Empirical Likelihood for Right Censored Lifetime Data

Shuyuan He, Wei Liang, Junshan Shen, and Grace Yang

ABSTRACT
When the empirical likelihood (EL) of a parameter \( \theta \) is constructed with right censored data, literature shows that \(-2 \log(\text{empirical likelihood ratio})\) typically has an asymptotic scaled chi-squared distribution, where the scale parameter is a function of some unknown asymptotic variances. Therefore, the EL construction of confidence intervals for \( \theta \) requires an additional estimation of the scale parameter. Additional estimation would reduce the coverage accuracy for \( \theta \). By using a special influence function as an estimating function, we prove that under very general conditions, \(-2 \log(\text{empirical likelihood ratio})\) has an asymptotic standard chi-squared distribution with one degree of freedom. This eliminates the need for estimating the scale parameter as well as eases some of the often demanding computations of the EL method. Our estimating function yields a smaller asymptotic variance than those of Wang and Jing (2001) and Qin and Zhao (2007). Thus, it is not surprising that confidence intervals using the special influence functions give a better coverage accuracy as demonstrated by simulations.

1. Introduction
Let \( Y \) and \( C \) be two independent, nonnegative random variables with respective distribution functions \( F(y) = P[Y \leq y] \) and \( G(s) = P[C \leq s] \). \( Y \) represents a lifetime whose observation is subject to right censoring by \( C \) such that instead of \( Y \) we observe \( Z = \min(Y, C) \) and the indicator \( \delta = 1[Y \leq C] \) of the event \( [Y \leq C] \).

This article is concerned with using the empirical likelihood (EL) method with right-censored data to construct confidence intervals for a parameter \( \theta \), a functional of the distribution \( F \), defined by

\[
\frac{\partial \log(Eg(Y, \theta))}{\partial \theta} = 0, \tag{1.1}
\]

for a specified function \( g(s, \theta) \). Both \( F \) and \( G \) are unknown.

A simple example is \( g(s, \theta) = \xi(s) - \theta \) for some function \( \xi(s) \). Then \( \theta = \mathbb{E}\xi(Y) \). Many other examples of \( g(s, \theta) \) are given in Section 2.

The problem of estimating \( \theta = \mathbb{E}\xi(Y) \) with a sample of \( n \) iid observations of \( (Z, \delta) \) has been studied by many authors using the estimator

\[
\hat{\theta}_n = \int_0^{\infty} \xi(s) d\hat{F}_n(s), \tag{1.2}
\]

where \( \hat{F}_n \) is the Kaplan-Meier (KM) or the product-limit estimator of \( F \) defined in (2.3) below. For continuous \( F \) and \( G \), under the finite second moment condition for \( \xi(s) \),

\[
\int_0^{\infty} \frac{\xi^2(s)}{1 - G(s)} dF(s) < \infty, \tag{1.3}
\]

the asymptotic distribution of \( \sqrt{n}(\hat{\theta}_n - \theta) \), as \( n \to \infty \), is normal \( N(0, \sigma^2) \), see, for example, Yang (1994), where

\[
\sigma^2 = \int_0^{\infty} \frac{\xi(s) - \psi(s)^2}{F(s)G(s)} dF(s),
\]

\[
\psi(s) = \int_s^{\infty} \xi(y) dF(y), s \geq 0, \tag{1.4}
\]

where \( F = 1 - G \) and \( \overline{G} = 1 - G \).

Confidence intervals for \( \theta \) can be constructed using the asymptotic normal distribution \( N(0, \sigma^2) \). To use \( N(0, \sigma^2) \), it is necessary to estimate the unknown variance \( \sigma^2 \) in (1.4). For instance, Stute (1996) proposes a jackknife estimator for \( \sigma^2 \). Although any consistent estimator \( \hat{\sigma}_n^2 \) of \( \sigma^2 \) can be employed, the convergence rate of \( \hat{\sigma}_n^2 \) is generally unknown. Substitution of \( \sigma^2 \) by \( \hat{\sigma}_n^2 \) tends to reduce the coverage accuracy for \( \theta \) as compare to the case of known \( \sigma^2 \). In the example of \( \theta = F(y) \) for a fixed \( y \), the coverage accuracy was studied by Thomas and Grunkemeier (1975), who proposed the EL method and illustrated by simulation its superiority over the method of normal approximation. The ensuing theoretical work on the EL method was carried out by Owen (1988) and followed by a rapid growth of literature. The usefulness of the EL method for constructing confidence interval/regions has been well established in a wide variety of situations, see for example, DiCiccio, Hall, and Romano (1991), Chen (1994), and a standard reference, Owen (2001) and references therein. An article by Zheng, Zhao, and Yu (2012) contains an excellent review of some recent development.

Let \( R(\theta) \) denote the EL ratio function of \( \theta \) and \( l(\theta) = -2 \log R(\theta) \). If \( R(\theta) \) is computed from a “complete” sample of \( n \) iid observations, then in many cases, \( l(\theta) \) has asymptotic \( \chi^2_1 \) distribution. Confidence intervals for \( \theta \) can be constructed using the asymptotic \( \chi^2_1 \) distribution. However, if \( R(\theta) \) is obtained...
from a right-censored sample, except for a few special cases (see, e.g., Owen 2001, Chap. 6), literature shows that the asymptotic distribution of \( l(\theta) \) is typically a scaled \( \chi^2 \), where the scale parameter \( r \) is a function of unknown asymptotic variances. See, for example, Wang and Jing (2001) and Qin and Zhao (2007). This is often also the case in the regression analysis (e.g., Li and Wang 2003), weighted EL analysis of a variety of censoring models and nonparametric problems involving infinite dimensional nuisance parameters in which the asymptotic distribution of the corresponding \( l(\theta) \) involves more complicated distributions of sums of weighted chi-squared random variables with unknown weights (Hjort, McKeague, and van Keilegom 2009). For example, when applying Theorems 2.1 and 2.2 in Hjort, McKeague, and van Keilegom (2009) with parameter \( \theta = \int \xi(t) dF(t) \), it can be seen from their Remark 2.2 that we will have

