi_p(x)\}/\widehat{\text{se}}_p(x)$ directly to construct confidence intervals for $\xi_p(x)$, they use it to provide a theoretical justification of the bootstrap-t method to construct confidence intervals. However, as pointed out by Efron and Tibshirani (1993, Sect.12.6), the bootstrap-t method requires stable estimates of standard errors for it to work well in practice. Therefore the difficulties in estimating the standard error of $\widehat{\xi}_p(x)$ also cause difficulties with bootstrap-t confidence intervals for $\xi_p(x)$.

In fact, even without censoring and covariate effects so that the problem reduces to that of confidence intervals for the pth quantile ξ_p of a distribution function based on a sample of independent and identically distributed survival times $t_1, \ldots t_n$ with common density function f that has a consistent kernel estimator $\widehat{f}$, the limiting normal distribution of

$$\widehat{f}(\widehat{\xi}_p)\{n/[p(1-p)]\}^{1/2}(\widehat{\xi}_p - \xi_p) \tag{1.3}$$

is seldom used in constructing confidence intervals for ξ_p. Besides issues with finite-sample performance of the density estimator $\widehat{f}(\widehat{\xi}_p)$, the adequacy of the linear approximation $f(\xi_p)(\widehat{\xi}_p - \xi_p)$ to $F(\widehat{\xi}_p) - F(\xi_p)$ used to derive the asymptotic normality of $\widehat{\xi}_p - \xi_p$ (where F is the distribution function whose derivative is f) is problematic when $\widehat{\xi}_p$ is not sufficiently near ξ_p. Instead, a standard nonparametric confidence interval is of the form $t_{(k_1)} < \xi_p < t_{(k_2)}$, where the $t_{(i)}$ denote the order statistics of the sample and $k_1 < k_2$ are integers such that

$$P\{t_{(k_1)} \leq \xi_p < t_{(k_2)}\} = P\{k_1 \leq \mathrm{Bi}(n,p) < k_2\} \geq 1 - 2\alpha; \tag{1.4}$$

the lower bound $1 - 2\alpha$ in (1.4) may not be attainable because of the discreteness of the binomial distribution Bi(n,p). As shown by Efron (1979) and Chen and Hall (1993, p. 1169), bootstrap percentile confidence intervals and empirical likelihood confidence intervals (obtained by inverting empirical likelihood ratio tests) for ξ_p are of this form; see also Efron and Tibshirani (1993, p. 174). Chen and Hall (1993) also showed that the inability of (1.4) to attain $1 - 2\alpha$ with an $O(n^{-1})$ error due to the discreteness of the binomial distribution can be overcome by using a smoothed version of empirical likelihood. An alternative method to achieve a coverage probability of $1 - 2\alpha + O(n^{-1})$ was proposed by Beran and Hall (1993) who used convex combinations of sample quantiles to develop interpolated confidence intervals. Recently Ho and Lee (2005) made use of smoothed bootstrap iterations to achieve more accurate one-sided coverage errors of the bootstrap percentile interval. Their method, however, is very computationally intensive and involves an additional layer of bootstrapping to determine the bandwidth used to smooth the empirical distribution.

For censored survival data without covariates, Li et al. (1996) made use of empirical likelihood to construct confidence bands for ξ_p, jointly in $p_1 \leq p \leq p_2$. Their results on coverage probabilities are based on weak convergence and do not provide convergence rates of the kind in Chen and Hall (1993). They have, however, not smoothed the empirical likelihood function, nor have they compared the empirical likelihood approach with other test-based methods to construct confidence intervals for ξ_p when the t_i are subject to censoring. These alternative test-based intervals date back to Brookmeyer and Crowley (1982) who invert a generalized sign test, leading to an approximate $1 - 2\alpha$ confidence set of the form

$$\{t : |\widehat{S}(t) - 1/2| \leq z_{1-\alpha}\widehat{\sigma}(t)\} \tag{1.5}$$

for the median $\xi_{1/2}$, where $\widehat{S}(t)$ is the Kaplan–Meier estimator of the survival function, $\widehat{\sigma}(t)$ is the estimated standard error of $\widehat{S}(t)$ and z_q denotes the qth quantile of the standard normal distribution. Instead of using the normal approximation, Strawderman et al. (1997) use Edgeworth expansions for the Studentized cumulative hazard function to derive more accurate test-based confidence limits for ξ_p.

In this paper we develop a new method to construct confidence intervals and confidence bands for the quantile $\xi_p(x)$ in the proportional hazards model (1.1). Unlike the methods of Dabrowska and Doksum (1987) and Burr and Doss (1993) that use $\{\widehat{\xi}_p(x) - \xi_p(x)\}/\widehat{\mathrm{se}}_p$ as an approximate pivot, we use a test-based approach, using $\widehat{\Lambda}(t|x)$ to test if $\Lambda(t|x) = \log(p^{-1})$, where $\widehat{\Lambda}(t|x) = \widehat{\Lambda}(t)\exp(\widehat{\beta}^T x)$ and $\widehat{\Lambda}(t)$ is Breslow's estimator of the baseline cumulative hazard function $\Lambda(t)$. Instead of using the normal approximation or its second-order refinement as in Strawderman et al. (1997) to find the quantiles of the test statistic, we use the bootstrap method to evaluate the quantiles of an approximate pivot obtained by Studentizing the test statistic. The details are described in Sect. 2 which also extends this approach to confidence bands for $\xi_p(x)$, jointly in x belonging to some given set K. The advantages of the proposed procedure are demonstrated in the asymptotic theory in Sect. 2 and the simulation studies in Sect. 3. In Sect. 3 we also apply the proposed methods to construct confidence intervals for median survival given a patient's covariates from the Stanford Heart Transplant data, and compare our results with those of Burr and Doss (1993). Section 4 concludes with some remarks and further discussion.