\[
-2 \log EL_n(\theta_0, \hat{h}) \rightarrow_d (V_1/V_2) \chi^2_1,
\]

where \( V_1 \) is the variance of \( U \) in their Condition A2 and \( V_2 \) in Condition A3. This example shows that the unknown scale \( V_1/V_2 \) remains in the asymptotic distribution. In their Section 3.3 on functionals of survival distributions, applying their Theorems 2.1 and 2.2 one gets the same result obtained by Wang and Jing (2001). In view of a complicated formula for the scale parameter \( r \) (see Remark 3.1) which is computationally demanding, a scale-free chi-squared distribution would be desirable.

For the EL interval estimation of \( \theta \) in (1.1) we propose a method of eliminating the estimation of the scale parameter \( r \). The elimination simplifies some of the intensive computations often required of the EL method for censored data. This approach also unifies the proof of the asymptotic distribution of the corresponding \( l(\theta) \). In addition, simulation shows that the method also gives a larger cover probability of the true \( \theta \) as compared to the results of Wang and Jing (2001) and Qin and Zhao (2007).

Empirical likelihood ratio function \( R(\theta) \) is obtained by utilizing auxiliary information on \( \theta \) through a set of estimating equations. In our investigation, the estimation equation is \( Eg(Y, \theta) = 0 \), where the estimating function \( g(y, \theta) \) is a real-valued function. In the case of complete sample, \( Y_1, \ldots, Y_n \), \( R(\theta) \) is formulated as

\[
R(\theta) = \sup_{\{p_i\}} \left\{ \prod_{i=1}^{n} n p_i \left( \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i g(Y_i, \theta) = 0, p_i \geq 0, i = 1, 2, \ldots, n \right) \right\},
\]

where \( \{p_i, i = 1, \ldots, n\} \) is a multinomial distribution which puts weight \( p_i \) on the observation \( Y_i \), and the constraint \( \sum_{i=1}^{n} p_i g(Y_i, \theta) = 0 \) mimics the estimating equation \( Eg(Y, \theta) = 0 \); see, for example, Owen (2001). In the case of right censoring, we observe not \( Y_1, \ldots, Y_n \), but a right-censored sample of iid random vectors \( (Z_1, \delta_1), \ldots, (Z_n, \delta_n) \). Our approach is to use the \((Z, \delta)\)—sample to construct another set of \( n \) random variables, \( \{W_{ni}, i = 1, \ldots, n\} \), and formulate a modified EL ratio function \( R(\theta) \) as follows,

\[
R(\theta) = \sup_{\{p_i\}} \left\{ \prod_{i=1}^{n} n p_i \left( \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i W_{ni} = 0, p_i \geq 0, i = 1, 2, \ldots, n \right) \right\}.
\]

The \( W_n \) given in (3.7) are constructed in a special way such that the asymptotic distribution of the corresponding \( l(\theta) \), that is, \( l(\theta) = -2 \log R(\theta) \), is the standard \( \chi^2 \) without involving any unknown scale parameter. Our proof of the asymptotic results requires some hard analysis due in part to the fact that the \( W_n \) are dependent random variables.

Our work is motivated by that of Wang and Jing (2001) and Qin and Zhao (2007). They used a modified \( R(\theta) \) similar to (1.6) but with different \( W_{ni} \), say \( V_{ni} \). Their \( V_{ni} \) (given in (3.13)) would result in a scaled \( \chi^2 \) distribution for the asymptotic distribution of the corresponding \( l(\theta) \). The \( W_n \) in (1.6), is based on a special influence function \( W(Z_i, \delta_i, \theta) = \mu_n = \int_0^\infty g(y, \theta) dF_n(y) \) for given \( \theta \). Our results generalize those of Wang and Jing (2001) and Qin and Zhao (2007) under a more relaxed assumption. Namely, throughout the article the only assumption we use is (1.3) which is stated as Assumption A in Section 3.

In comparison with several other recent results, we have the following observations. Encountered with the same challenging or prohibitive computational problem of estimating the scale (or asymptotic variances/covariances) using other existing procedures, Zhou and Jeong (2011) find a way to avoid the estimation of asymptotic variances/covariances. Their method is different from ours. Their parameter of interest is also not identical to ours.

To test the null hypothesis \( H_0 : \theta = Eg(Y) \) for a given \( g \), Zhou and Jeong (2011) put the probability mass, \( p_i \), on the \( Z_i \) which is \( Z_i = \min(Y_i, C_i) \) in our notation and define the maximum EL, EL(KM), to be the maximum value of

\[
EL = \prod_{i=1}^{n} \left( \sum_{j > Z_i} p_j \right)^{1 - b_i},
\]

where EL(KM) is the maximum over all \( p_i \) and attained at the KM estimator for right censored data. Let maxEL be the maximum EL under the extra constraint \( \sum_{i=1}^{n} g(Z_i) p_i = \theta \). These authors formulated the likelihood ratios maxEL/EL(KM) and proved that \( W(\theta) = -2 \log[\max EL/EL(KM)] \) converges to the standard \( \chi^2 \). Pan and Zhou (2002) formulated a different EL ratio \( \alpha CR(\theta) \) and proved that \( -2 \log \alpha CR(\theta) \) converges in distribution to the standard \( \chi^2 \), but for the parameter \( \theta = \int g(t) d\Lambda(t) \) defined in terms of cumulative hazard function \( \Lambda(t) \) of \( F \). There are some recent results in the regression analysis using the EL method. For right censored observations and \( p \)-dimensional regression parameter \( \beta \) in the context of accelerated failure time models, Zhou, Kim and Bathke (2012) formulated certain likelihood ratios as \( R_{xy}(\beta) \). Under \( H_0 : \beta = \beta_0 \), they showed that \( -2 \log R_{xy}(\beta_0) \) converges in distribution to the standard \( \chi^2_p \), and their method is applicable to the related censored quantile regression. In the Bayesian framework, Yang and He (2012) used the EL analysis for quantile regression. The application of our method to censored regression problems will be reported in a future publication.

In our article we use \( R(\theta) \) defined in (1.6). The probability \( p_i \) is put on \( W_{ni} \). Maximization shows that \( R(\theta) \) equals to (3.11). In addition to the consideration of efficient estimation for which
$W_n$ are responsible, (3.11) seems to be easier to investigate than $W(\theta)$ of Zhou and Jeong (2011) in which the KM estimator is used. We prove the weak convergence of $-2 \log R(\theta)$ in Theorem 3.2 for parameter $\theta$ defined by $\int_0^{2} g(s, \theta) dF(s) = 0$. Our method yields an efficient estimation equation in the sense that $W_n$ in (3.2) has the minimum variance $\sigma^2$ as shown in Remark 3.1. The class of parameters considered in this article is more general than that of Pan and Zhou (2002) and Zhou and Jeong (2011). Our method can be used to test and construct confidence intervals for a parameter $\theta$ defined by $Eg(Y, \theta)$ while Zhou and Jeong (2011)'s method cannot.

The EL method based on influence functions was studied by Zheng, Zhao, and Yu (2012). They obtained a chi-squared limit for finite dimensional nuisance parameters, and no proofs were given for infinite dimensional nuisance parameters. Our nuisance parameters $F$ and $G$ are infinite dimensional. For right censored data, the infinite dimensional nuisance parameters $F$ and $G$ involved in the estimating functions cannot be dealt with by their profile EL. The proposed influence function based EL methods provide a way to retain the nonparametric Wilks property of a chi-squared limit.

The article is organized as follows. Preliminaries, notations, examples of $\theta$, $g(y, \theta)$ are given in Section 2. In Section 3, the construction of $W_n$ from $W(Z_i, \delta_i, \theta)$ is carried out. It is shown in Theorem 3.1 that asymptotically $\int_0^\infty g(y, \theta) dF_n(y)$ is a partial sum of $n$ independent influence functions $W(Z_i, \delta_i, \theta)$, or is asymptotically linear. The weak convergence of $-2 \log R(\theta)$ to the standard $\chi^2$ distribution without any scale parameter is established in Theorem 3.2. Theorem 3.2 is used for the EL construction of confidence intervals for right-censored data. In Section 4, simulation comparisons of the new method with that of the scaled $\chi^2$ distribution is presented. If $\theta$ is the survival function $F(y)$ at a fixed $y$ or the expected value $EY$, the coverage ratios computing from Wang and Jing (2001), Qin and Zhao (2007), and our EL method are about the same. For more complicated $\theta$, the proposed EL procedure performs better. The amount of improvement depends on the form of $\theta$. In Section 5, an example with real data is provided. The data are taken from Andrews and Herzberg (1985). The proofs of Theorems 3.1, 3.2, and Lemma 3.1 are relegated to the Appendix.

2. Preliminaries and Examples

For $Z = \min(Y, C)$, let

$$H(x) = P(Z \leq x), \quad b_H = \sup\{x : H(x) < 1\}. \quad (2.1)$$

So that $[0, b_H]$ is the range of $H$. The symbols $b_F$ and $b_G$ are similarly defined for $F$ and $G$.

Given a sample of $n$ iid random vectors $(Z_i, \delta_i), \ i = 1, 2, \ldots, n$, of $(Z, \delta)$, their empirical distribution functions are given by

$$H_n^1(x) = \frac{1}{n} \sum_{j=1}^{n} I[Z_j \leq x, \delta_j = 1],$$
$$H_n^0(x) = \frac{1}{n} \sum_{j=1}^{n} I[Z_j \leq x, \delta_j = 0]. \quad (2.2)$$

For any right continuous monotone function $\alpha(x)$, let $\alpha(x^-)$, or $\alpha_-(x)$ denote the left continuous version of $\alpha(x)$ and the curly brackets $\alpha[x]$ denote the difference $\alpha(x) - \alpha(x^-)$. Then $\alpha(x) = d\alpha(x)$. For any cumulative distribution function $F$, let $\overline{F} = 1 - F$.

Asymptotic optimal nonparametric estimators of $F(x)$ and $G(x)$ are the well-known KM estimators given by

$$F_n(x) = 1 - \prod_{s \leq x} \left[ 1 - \frac{H_n^1(s)}{H_n(s^-)} \right]$$
$$G_n(x) = 1 - \prod_{s \leq x} \left[ 1 - \frac{H_n^0(s)}{H_n(s^-)} \right], \quad (2.3)$$

respectively, where an empty product is set equal to one. It can be checked that for all $x$,

$$H_n(x) = F_n(x)G_n(x). \quad (2.4)$$

Applying (2.3) and (2.4), we get

$$dF_n(x) = \frac{dH_n^1(x)}{F_n(x^-)G_n(x)}. \quad (2.5)$$

It follows that

$$dH_n^1(x) = dG_n(x^-), \quad dH_n^0(x) = dF_n(x^-) dG_n(x). \quad (2.6)$$

Put

$$H^0(x) = P(Z \leq x, \delta = 0), \quad H^1(x) = P(Z \leq x, \delta = 1). \quad (2.7)$$

Then

$$H^0(x) = \int_0^x \overline{F}(s) dG(s), \quad H^1(x) = \int_0^x \overline{G}(s) dF(s). \quad (2.8)$$

Here and after, for simplicity, the integral sign $\int_{a}^{b}$ stands for $\int_{(a,b]}$ and $\int$ stands for $\int_{0}^{\infty}$.