2 Methodology

2.1 A new test-based bootstrap confidence interval

In this section we propose a new test-based confidence interval for the pth quantile $\xi_p(x)$ given a subject's covariate vector x in the Cox model, and provide an associated algorithm for computing the endpoints of the interval. An obvious generalization of the Brookmeyer–Crowley confidence interval (1.5) for $\xi_{1/2}$ to $\xi_p(x)$ in the Cox model is

$$\{t : |\widehat{S}(t|x) - (1-p)| \leq z_{1-\alpha}\widehat{\sigma}(t|x)\}, \tag{2.1}$$

where $\widehat{\sigma}^2(t|x)$ is the asymptotic variance of

$$\widehat{S}(t|x) = \exp\{-\widehat{\Lambda}(t)e^{\widehat{\beta}^T x}\}, \tag{2.2}$$

in which $\widehat{\beta}$ is the maximizer of (1.2) and $\widehat{\Lambda}$ is Breslow's (1974) estimate of the baseline cumulative hazard function based on $(\tilde{t}_i, \delta_i, x_i)$, $1 \leq i \leq n$. The asymptotic variance formula was derived by Tsiatis (1981) using the delta method; see (2.6) below for its consistent estimate $\widehat{\sigma}^2(t|x)$. Note that this asymptotic variance is a nonlinear function of the asymptotic covariance matrix of $(\widehat{\Lambda}(t) - \Lambda(t), (\widehat{\beta} - \beta)^T x)$. Although $\widehat{S}(t|x)$ takes values in [0, 1], $(\widehat{\beta} - \beta)^T x$ does not have such constraints and its variance in finite samples can be substantial. Moreover, the normal approximation to $|\widehat{S}(t|x) - S(t|x)|/\widehat{\sigma}(t|x)$ used in (2.1) may be inadequate when the sample size is not large enough; in particular, its symmetry about $S(t|x)$ fails to incorporate skewness that is especially relevant for censored data.

Instead of using $\widehat{S}(t|x) - (1-p)$ as the test statistic, we use the logarithmic transformation to transform it into $\widehat{\Lambda}(t|x) - \log(1-p)^{-1}$. An advantage of this transformation is that unlike $\widehat{S}(t|x)$, $\widehat{\Lambda}(t|x)$ is no longer constrained to belong to [0, 1] and therefore the variability due to $(\widehat{\beta} - \beta)^T x$ in its asymptotic variance formula can be more compatible with its magnitude. Another advantage is that the asymptotic variance of $\widehat{S}(t|x)$ involves further linear approximation around $\widehat{\Lambda}(t|x)$. In fact, after deriving the asymptotic variance $v(t|x)$ of

$$\widehat{\Lambda}(t|x) = \widehat{\Lambda}(t)\exp(\widehat{\beta}^T x) \tag{2.3}$$

from the asymptotic covariance matrix of $(\widehat{\Lambda}(t) - \Lambda(t), (\widehat{\beta} - \beta)^T x)$, Tsiatis (1981) used it to derive the asymptotic variance of $\widehat{S}(t|x)$ via the nonlinear transformation $\widehat{S}(t|x) = e^{-\widehat{\Lambda}(t|x)}$. Letting $x_i = (x_{i1}, \ldots, x_{ik})^T$ and $a = (a_1, \ldots, a_k)^T$, define

$$
\begin{aligned}
W(t) &= \sum_{j:\tilde{t}_j \geq t} \exp(\widehat{\beta}^T x_j), \quad W_l(t) = \sum_{j:\tilde{t}_j \geq t} x_{jl} \exp(\widehat{\beta}^T x_j), \\
Q_l(t,a) &= \sum_{i:\tilde{t}_i \leq t} \delta_i \{W_l(\tilde{t}_i)/W(\tilde{t}_i) - a_l\}/W(\tilde{t}_i),
\end{aligned}
\tag{2.4}
$$

and $Q(t,a) = (Q_1(t,a), \ldots, Q_k(t,a))^T$. Replacing the unknown parameters in $v(t|x)$ by their consistent estimates yields

$$
\widehat{v}(t|x) = e^{2\widehat{\beta}^T x} \left\{ \sum_{i:\tilde{t}_i \leq t} \delta_i / W^2(\tilde{t}_i) + (Q(t,x))^T (-\ddot{l}(\widehat{\beta}))^{-1} Q(t,x) \right\}, \tag{2.5}
$$

which in turn yields

$$
\widehat{\sigma}^2(t|x) = (\widehat{S}(t|x))^2 v(t|x), \tag{2.6}
$$

by applying the delta method to the transformation $\widehat{S}(t|x) = e^{-\widehat{\Lambda}(t|x)}$; see Tsiatis (1981).

Instead of the normal quantiles $z_{1-\alpha}$ and z_α $(= -z_{1-\alpha})$ used in (2.1), we approximate the αth and $(1-\alpha)$th quantiles $c_\alpha(t)$ and $c_{1-\alpha}(t)$ by the quantiles $\widehat{c}_\alpha(t)$ and $\widehat{c}_{1-\alpha}(t)$ of the bootstrap distribution of $\{\widehat{\Lambda}(t|x) - \Lambda(t|x)\}/\widehat{v}^{\frac{1}{2}}(t|x)$. Define the test-based confidence set

$$
T = \{t : \widehat{c}_\alpha(t) \leq [\widehat{\Lambda}(t|x) - \log(1-p)^{-1}]/\widehat{v}^{\frac{1}{2}}(t|x) \leq \widehat{c}_{1-\alpha}(t)\} \tag{2.7}
$$

for the pth quantile $\xi_p(x)$ at a given covariate vector x.

2.2 Asymptotic theory

When there are no covariates, Lai and Wang (1993) have derived Edgeworth expansions for the sampling distribution and also for the bootstrap distribution of $\{\widehat{\Lambda}(t) - \Lambda(t)\}/\widehat{v}^{\frac{1}{2}}(t)$. In the Cox regression model with univariate covariates, Gu (1992) has derived an Edgeworth expansion, with $o(n^{-1/2})$ error, for $Z := (-\ddot{l}(\widehat{\beta}))^{1/2}(\widehat{\beta} - \beta)$ and also for its bootstrap counterpart $Z^* := (-\ddot{l}^*(\widehat{\beta}^*))^{1/2}(\widehat{\beta}^* - \widehat{\beta})$ under certain regularity conditions; his arguments can be readily extended to multidimensional covariates. His derivation involves showing that Z and Z^* are asymptotic U-statistics (see Sect. 2 of Lai and Wang 1993) and applying Helmers' (1991) result for U-statistics. Since Breslow's estimate of the baseline hazard function has the form $\widehat{\Lambda}(t) = \sum_{i:\tilde{t}_i \leq t}\{\delta_i / \sum_{j:\tilde{t}_j \geq \tilde{t}_i} \exp(\widehat{\beta}^T x_j)\}$, arguments similar to those in Example 1 of Lai and Wang (1993) can be used to show that $\widehat{\Lambda}(t) - \Lambda(t)$ is an asymptotic U-statistic. Since $\widehat{\Lambda}(t|x) = e^{\widehat{\beta}^T x}\widehat{\Lambda}(t)$, arguments similar to those in the proof of Lemma 4.4 of Gu (1992) and in Example 2 of Gross and Lai (1996) can be used to prove that $\{\widehat{\Lambda}(t|x) - \Lambda(t|x)\}/\widehat{v}^{\frac{1}{2}}(t|x)$ is an asymptotic U-statistic that has an Edgeworth expansion with $o(n^{-1/2})$ error; the $o(n^{-1})$ error in Lai and Wang (1993) and Gross and Lai (1996) requires stronger assumptions than those in Gu (1992).

As in Gu (1992), we assume here the following regularity conditions:

(A1) (x_i, t_i, c_i) are i.i.d., x_i is bounded, and t_i and c_i are conditionally independent given x_i.

(A2) Λ has a continuous derivative λ.
Moreover, following Gu (1992), we assume that $\|\beta\| < B$ for some known B and that
(A3) $P\{\tilde{t}_i \geq \tau\} > 0$
for some known $\tau > \xi_p(x)$, and redefine (1.2) by

$$\ell(\beta) = \sum_{i:\tilde{t}_i \leq \tau} \delta_i \left\{ \beta^T x_i - \log\left(\sum_{j:\tilde{t}_j \geq \tilde{t}_i} \exp(\beta^T x_j) \right) \right\},$$

so that $\widehat{\beta}$ is the maximizer of this modification of (1.2) within the bounded set $\{b : \| b \| \leq B\}$. Since $\tau > \xi_p(x)$, we can also modify the confidence set T, defined by (2.7), by restricting it within $\{t : t \leq \tau\}$ so that the results of Lai and Wang (1993) on asymptotic U-statistics and Edgeworth expansions can be applied to $\widehat{\Lambda}(t) - \Lambda(t)$ for every $t \in T$. In addition, assume as in Gu (1992) that

(A4) $\int_0^\tau \{\alpha_2(t) - \alpha_1(t)\, \alpha_1^T(t)/\alpha_0(t)\}\lambda(t)\, dt$ is positive definite,

where $\alpha_k(t) = E(x^k e^{\beta^T x} I_{\{\tilde{t} \geq t\}})$ for k = 0, 1, 2, with $x^0 = 1$ and $x^2 = xx^T$.

Under these assumptions, not only does $\{\widehat{\Lambda}(t|x) - \Lambda(t|x)\}/\hat{v}^{\frac{1}{2}}(t|x)$ have an Edgeworth expansion with $o(n^{-1/2})$ error, but the coefficients of this Edgeworth expansion also differ from those of the bootstrap counterpart $\{\widehat{\Lambda}^*(t|x) - \widehat{\Lambda}(t|x)\}/\widehat{v}^{*\frac{1}{2}}(t|x)$ by $o_p(n^{-1/2})$ by standard arguments; see Theorem 3.2 of Gu (1992). Hence

$$\widehat{c}_\alpha(t) - c_\alpha(t) = o_p(n^{-1/2}), \quad \widehat{c}_{1-\alpha}(t) - c_{1-\alpha}(t) = o_p(n^{-1/2}) \tag{2.8}$$

for every fixed t. Applying (2.8) and an argument similar to that of Hall (1992, Sect. 5.3) then yields from (2.7) that

$$\begin{aligned} P\{\xi_p(x) \in T\} &= P\{\widehat{c}_\alpha(\xi_p(x)) \leq [\widehat{\Lambda}(\xi_p(x)|x) - \Lambda(\xi_p(x)|x)]/\widehat{v}^{\frac{1}{2}}(\xi_p(x)|x) \leq \widehat{c}_{1-\alpha}(\xi_p(x))\} \\ &= 1 - 2\alpha + o(n^{-1/2}). \end{aligned} \tag{2.9}$$

2.3 Computation of confidence limits

The set (2.7) may not be an interval, as has already been noted by Brookmeyer and Crowley (1982, p. 32) for their test-based confidence set (1.5) when there are no covariates. In practice, it often suffices to give only the upper and lower limits of (2.7), thereby obtaining a confidence interval. Let $q = \alpha$ or $1 - \alpha$. Note that for fixed x, the cumulative hazard function $\widehat{\Lambda}$ is a step function with jumps at the uncensored ($\delta_i = 1$) observations $\tilde{t}_i$, and so is the function $\widehat{v}$. The jumps at the uncensored $\tilde{t}_i$'s also cause discontinuities of $\widehat{c}_q$ at these points. Let $[\widetilde{\Lambda}(\cdot|x) - \log(1-p)^{-1}]/\widetilde{v}^{\frac{1}{2}}(\cdot|x) - \widetilde{c}_q(\cdot)$ denote the modification of $[\widehat{\Lambda}(\cdot|x) - \log(1-p)^{-1}]/\widehat{v}^{\frac{1}{2}}(\cdot|x) - \widehat{c}_q(\cdot)$ that linearly interpolates between the corresponding values at two adjacent uncensored $\tilde{t}_i$'s.