Examples of $\theta$ and $g(x, \theta)$

1. $g(x, \theta) = [x > y] - \theta$ with $y$ fixed. Then solving the equation $Eg(Y, \theta) = 0$ for $\theta$ (cf (1.1)) yields $\theta = \overline{F}(y)$, the survival function of $Y$.
2. $g(x, \theta) = x^\theta - \theta$. Then $\theta = E \overline{Y}^\theta$, the $k$-th moments of $Y$.
3. $g(x, \theta) = (x - t_0 \delta) I[x \geq t_0]$ with $t_0$ fixed. Then
$$\theta = E(Y - t_0|Y \geq t_0) = \frac{E(Y - t_0) I[Y \geq t_0]}{P(Y \geq t_0)},$$

the mean residual life of $Y$.
4. $g(x, \theta) = x([x > y] - \theta)$. Then
$$\theta = \frac{1}{EY} \int_y^{\infty} s dF(s),$$

the length-biased survival function of $Y$. See Vardi (1982), for example.
5. \( g(x, \theta) = x^2 - \theta x \). Then
\[
\theta = \frac{1}{E(Y)} \int_0^\infty x^2 \, dF(x),
\]
the mean of the length-biased residual lifetime.

6. \( g(x, \theta) = x^2 - 2\theta x \). Then
\[
\theta = \frac{1}{2E(Y)} \int_0^\infty x^2 \, dF(x),
\]
the mean of the length-biased residual lifetime.

7. \( g(x, \theta) = |x| \leq \theta - p \) with \( p \in (0, 1) \). Then
\[
\theta = F^{-1}(p), \text{ the } p\text{th quantile of } Y.
\]
Examples (4)–(6) often appear in renewal processes and their applications.

3. Empirical Likelihood Ratios and Confidence Intervals for \( \theta \)

The following assumption will be used throughout the article.

Assumption A. \( F \) and \( G \) are continuous, \( g^2(Y, \theta) / G(Y) \) has finite moment
\[
\int_0^\infty \frac{g^2(s, \theta)}{G(s)} \, dF(s) < \infty.
\]
To develop an EL inference procedure, for a specific \( \theta \) and \( g(x, \theta) \), let us define
\[
\psi(s) = \int_{s \geq s} g(x, \theta) \, dF(x), \quad \mu = \int g(x, \theta) \, dF(x).
\]
To obtain an estimating equation for the modified EL ratio in (1.6), we utilize the iid random functions in Theorem 6 of Akritas (2000) or He and Huang (2003),
\[
W_i = W(Z_i, \delta_i, \theta) = \frac{g(Z_i, \theta)\delta_i}{G(Z_i)} + \frac{\delta_i}{H(Z_i)}\psi(Z_i) - \int \psi(s) \frac{I[Z_i \geq s]}{H(s)} \, dH^0(s).
\]
The \( W_i \) are functions of random variables \( Z_i, \delta_i \), and the parameter \( \theta \). It can be calculated that
\[
EW_i = E\left( \frac{g(Z_i, \theta)\delta_i}{G(Z_i)} \right) = \mu,
\]
\[
\sigma_i^2 = \text{var}(W_i) = \int \frac{g^2(s, \theta)}{G(s)} \, dF(s) - \mu^2 - \int \frac{\psi^2(s)}{F(s)G(s)} \, dG(s)
\]
\[
= \int \frac{(F(s)g(s, \theta) - \psi(s))^2}{F(s)G(s)} \, dF(s).
\]
If the true parameter \( \theta_0 \) is the solution of the equation \( E g(Y, \theta) = 0 \), then
\[
EW(Z_i, \delta_i, \theta_0) = \int g(x, \theta_0) \, dF(x) = 0.
\]
Regarding \( W_i \) for \( i = 1, \ldots, n \) as a “complete” random sample, one could formulate an EL likelihood ratio \( R(\theta_0) \) with multinomial probability \( p_i \) assigned to \( W_i \) and the constraint \( \sum_{i=1}^n W_ip_i = 0 \). To proceed, we shall replace the unknown distribution functions \( F, G, H, H^0 \) in (3.2) by their respective estimates \( F_n, G_n \) given by (2.3) and (2.2). Likewise, \( \psi(s) \) will be replaced by the estimate,
\[
\psi_n(s) = \int_{s \geq s} g(s, \theta) \, dF_n(s).
\]
We arrived at an approximation of \( W_i \) in (3.2) as
\[
W_{ni} = \frac{g(Z_i, \theta)\delta_i}{G_n(Z_i)} + \frac{\delta_i}{H_n(Z_i)}\psi_n(Z_i) - \int \psi_n(s) \frac{I[Z_i \geq s]}{H_n(s)} \, dH_n^0(s).
\]
The price to pay for the approximation is that \( W_{ni}'s \) are not stochastically independent which complicates the ensuing analysis.

The following theorem indicates the possibility of using \( W_{ni} \) to construct EL ratio and to obtain asymptotically a standard \( \chi^2 \) distribution.