Suppose the covariates x_i are independent and identically distributed, as is often the case in randomized clinical trials. Then the bootstrap distribution of the asymptotic pivot $\{\widehat{\Lambda}(t|x) - \Lambda(t|x)\}/(\widehat{v}(t|x))^{\frac{1}{2}}$ can be evaluated by resampling from $\{(\tilde{t}_i, \delta_i, x_i) : 1 \le i \le n\}$ to obtain B bootstrap samples $\{(\tilde{t}_i^*, \delta_i^*, x_i^*)_b, 1 \le i \le n\}, 1 \le b \le B$. At each given value of t that will be specified below, $\omega_b^*(t) := \{\widehat{\Lambda}_b^*(t|x) - \widehat{\Lambda}(t|x)\}/(\widehat{v}_b^*(t|x))^{\frac{1}{2}}$ is computed from the bth bootstrap sample, and the αth and $(1-\alpha)$th quantiles of $\{\omega_1^*(t), \ldots, \omega_B^*(t)\}$ are computed to yield $\widehat{c}_\alpha(t)$ and $\widehat{c}_{1-\alpha}(t)$. We can use the following iterative procedure to choose the values of t, belonging to the ordered set U of uncensored $\tilde{t}_i$'s, at which $\widehat{c}_\alpha(t)$ or $\widehat{c}_{1-\alpha}(t)$ is computed. For definiteness, we consider $\widehat{c}_\alpha(t)$. The objective of the iterative procedure is to solve the equation $g(t) = 0$, where

$$g(t) = \{\widetilde{\Lambda}(t|x) - \log(1-p)^{-1}\}/\widetilde{v}^{\frac{1}{2}}(t|x) - \widetilde{c}_\alpha(t). \tag{2.10}$$

Let a be the smallest and b be the largest element of U. With $g(a) < 0$ and $g(b) > 0$, we can use the bisection method, to find two adjacent elements of U where g changes sign. Then we either linearly interpolate between these two points to find the solution of $g(t) = 0$ or simply take the larger element to be the confidence limit. Note that this procedure can also be used to compute test-based bootstrap confidence intervals for the quantiles ξ_p in the absence of covariates and also in the case of complete i.i.d. observations, which we study in Sect. 3.1.1.

Burr and Doss (1993) use another resampling scheme under the assumption that the censoring variables c_i have the same distribution function C. Let $\widehat{C}$ be the Kaplan–Meier estimate of C. A bootstrap sample is of the form $\{(\tilde{t}_i^*, \delta_i^*, x_i) : 1 \le i \le n\}$, where $\tilde{t}_i^* = \min(t_i^*, c_i^*)$ and $\delta_i^* = I_{\{t_i^* \le c_i^*\}}$, in which c_i^* is generated from $\widehat{C}$ and t_i^* is generated from $\widehat{S}(\cdot|x_i)$ independently of c_i^*. This resampling scheme does not need the x_i to be identically distributed but assumes the c_i to be identically distributed instead.

2.4 Extension to confidence bands

Let K be a compact subset of the covariate space. Noting that $\{\sqrt{n}(\widehat{\xi}_p(x) - \xi_p(x)), x \in K\}$ converges weakly to a Gaussian process indexed by $x \in K$ as $n \to \infty$, Burr and Doss (1993) used

$$\sqrt{n} \max_{x \in K} |\widehat{\xi}_p(x) - \xi_p(x)|/\widehat{\text{se}}_p(x) \tag{2.11}$$

as an approximate pivot to construct bootstrap confidence bands for $\{\xi_p(x), x \in K\}$. We can also modify the approach in Sect. 2.1 to construct test-based bootstrap confidence bands as follows. Let d_α denote the αth quantile of the distribution of

$$\min_{x \in K}[\widehat{\Lambda}(\xi_p(x)|x) - \log(1-p)^{-1}]/\widehat{v}^{\frac{1}{2}}(\xi_p(x)|x), \tag{2.12}$$

noting that $\Lambda(\xi_p(x)|x) = \log(1-p)^{-1}$. We can estimate d_α by the αth quantile $\widehat{d}_\alpha$ of the bootstrap distribution of $\min_{x \in K}[\widehat{\Lambda}^*(\widehat{\xi}_p(x)|x) - \widehat{\Lambda}(\widehat{\xi}_p(x)|x)]/(\widehat{v}^*(\widehat{\xi}_p(x)|x))^{\frac{1}{2}}$; see the last two paragraphs of Sect. 2.1. Similarly use the bootstrap quantile $\widehat{d}'_{1-\alpha}$ to estimate the $(1-\alpha)$th quantile $d'_{1-\alpha}$ of

$$\max_{x \in K}[\widehat{\Lambda}(\xi_p(x)|x) - \log(1-p)^{-1}]/\widehat{v}^{\frac{1}{2}}(\xi_p(x)|x). \tag{2.13}$$

With the same notation as that in (2.7), let $T_x = \{t : \widehat{d}_\alpha \le [\widetilde{\Lambda}(t|x) - \log(1-p)^{-1}]/\widetilde{v}^{\frac{1}{2}}(\xi_p(x)|x) \le \widehat{d}'_{1-\alpha}\}$. Then $\{T_x, x\in K\}$ is a confidence band for $\{\xi_p(x), x\in K\}$ satisfying

$$\begin{aligned} & P\{\xi_p(x) \in T_x \text{ for all } x \in K\} \\ =& P\{\widehat{d}_\alpha \le \min_{x\in K}[\widetilde{\Lambda}(\xi_p(x)|x) - \log(1-p)^{-1}]/\widetilde{v}^{\frac{1}{2}}(\xi_p(x)|x), \\ & \widehat{d}'_{1-\alpha} \ge \max_{x\in K}[\widetilde{\Lambda}(\xi_p(x)|x) - \log(1-p)^{-1}]/\widetilde{v}^{\frac{1}{2}}(\xi_p(x)|x)\} = 1 - 2\alpha + O(n^{-1/2}), \end{aligned} \tag{2.14}$$

under the same regularity conditions as those for (2.9).