Theorem 3.1. Let \( W_{ni} \) be given by (3.7) and \( E g(Y, \theta) = 0 \). Then under Assumption A, as \( n \to \infty \), we have
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^n W_{ni} = \sqrt{n} \int g(y, \theta) \, dF_n(y) = \frac{1}{\sqrt{n}} \sum_{i=1}^n W_i + o_p(1).
\]
Following literature, we call \( W(Z_i, \delta_i) \) the \( i \)-th influence function of \( \int g(s, \theta) \, dF_n(s) \) for given \( \theta \).
We define the modified EL ratio of \( \theta \) by a multinomial likelihood subject to constraints as
\[
R(\theta) = \sup \left\{ \sum_{i=1}^n p_i \left| \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i W_{ni} = 0, \right. \right. \}
\]
\[
\left. \left. p_i \geq 0, \ i = 1, 2, \ldots, n \right\} \right\}.
\]
To determine \( R(\theta) \), we solve, as usual, for the Lagrange multipliers \( \beta \) and \( \lambda \) in
\[
A = \sum_{i=1}^n \log(n p_i) - n \lambda \left( \sum_{i=1}^n p_i W_{ni} \right) - \beta \left( 1 - \sum_{i=1}^n p_i \right).
\]
Then \( \beta = -n \) and \( p_i = \frac{1}{n} (1 + \lambda W_{ni})^{-1} \), \( i = 1, 2, \ldots, n \), where \( \lambda \) is the solution of the equation
\[
\vartheta(\lambda) = \frac{1}{n} \sum_{i=1}^n W_{ni} / (1 + \lambda W_{ni}) = 0.
\]
Lemma 3.1 shows that with probability 1, for large \( n \) the set \( \{ W_{ni} \} \) contains a positive and a negative value. It follows that for large \( n \), (3.10) has a unique solution \( \lambda_n \) such that \( \lambda_n W_{ni} > -1 \). Now, the EL ratio of \( \theta \) can be written as
\[
R(\theta) = \prod_{i=1}^n (np_i) = \prod_{i=1}^n (1 + \lambda_n W_{ni})^{-1}.
\]

Theorem 3.2. Suppose that \( \theta_0 \) is the unique solution of \( E g(Y, \theta) = 0 \). Then under Assumption A, \( I(\theta_0) = -2 \log R(\theta_0) \) converges in distribution to a \( \chi^2 \) random variable with one degree of freedom, as \( n \to \infty \).
Applying Theorem 3.2, confidence intervals for $\theta$ can be constructed as
\[
I_1 = \{ \theta : l(\theta) \leq c_{1-\alpha} \},
\]
where $c_{1-\alpha}$ is the $(1 - \alpha)$th quantile of the $\chi^2_1$ distribution. $I_1$ has asymptotic coverage probability of $1 - \alpha$, as $n \to \infty$.

The following lemma is needed for proving Theorem 3.2.

**Lemma 3.1.** Under the conditions of Theorem 3.2 and $\theta = \theta_0$, as $n \to \infty$,
\[
\begin{align*}
(1) & \quad \frac{1}{n} \sum_{i=1}^{n} (W_m - W_i)^2 = o_p(1), \\
(2) & \quad \max_{1 \leq i \leq n} |W_i| = o_p(\sqrt{n}), \quad \frac{1}{n} \sum_{i=1}^{n} W_i^2 = \sigma^2 + o_p(1).
\end{align*}
\]

**Remark 3.1.** We are able to obtain the standard asymptotic $\chi^2_1$ distribution for $-2 \log R(\theta_0)$. This is because according to Lemma 3.1 the asymptotic variance, $\sigma^2$, of $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i$ is equal to the limit of $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_i^2$.

In Qin and Zhao (2007), the constraint in their EL ratio is $\sum_{i=1}^{n} P_i V_i = 0$, where
\[
V_i = \frac{g(Z_i, \theta) \delta_i}{1 - G_n(Z_i)}
\]
is the estimate of $V = g(Z, \theta) \delta/(1 - G(Z))$. Theorem 3.1 remains true for the weights $V_i$, that is
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_i + o_p(1)
\]
and, therefore, the asymptotic variance of $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i$ is $\sigma^2$ as given in (3.4). However, Lemma 3.1(1) does not hold for $\{V_i\}$. The limit of $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i^2$ or $\frac{1}{n} \sum_{i=1}^{n} (V_i - \frac{1}{n} \sum_{i=1}^{n} V_i)^2$ is
\[
\sigma^2 = \int \frac{q(s, \theta)}{G(s)} \, dF(s) - \mu^2 > \sigma^2, \quad \mu = Eg(Y, \theta).
\]
Therefore, a scaled parameter $r = \sigma^{-1}_1 / \sigma^2$ must be introduced to obtain an asymptotic distribution for $-2 \log(EL \ ratio)$.

### 4. Simulation

Simulations are carried out to study and compare finite sample performance of confidence intervals $I_1$ in (3.12) derived from Theorem 3.2 and $I_2$ from the scaled $\chi^2_1$ distribution given by Wang and Jing (2001) and Qin and Zhao (2007).

Confidence intervals $I_2$ are calculated as follows. Let $F_n$ and $G_n$ be the KM estimators defined by (2.3). Suppose $\hat{\theta}$ is the unique solution of $\int g(s, \theta) \, df_n(s) = 0$. Set
\[
\begin{align*}
\xi_i &= g(Z_i, \theta), \\
\hat{\xi}_i &= g(Z_i, \hat{\theta}), \\
V_m &= \frac{\hat{\xi}_i \delta_i}{1 - G_n(Z_i)} \quad \text{and} \\
V_m &= \frac{\hat{\xi}_i \delta_i}{1 - G_n(Z_i)},
\end{align*}
\]
\[
\begin{align*}
\hat{V}_m &= \frac{\hat{\xi}_i \delta_i}{1 - G_n(Z_i)}, \\
\sigma^2 &= \frac{1}{n} \sum_{i=1}^{n} (\hat{V}_m - \overline{V}_m)^2, \\
\hat{r} &= \frac{\sigma^2}{n \var^* (\text{jack})},
\end{align*}
\]
where $n \var^* (\text{jack})$ is the modified jackknife estimator of the asymptotic variance of $\xi$ given in Stute (1996). Then, the confidence interval for $\theta$ is
\[
I_2 = \left\{ \theta : 2 \hat{r} \sum_{i=1}^{n} \log (1 + \lambda V_m) \leq c_{1-\alpha} \right\},
\]
where $\lambda$ is the solution of $\sum_{i=1}^{n} V_m/(1 + \lambda V_m) = 0$.

Simulations were performed in two scenarios. In Scenario I, the parameter of interest is $\theta_0 = E Y$ and in Scenario II, the mean residual lifetime.

Scenario I: The parameter of interest is $\theta_0 = E Y$ and $\xi(x) = g(x, \theta) = x - \theta$ is used for calculating $I_1$. Two cases (i) and (ii) were simulated:

(i) The lifetime $Y$ is uniformly distributed on $(0, 1)$ and the censoring time $C$ is uniformly distributed on $(0, c)$. We selected $c = 2.5$ and $c = 1.3$ which corresponds, respectively, to 20% and 30% censoring proportions.

(ii) $Y$ has a Weibull$(1, 10)$ distribution and $C$ has an Exponential($\lambda$) distribution. Then for $\lambda = 4.3$ and $\lambda = 2.7$, the corresponding censoring proportions are 20% and 30%. The simulated sample is $n$ iid copies of $Z = \min(Y, C)$, $\delta = I(Y \leq C)$. Based on the simulated observations, confidence intervals $I_1$ derived from Theorem 3.2 and $I_2$ from (4.2) were calculated. The simulation of a sample $n$ was repeated $N$ times for $N = 2 \times 10^4$. The coverage proportions and the average width of the $N$ confidence intervals were calculated using the $N$ datasets. The results are summarized in Tables 1 and 2. The following are noted.

1. As the sample size $n$ increases from 20 to 80, all of the coverage proportions converge to the nominal level $1 - \alpha$.
2. For Uniform(0, 1) distribution, $I_1$ has better coverage proportions. In 8/16 of the cases, the average width of $I_2$ is slightly shorter than that of $I_1$. In 8/16 of the cases, $I_2$ and $I_1$ have the same average width.
3. For Weibull(1, 10) distribution, $I_1$ has better coverage proportion and width.

Scenario II: Let $g(x, \theta) = (x - t_0 - \theta) I[x \geq t_0]$. Then the parameter of interest $\theta$ is the mean residual life of $Y$, $\theta = E(Y - t_0|Y \geq t_0)$, as considered in Qin and Zhao (2007) and in Example 3 of Section 2.

We used the same Weibull (1, 10) distribution for $Y$ and the same exp($\beta$) distribution for the censoring variable $C$ as in Scenario I. As in Scenario I, a sample of $W_m$, $i = 1, \ldots, n$ was simulated repeatedly $N$ times for $N = 2 \times 10^4$. The coverage proportions and the average width of the $N$ confidence intervals were calculated using the $N$ datasets. The results are summarized in Table 3 and Table 4. The following are noted.

1. All of the coverage proportions increase with the sample size $n$ and are close to the nominal levels. In fact when sample size is 80 (which is moderate), the coverage proportions are about the nominal level except when $P[Y \geq t_0] = 0.3$.
2. The coverage proportions of $I_1$ are much better than that of $I_2$.
3. In 15/32 of the samples, the average width of $I_2$ is slightly shorter than that of $I_1$. In 17/32 of the cases, $I_2$ and $I_1$ have the same average width.
Table 1. The coverage proportions for the true $\theta_0 = E(Y)$.

| Censoring proportion | Nominal value $1 - \alpha$ | Sample size $n$ | Uniform($0, 1$) | Weibull($1, 10$) |
|----------------------|----------------------------|----------------|----------------|----------------|
|                      | $l_2$                       | $l_1$          | $l_2$          | $l_1$          |
| 20%                  | 0.90                        | 20             | 0.876          | 0.881          | 0.871          | 0.871          |
|                      | 0.95                        | 20             | 0.928          | 0.935          | 0.922          | 0.924          |
| 30%                  | 0.90                        | 20             | 0.841          | 0.861          | 0.869          | 0.869          |
|                      | 0.95                        | 20             | 0.897          | 0.897          | 0.893          | 0.893          |

Table 2. The average width of confidence intervals for $\theta_0 = E(Y)$.

| Censoring proportion | Nominal value $1 - \alpha$ | Sample size $n$ | Uniform($0, 1$) | Weibull($1, 10$) |
|----------------------|----------------------------|----------------|----------------|----------------|
|                      | $l_2$                       | $l_1$          | $l_2$          | $l_1$          |
| 20%                  | 0.90                        | 20             | 0.217          | 0.218          | 0.092          | 0.091          |
|                      | 0.95                        | 20             | 0.157          | 0.157          | 0.066          | 0.065          |
| 30%                  | 0.90                        | 20             | 0.258          | 0.259          | 0.109          | 0.109          |
|                      | 0.95                        | 20             | 0.154          | 0.154          | 0.064          | 0.064          |

Table 3. The coverage proportions and average width of confidence intervals for $\theta_0 = E(Y - Y_{\Delta})$ under the assumptions of $Y \sim$ Weibull($1, 10$), 20% censoring proportion, $1 - \alpha = 0.90$, and specified value for $P(Y \geq t_0)$.

| Sample size $n$ | CI  | 0.90 | 0.70 | 0.50 | 0.30 | 0.90 | 0.70 | 0.50 | 0.30 |
|----------------|-----|------|------|------|------|------|------|------|------|
| 20             | $l_2$ | 0.878 | 0.851 | 0.795 | 0.659 | 0.074 | 0.062 | 0.054 | 0.044 |
|                | $l_1$ | 0.881 | 0.863 | 0.820 | 0.701 | 0.074 | 0.062 | 0.056 | 0.048 |
| 40             | $l_2$ | 0.889 | 0.878 | 0.859 | 0.800 | 0.053 | 0.046 | 0.042 | 0.039 |
|                | $l_1$ | 0.891 | 0.884 | 0.874 | 0.833 | 0.053 | 0.046 | 0.043 | 0.041 |
| 60             | $l_2$ | 0.897 | 0.892 | 0.877 | 0.839 | 0.044 | 0.037 | 0.035 | 0.034 |
|                | $l_1$ | 0.898 | 0.897 | 0.888 | 0.863 | 0.044 | 0.038 | 0.035 | 0.035 |
| 80             | $l_2$ | 0.895 | 0.888 | 0.884 | 0.853 | 0.038 | 0.033 | 0.031 | 0.030 |
|                | $l_1$ | 0.896 | 0.892 | 0.892 | 0.871 | 0.038 | 0.033 | 0.031 | 0.031 |