3 Numerical examples

3.1 Simulation studies

This subsection contains simulation studies of the coverage errors of the test-based confidence set (2.7) with $p = \frac{1}{2}$ and compares them with those of Dabrowska and Doksum (1987) and Burr and Doss (1993). While Dabrowska and Doksum have described their procedure explicitly for us to implement in the comparative study in Sect. 3.2, Burr and Doss (1993, p. 1333) "use the biweight kernel and choose bin width subjectively" in the numerical studies of their procedure, for which there are many possible choices of the kernel and the smoothing parameter in estimating λ. To simplify matters, we assume λ to be known in the estimation of $\widehat{\mathrm{se}}_p(x)$ (see the first two paragraphs of Sect. 1) for their bootstrap-*t* confidence intervals; this circumvents issues concerning how the bandwidth and kernel should be chosen for their procedure to compare with ours which does not require kernel smoothing. In Sect. 3.1.1 we simplify the simulation study even further by considering the case in which covariates are absent so that the problem reduces to interval estimation of the median based on a sample of i.i.d. t_i when there is no censoring, or on $(\widetilde{t}_i, \delta_i)$ when there are censoring variables c_i.

3.1.1 Case without covariates

We first consider the case where there is no censoring. In this case, a test-based confidence set of the type (2.7) can be re-expressed in the form

$$\{t : \widehat{w}_\alpha(t) \le \sqrt{n}[\widehat{F}(t) - 1/2]/[\widehat{F}(t)(1 - \widehat{F}(t))]^{1/2} \le \widehat{w}_{1-\alpha}(t)\}, \tag{3.1}$$

where $w_q(t)$ denotes the qth quantile of the Studentized variate $\sqrt{n}[\widehat{F}(t) - F(t)]/[\widehat{F}(t)(1 - \widehat{F}(t))]^{1/2}$ and $\widehat{w}_q(t)$ denotes the estimate of $w_q(t)$ via bootstrap resampling. The coverage errors of the confidence limits of (3.1) computed by the algorithm in Sect. 2.3 are given in Table 1a,b for n = 30, 100 and for the three distributions considered in the simulation study of Ho and Lee (2005): standard normal F, double exponential with density function $f(x) = e^{-|x|}/2$, and lognormal F which is the distribution function of $\exp\{N(0,1)\}$. As in Ho and Lee, α = 5% and each result is based on 1,000 simulations. Moreover, 1,000 bootstrap samples are used to compute the bootstrap quantiles.

Table 1 Coverage errors (in %) of confidence intervals for median of three distributions: Lognormal, Exponential (or Double Exponential when there is no censoring), Weibull (or Normal when there is no censoring)

Interval	Lognormal			(Double) Exponential			(Normal) Weibull		
	Lower	Upper	Total	Lower	Upper	Total	Lower	Upper	Total
(a) n = 30, no censoring									
(3.1)	4.5	5.0	9.5	5.0	4.4	9.4	4.5	5.0	9.5
BeHa	5.1	5.0	10.1	5.9	4.9	10.8	5.1	5.0	10.1
SmoEL	5.2	5.5	10.7	6.2	5.7	11.9	4.9	5.7	10.6
Ho-Lee	3.6	6.6	10.2	4.0	5.4	9.4	3.7	7.3	11.0
Boot-*t*	2.6	8.9	11.5	2.8	3.0	5.8	5.4	5.5	10.9
(b) n = 100, no censoring									
(3.1)	5.1	5.0	10.1	5.1	4.5	9.6	5.0	5.2	10.2
BeHa	4.6	4.9	9.5	4.6	4.3	8.9	4.6	5.0	9.6
SmoEL	5.3	5.5	10.8	5.3	4.2	9.5	4.9	5.6	10.5
Ho-Lee	4.9	5.9	10.8	4.2	5.3	9.5	5.1	6.0	11.1
Boot-*t*	3.0	9.4	12.4	2.9	3.6	6.5	4.5	6.1	10.6
(c) n = 60, $C\sim$ Exp(1/4)									
(3.2)	5.4	3.5	8.9	4.8	4.5	9.3	4.3	4.5	8.8
BrCr	6.0	6.6	12.6	6.0	5.8	11.8	5.7	5.7	11.4
SPW	7.3	3.7	11.0	5.8	4.6	10.4	6.3	4.5	10.8
Boot-*t*	1.8	11.9	13.7	2.1	7.0	9.1	2.0	10.1	12.1
(d) n = 100, $C\sim$ Exp(1/2)									
(3.2)	5.1	4.0	9.1	4.6	5.8	10.4	5.1	4.1	9.2
BrCr	6.1	5.8	11.9	5.4	5.8	11.2	6.1	4.8	10.9
SPW	6.9	3.9	10.8	6.2	4.7	10.9	5.9	3.3	9.2
Boot-*t*	2.0	10.2	12.2	2.8	6.5	9.3	2.5	8.8	11.3

Following Burr and Doss (1993), a bootstrap-*t* confidence interval in the present setting without censoring and covariates uses $\{\text{med}(\widehat{F}) - \text{med}(F)\}/\widehat{\text{se}}$ as an approximate pivot and bootstrap resampling to estimate its αth and $(1-\alpha)$th quantiles. As pointed out in (1.3), the asymptotic standard error is $\{2\sqrt{n}f(\xi_{1/2})\}^{-1}$ and its estimate requires a density estimator $\widehat{f}$. To simplify matters in comparing this method with (3.1), we consider a more favorable version of the method that can use the true f to define $\widehat{\text{se}} = \{2\sqrt{n}f(\widehat{\xi}_{1/2})\}^{-1}$. Parts (a) and (b) of Table 1 consider these bootstrap-*t* (abbreviated by Boot-*t*) confidence intervals. Also given for comparison are the coverage errors, taken from Table 3 of Ho and Lee (2005), of the confidence intervals of Ho and Lee, Beran and Hall (1993, abbreviated by BeHa) and Chen and Hall (1993) who construct the confidence intervals by using smoothed empirical likelihood (abbreviated by SmoEL). The results of Table 1a,b show that (3.1) and the Beran–Hall confidence limits have coverage errors that are close to the nominal value of 5%. The smoothed empirical likelihood and Ho–Lee confidence limits also perform well except for a couple of cases. In contrast, the bootstrap-*t* confidence intervals based on the approximate pivot $\{\text{med}(\widehat{F}) - \text{med}(F)\}/\widehat{\text{se}}$ have coverage errors that are markedly different from 5% in most cases.