Table 4. The coverage proportions and average width of confidence intervals for $\theta_0 = E(Y - Y_{\Delta})$ under the assumptions of $Y \sim$ Weibull($1, 10$), 30% censoring proportion, $1 - \alpha = 0.90$, and specified value for $P(Y \geq t_0)$.

| Sample size $n$ | CI  | 0.90 | 0.70 | 0.50 | 0.30 | 0.90 | 0.70 | 0.50 | 0.30 |
|----------------|-----|------|------|------|------|------|------|------|------|
| 20             | $l_2$ | 0.864 | 0.833 | 0.760 | 0.605 | 0.079 | 0.065 | 0.055 | 0.043 |
|                | $l_1$ | 0.872 | 0.851 | 0.793 | 0.659 | 0.079 | 0.065 | 0.058 | 0.048 |
| 40             | $l_2$ | 0.887 | 0.872 | 0.846 | 0.777 | 0.057 | 0.048 | 0.045 | 0.041 |
|                | $l_1$ | 0.891 | 0.892 | 0.867 | 0.818 | 0.057 | 0.049 | 0.046 | 0.043 |
| 60             | $l_2$ | 0.892 | 0.888 | 0.870 | 0.822 | 0.046 | 0.040 | 0.037 | 0.036 |
|                | $l_1$ | 0.895 | 0.895 | 0.884 | 0.851 | 0.046 | 0.040 | 0.038 | 0.037 |
| 80             | $l_2$ | 0.892 | 0.888 | 0.878 | 0.845 | 0.040 | 0.035 | 0.033 | 0.032 |
|                | $l_1$ | 0.895 | 0.895 | 0.887 | 0.869 | 0.040 | 0.035 | 0.033 | 0.033 |
5. Example with Real Data

In this section, we apply our method to a well-known dataset taken from Andrews and Herzberg (1985). The data are the survival times of heart transplant patients in the Stanford Heart Transplant Program. The survival data from the Program have been analyzed and reanalyzed by many authors using different methods. A comparison of performances of several different methods was carried out in Miller and Halpern (1982) using the survival times of 184 patients who received a transplant among all of those patients admitted to the program during the period from October 1967 to February 1980. For each of the 184 patients, his/her survival time in days after transplant and an indicator on whether the patient is dead or alive were recorded.

We are interested in the mean survival time after transplant, $\theta$. We shall use the same dataset to compute the confidence intervals for $\theta$.

Based on our method described in the simulations, the 95% confidence interval for $\theta$ is [1038, 1478]. The 95% confidence interval computed using the method of Wang and Jing (2001) is [1045, 1232]. A modified EM algorithm to calculate the EL method was carried out in Miller and Halpern (1982) using survival times of heart transplant patients in the Stanford Heart lifetime. We treated it as an uncensored lifetime to compute the mle.

Appendix

Throughout Section 6, $\theta$ is fixed and such that $Eg(Y, \theta) = 0$. Then it is convenient to suppress $\theta$ in the exposition, by setting $\xi(y) = g(y, \theta)$. It follows that

$$E\xi(Y) = \int \xi(y)dF(y) = 0.$$ 

The following Lemma will be used repeatedly.

Lemma A.1. For $b < b_H$, let $\{h_n(b)\}$ be a random sequence such that $h_n(b) \rightarrow h(b)$ in distribution as $n \rightarrow \infty$, and $h(b) = o_p(1)$ as $b \rightarrow b_H$. As $n \rightarrow \infty$, if $V_n = O_p(1)$ and the random sequence $\{s_n\}$ can be written as $s_n = O_p(1) + V_nh_n(b)$ for any $b < b_H$, then $s_n = O_p(1)$.

Proof. Put $\eta_n = o_p(1)$. By assumptions, for any $\varepsilon > 0$ and $\delta > 0$, there exist $M > 0$, $b < b_H$ and $n_0 > 1$ such that for $n \geq n_0$, $P(|V_n| \geq M) \leq \delta$, $P(|h_n(b)| \geq \varepsilon) \leq P(|h(b)| \geq \varepsilon/M) + \delta/2 \leq \delta$ and $P(|\eta_n| \geq \varepsilon) < \delta$. It follows that for $n \geq n_0$,

$$P(|s_n| \geq 2\varepsilon) \leq P(|\eta_n| \geq \varepsilon) + P(|V_nh_n(b)| \geq \varepsilon) \leq \delta + P(|V_nh_n(b)| \geq \varepsilon) \leq P(|V_n| \geq M) + \delta \leq 3\delta.$$ 

As an example of Lemma A.1, put

$$h_n(b) = \int_b^{b_H} \frac{\xi^2(s)}{G(s)} dH_n^1(s), \ b < b_H.$$ 

Then, by the SLLN and (2.8),

$$\lim_{n \rightarrow \infty} h_n(b) = \int_b^{b_H} \frac{\xi^2(s)}{G(s)} dH^1(s) = h(b) \rightarrow 0, \ as \ b \rightarrow b_H. \text{(A.1)}$$

Proof of Theorem 3.1. The first equation follows from

$$\frac{1}{n} \sum_{i=1}^n W_{ni} = \frac{1}{n} \sum_{i=1}^n \left( \frac{\xi(Z_i)\delta_i}{G_n(Z_i-)} + \frac{\delta_i}{H_n(Z_i-)} \psi_n(Z_i) \right) - \int \psi_n(s) \frac{I[Z_i \geq s]}{H_n(s-)} dH^0_n(s) + \int \psi_n(s) \frac{H_n(s-)}{H_n(s-)} dH^0_n(s) - \int \psi_n(s) \frac{H_n(s-)}{H_n(s-)} dH^0_n(s) = \int \frac{\xi(s)}{G_n(s-)} dH^1_n(s) = \int \xi(s) dF_n(s).$$

The second equation follows from Theorem 6 in Akritas (2000) and the fact that $W_i$ equals $Z_i$ defined by (12) in Akritas (2000).