Parts (c) and (d) of Table 1 consider the case where there are independent censoring variables c_i that have a common distribution C which is assumed to be exponential with intensity parameter 1/4 or 1/2. The three baseline survival functions considered are Lognormal, Exponential with intensity parameter 1, and Weibull with scale parameter 1 and shape parameter 0.7. The observations are

$(\widetilde{t}_i, \delta_i), i = 1, \ldots, n$. Let $\widehat{\sigma}(t)$ be the estimated standard error of $\widehat{S}(t)$ given by Greenwood's formula; see Andersen et al. (1993, p. 258). Since we do not have to estimate β and to adjust for the variability of $\widehat{\beta} - \beta$ here, we can use $\widehat{S}(t) - 1/2$ as in the Brookmeyer–Crowley interval (1.5) (abbreviated by BrCr) instead of transforming it to $\widehat{\Lambda}(t) - \log 2$ as in (2.7). This leads to a test-based confidence set of the form

$$\{t : \widehat{w}_\alpha(t) \leq [\widehat{S}(t) - 1/2]/\widehat{\sigma}(t) \leq \widehat{w}_{1-\alpha}(t)\}, \tag{3.2}$$

which we implement by the procedure in Sect. 2.3. Besides (1.5) and (3.2), parts (c) and (d) of Table 1 also consider the bootstrap-*t* confidence intervals (abbreviated by Boot-*t*) of the type considered by Burr and Doss (1993). Since there are no covariates, the asymptotic standard error of $\widehat{\xi}_{1/2} - \xi_{1/2}$ is simpler than that in their Theorem 1. It is equal to $\sigma(\xi_{1/2})/f(\xi_{1/2})$, where $\sigma^2(t)$ is the asymptotic variance of $\widehat{S}(t)$. Assuming f to be known, the estimated standard error of $\widehat{\xi}_{1/2} - \xi_{1/2}$ is $\widehat{\text{se}} = \widehat{\sigma}(\widehat{\xi}_{1/2})/f(\widehat{\xi}_{1/2})$, where $\widehat{\sigma}(t)$ is the estimated standard error of $\widehat{S}(t)$ given by Greenwood's formula; see Andersen et al. (1993, pp. 257, 258, 276 which assume t_i to be nonnegative and use $\sigma^2(t)$ to denote the asymptotic variance of $\widehat{S}(t)/S(t)$ instead).

Strawderman et al. (1997) gave a review of confidence intervals for ξ_p based on censored observations in the earlier literature and proposed a new test-based confidence interval for ξ_p, which uses an Edgeworth expansion for $[\widehat{\Lambda}(t) - \Lambda(t)]/\widehat{v}^{\frac{1}{2}}(t)$ to improve the normal approximation and which they denote by I_2. Their simulation study shows that I_2, which Table 1 refers to as SPW (abbreviation for the authors), "is superior to all others considered in terms of maintaining coverage accuracy." The results in Table 1c,d, however, contain cases where SPW has worse coverage accuracy than the Brookmeyer–Crowley confidence limits, and show that (3.2) has better coverage accuracy. Parts (c) and (d) of Table 1 also show that the bootstrap-*t* confidence intervals based on the approximate pivot $(\widehat{\xi}_{1/2} - \xi_{1/2})/\widehat{\text{se}}$ have poor coverage accuracy.

3.1.2 *Case with covariates*

Consider the proportional hazards model (1.1) in which the baseline survival function is that of a Weibull distribution with scale parameter 1 and shape parameter θ. The covariates x_i are independent and uniformly distributed in [0, 1] and $\beta = 1$. The censoring times c_i are i.i.d. exponential with intensity parameter 0.5, 1 or 2. The sample size is 80. Table 2 considers three different values of the Weibull shape parameter θ and gives the censoring proportion π_a for each value a of the intensity parameter of the exponential censoring distribution. Besides the coverage errors of our proposed test-based bootstrap confidence intervals, Table 2 also gives those of the confidence intervals of Dabrowska and Doksum (in brackets) and those of the test-based confidence intervals using the normal approximation (in parentheses). Each result is based on 1,000 simulations, and 1,000 bootstrap samples are used to compute the bootstrap quantiles for our proposed procedure. Table 2 shows that the coverage errors of our proposed confidence intervals are close to the nominal value $\alpha = 5\%$ but those using the normal approximation (instead of the bootstrap) or the Dabrowska–Doksum method differ markedly from 5%.

Table 2 Coverage errors (in %) of (2.7), of its normal approximation counterpart (in parentheses) and of the Dabrowska–Doksum confidence intervals (in brackets) for median survival

x	C~Exp(0.5)			C~Exp(1)			C~Exp(2)		
	Lower	Upper	Total	Lower	Upper	Total	Lower	Upper	Total
(a) $\theta = .7$,	$\pi_{0.5} = 22\%$		$\pi_1 = 33\%$	$\pi_2 = 47\%$					
.25	4.7	5.2	9.9	4.2	5.4	9.6	4.9	4.8	9.7
	(8.3)	(3.5)	(11.8)	(7.1)	(3.4)	(10.5)	(8.9)	(2.2)	(11.1)
	[3.3]	[3.2]	[6.5]	[4.1]	[2.8]	[6.9]	[3.2]	[4.5]	[7.7]
.5	4.1	5.8	9.9	5.8	5.5	11.3	5.1	3.8	8.9
	(7.4)	(3.8)	(11.2)	(7.5)	(3.9)	(11.4)	(8.8)	(2.4)	(11.2)
	[4.3]	[4.2]	[8.5]	[3.7]	[3.8]	[7.5]	[4.1]	[2.6]	[6.7]
.75	5.4	5.1	10.5	5.0	5.4	10.4	4.7	5.6	10.3
	(8.1)	(2.7)	(10.8)	(7.8)	(2.9)	(10.7)	(8.2)	(1.5)	(9.7)
	[3.3]	[3.3]	[6.6]	[4.1]	[3.2]	[7.3]	[2.1]	[4.4]	[6.5]
(b) $\theta = 1$,	$\pi_{0.5} = 24\%$		$\pi_1 = 38\%$	$\pi_2 = 54\%$					
.25	5.3	5.4	10.7	4.9	6.0	10.9	5.4	6.1	11.5
	(8.0)	(2.9)	(10.9)	(8.7)	(1.9)	(10.6)	(10.7)	(1.9)	(12.6)
	[3.1]	[3.7]	[6.8]	[2.8]	[4.8]	[7.6]	[2.7]	[3.7]	[6.4]
.5	5.5	5.3	10.8	5.7	5.1	10.8	5.4	4.9	10.3
	(7.0)	(1.9)	(8.9)	(8.3)	(3.5)	(11.8)	(8.0)	(3.0)	(11.0)
	[3.3]	[2.9]	[6.2]	[4.1]	[2.5]	[6.6]	[2.9]	[3.8]	[6.8]
.75	4.6	5.0	9.6	5.5	5.1	10.6	4.8	5.3	10.1
	(8.4)	(2.2)	(10.6)	(9.4)	(2.9)	(12.3)	(10.4)	(1.7)	(12.1)
	[2.9]	[4.0]	[6.9]	[4.0]	[3.0]	[7.0]	[3.2]	[4.1]	[6.3]
(c) $\theta = 1.3$,	$\pi_{0.5} = 25\%$		$\pi_1 = 41\%$	$\pi_2 = 60\%$					
.25	4.6	5.1	9.7	4.4	5.0	9.4	3.8	6.6	10.4
	(8.1)	(3.4)	(11.5)	(10.1)	(2.5)	(12.6)	(11.3)	(1.4)	(12.7)
	[4.0]	[2.9]	[6.9]	[3.4]	[2.7]	[6.1]	[2.5]	[3.5]	[6.0]
.5	5.3	5.9	11.2	5.1	5.7	10.8	4.6	6.5	11.1
	(8.6)	(3.6)	(12.2)	(8.9)	(3.0)	(11.9)	(8.4)	(2.6)	(11.0)
	[3.3]	[3.9]	[7.2]	[3.2]	[2.8]	[6.0]	[3.1]	[2.4]	[5.5]
.75	5.0	5.3	10.3	5.3	5.4	10.7	4.2	5.5	9.7
	(7.1)	(1.5)	(8.6)	(8.4)	(2.5)	(10.9)	(8.6)	(1.1)	(9.7)
	[3.8]	[3.1]	[6.9]	[3.2]	[3.6]	[6.8]	[2.5]	[2.4]	[4.9]

Besides the confidence set (2.7), this simulation study also considers the coverage errors of 90% confidence bands of the type in Sect. 2.3 for $.25 \le x \le .75$. The coverage errors (in %) under the three censoring patterns are 10.5, 9.6, 9.2 for $\theta = .7$; 11.4, 10.5, 9.4 for $\theta = 1$; and 11.1, 10.5, 10.0 for $\theta = 1.3$, in close agreement with the nominal coverage error of 10%.

3.2 Application to Stanford Heart Transplant Data

We illustrate the methods in Sect. 2.1 for constructing confidence intervals on the 1980 version of the Stanford Heart Transplant (SHT) data as given in Miller and Halpern (1982), who fitted a proportional hazards regression model to the data involving 152 patients that had survived at least 10 days, and who chose quadratic regression of $\log_{10}$ (survival time in days) on age (in years) as the predictor variable for the final model. Burr and Doss (1993, p. 1338) have applied their bootstrap-*t* method to construct confidence intervals and bands for median survival (days in the $\log_{10}$ scale) from these data. In addition, they have also used the limiting Gaussian process for $\{\widehat{\xi}_{1/2}(x) - \xi_{1/2}(x)\}/\widehat{\text{se}}_{1/2}(x)$ to construct confidence intervals and

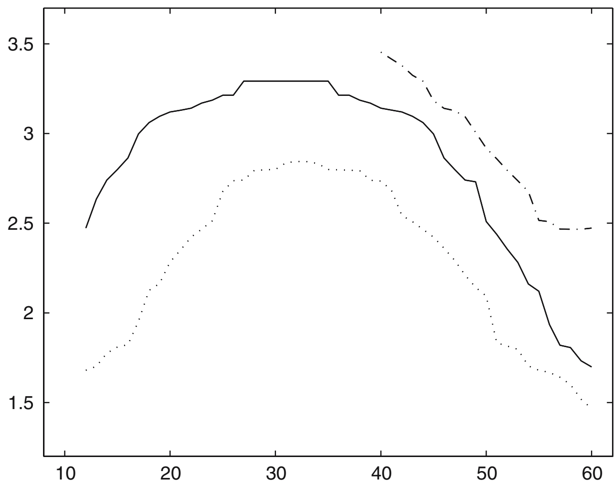

Fig. 1 95% confidence bands for the median survival time of the SHT data. The solid curve represents the estimate of median survival time (days in the $\log_{10}$ scale) as a function of age (in years)

Table 3 95% confidence intervals and bands for median survival (in years) from SHT data

		Age = 38.5				Age = 48.7			
		SP	Boot	$\text{Tb}_{\widehat{\Lambda}}$	$\text{Tb}_{\widehat{S}}$	SP	Boot	$\text{Tb}_{\widehat{\Lambda}}$	$\text{Tb}_{\widehat{S}}$
Interval	L	2.4	2.3	2.8	2.7	.8	.6	.7	.7
	U	6.7	5.6	6.4	6.7	2.8	2.2	2.4	2.5
Band	L	1.6	1.8	1.7	1.7	.5	.3	.4	.4
	U	9.7	9.1	7.9+	7.9+	4.8	6.4	3.4	3.8

confidence bands for $\xi_{1/2}(x)$, simulating the Gaussian process to determine the half-width of the band. Their results at 38.5 and 48.7 years of age for 95% simulated process (SP) and bootstrap-*t* (Boot) confidence bands and intervals are included in Table 3. Also given in Table 3 are the test-based ($\text{Tb}_{\widehat{\Lambda}}$) confidence intervals and bands in Sects. 2.1 and 2.3 that use $\{\widehat{\Lambda}(t|x) - \log(1-p)^{-1}\}/\widehat{v}^{\frac{1}{2}}(t|x)$ as the test statistic. Instead of $\widehat{\Lambda}(t|x)$, an alternative is to use $\{\widehat{S}(t|x) - (1-p)\}/\widehat{\sigma}(t|x)$ as pointed out in the first paragraph of Sect. 2.1, and its associated test-based ($\text{Tb}_{\widehat{S}}$) confidence intervals and bands are also given in Table 3 for comparison. Table 3 shows $\text{Tb}_{\widehat{\Lambda}}$ to yield somewhat shorter confidence intervals and bands than $\text{Tb}_{\widehat{S}}$. Moreover, $\text{Tb}_{\widehat{\Lambda}}$ yields markedly shorter confidence bands than SP and Boot. Figure 1 plots the entire confidence band for $\text{Tb}_{\widehat{\Lambda}}$; note that the upper band ends at $\log_{10}$(2878 days). In view of the interpolation scheme used to evaluate the upper limits of the confidence intervals and bands in Section 2.1 and 2.3, the value of the upper limit is undetermined if it exceeds the largest uncensored observation. Therefore the entry 7.9+ for the upper limit of the confidence band $\text{Tb}_{\widehat{\Lambda}}$ or $\text{Tb}_{\widehat{S}}$ in Table 3 indicates that it exceeds the largest observed survival of 7.9 years, beyond which there are no data to estimate the hazard function nonparametrically.

4 Discussion

An important ingredient in the test-based confidence intervals/bands developed herein for survival quantiles in the Cox regression model is the use of bootstrap quantiles to approximate the quantiles of $\{\widehat{\Lambda}(t|x) - \Lambda(t|x)\}/\widehat{v}^{\frac{1}{2}}(x)$, instead of using the normal approximation (or Edgeworth expansions) as in previous works on test-based confidence intervals for ξ_p (in the absence of covariates) from censored survival data. A novelty here is that we work with $\widehat{\Lambda}(t|x) - \log(1-p)^{-1}$, instead of $\widehat{S}(t|x) - (1-p)$ that has been used by Brookmeyer and Crowley (1982) and subsequent authors for the case without covariates. In the presence of covariates, there is additional variability due to the estimation of the regression parameter β and it is useful to transform $\widehat{S}(t|x)$, which is constrained to belong to [0, 1], to the unconstrained $\widehat{\Lambda}(t|x) - \log(1-p)^{-1}$. This transformation often leads to shorter confidence intervals. Another useful ingredient for implementation is the interpolation scheme in Sect. 2.3. Beran and Hall (1993) have used similar interpolation ideas to circumvent the discreteness of the binomial distribution in constructing confidence intervals for ξ_p from sample quantiles when there are no covariates and no censoring.

References

Andersen PK, Gill RD (1982) Cox's regression model for counting processes: a large sample study. Ann Statist 10:1100–1120

Andersen PK, Borgan O, Gill RD, Keiding N (1993) Statistical models based on counting processes. Springer, New York

Beran R, Hall P (1993) Interpolated nonparametric prediction intervals and confidence intervals. J R Stat Soc Ser B 55:643–652

Breslow NE (1974) Covariance analysis of censored survival data. Biometrics 30:89–99

Brookmeyer R, Crowley J (1982) A confidence interval for the median survival time. Biometrics 38:29–41

Burr D, Doss H (1993) Confidence bands for the median survival time as a function of the covariates in the Cox model. J Am Stat Assoc 88:1330–1340

Chen X, Hall P (1993) Smoothed empirical likelihood confidence intervals for quantiles. Ann Statist 21:1166–1181

Cox DR (1972) Regression models and life-tables (with discussion). J R Stat Soc Ser B 34:187–220

Dabrowska D, Doksum K (1987) Estimates and confidence intervals for median and mean life in the proportional hazard model. Biometrika 74:799–807

Efron B (1979) Bootstrap methods: another look at the jackknife. Ann Statist 7:1–21

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman & Hall, New York

Gross S, Lai TL (1996) Bootstrap methods for truncated and censored data. Statist Sin 6:509–530

Gu M (1992) On the Edgeworth expansion and bootstrap approximation for the Cox regression model under random censorship. Can J Statist 20:399–414

Hall P (1992) The bootstrap and Edgeworth expansion. Springer, New York

Helmers R (1991) On the Edgeworth expansion and the bootstrap approximation for a Studentized *U*-statistic. Ann Statist 19:470–484

Ho Y, Lee M (2005) Iterated smoothed bootstrap confidence intervals for population quantiles. Ann Statist 33:437–462

Lai TL, Wang Q (1993) Edgeworth expansions for symmetric statistics with applications to bootstrap methods. Statist Sin 3:517–542

Li G, Hollander M, McKeague IW, Yang J (1996) Nonparametric likelihood ratio confidence bands for quantile functions from incomplete data. Ann Statist 24:628–640

Miller R, Halpern J (1982) Regression with censored data. Biometrika 69:521–531

Strawderman RL, Parzen MI, Wells MT (1997) Accurate confidence limits for quantiles under random censoring. Biometrics 53:1399–1415

Tsiatis AA (1981) A large sample study of Cox's regression model. Ann Statist 9:93–108