Proof of Lemma 3.1. The differences $W_{ni} - W_i$ in (3.7) and (3.2) can be expressed in terms of

$$\gamma_i = \frac{\xi(Z_i)\delta_i}{G_n(Z_i-)} - \frac{\xi(Z_i)\delta_i}{G(Z_i)}$$

$$\eta_i = \frac{\delta_i}{H_n(Z_i-)} \psi_n(Z_i) - \frac{\delta_i}{H_n(Z_i-)} \psi(Z_i),$$

$$v_i = \int \psi_n(s) \frac{I[Z_i \geq s]}{H_n(s-)} dH^0_n(s) - \int \psi(s) \frac{I[Z_i \geq s]}{H_n(s-)} dH^0_n(s),$$

as

$$W_{ni} - W_i = \gamma_i + \eta_i - v_i.$$ 

Applying an elementary inequality $(a + b + c)^2 \leq 3(a^2 + b^2 + c^2)$, we obtain

$$(W_{ni} - W_i)^2 = (\gamma_i + \eta_i - v_i)^2 \leq 3(\gamma_i^2 + \eta_i^2 + v_i^2). \text{ (A.2)}$$

The lemma will be proven by showing that the sample means of $\gamma_i^2$, $\eta_i^2$, and $v_i^2$ tend to zero in probability. The proofs will be presented in (A), (B), and (C) below.

(A) The sample mean of $\gamma_i^2$ is $o_p(1)$.

Proof: Let $G_n(x)$ be the KM estimator defined in (2.3) and $b < b_H$. Then as $n \rightarrow \infty$,

$$U_n = \sup_{s \leq b} \left| \frac{G_n(s) - G(s)}{G_n(s)} \right| = o_p(1),$$
\[ V_n = \sup_{s \leq \max\{b, b_H\}} \frac{|G_n(s^-) - G(s)|}{G_n(s^-)} = O_P(1). \quad (A.3) \]

See Zhou (1992). To apply this result, we shall split the first integral in the following equation into two intervals \([0, b]\) and \([b, b_H]\), accordingly. For \(H_n^b\) defined in (2.2), we have

\[
\frac{1}{n} \sum_{i=1}^{n} \gamma_i^2 = \int \left( \frac{\xi(s)}{G_n(s^-)} \right)^2 dH_n(s) 
\leq U_n^2 \int_0^b \frac{\xi^2(s)}{G(s)} dH_n^b(s) + U_n^2 \int_b^{b_H} \frac{\xi^2(s)}{G(s)} dH_n^b(s) 
= o_P(1) + O_P(1) h_n(b) = o_P(1), \quad \text{as } n \to \infty, \quad \text{(A.4)}
\]

where

\[
h_n(b) = \int_b^{b_H} \frac{\xi^2(s)}{G(s)} dH_n^b(s). \quad \text{(A.5)}
\]

By (A.1), \(h_n(b)\) satisfies Lemma A.1. The proof of (A) follows.

(B) The sample mean of \(\gamma_i^2\) is \(O_P(1)\).

To split the integral that defines \(\sum \gamma_i^2/n\), let us define

\[
T_n(b, t) = \int_t^b \frac{\psi_n(s)}{H_n(s^-)} dH_n^b(s),
\]

\[
S_n(b, t) = \int_t^b \frac{\psi_n(s)}{H_n(s^-)} dH_n^b(s), \quad b < b_H. \]

Observe that

\[
\psi_n^2(s) = \left( \int_{u<s} \xi(u) dF_n(u) \right)^2 \leq F_n(s^-) \int_{u<s} \xi^2(u) dF_n(u),
\]

and \(F_n\) and \(G_n\) have no common jumps. It follows that

\[
T_n(b, b_H) \leq \int_b^{b_H} \left( \int_{u<s} \xi^2(u) dF_n(u) \right) dF_n(s^-) \frac{dG_n(s)}{H_n(s^-)} 
\leq \int_b^{b_H} \left( \int_{u<s} \xi^2(u) dF_n(u) \right) \frac{dG_n(s)}{H_n(s^-)} 
\leq \frac{1}{G_n(s^-)} \int_b^{b_H} \xi^2(u) dF_n(u) + \int_b^{b_H} \frac{\xi^2(s)}{G_n(s^-)} dF_n(s) 
\leq 2 \int_b^{b_H} \left[ \left( \frac{1}{G_n(s^-)} - \frac{1}{G(s)} \right)^2 \left( \frac{1}{G_n(s^-)} + \frac{1}{G(s)} \right) \right] dH_n^b(s) 
= o_P(1) + O_P(1) h_n(b). \quad \text{(A.6)}
\]

The first inequality follows from (2.6), the second and the third from integration by parts, and the last equality from \((a + b)^2 \leq 2(a^2 + b^2), (A.3), \text{ and } (A.1).\)

By the same token, we conclude that \(S_n(b, b_H) = o_P(1) + O_P(1) h_n(b).\).

Write

\[
\xi = \xi^+ - \xi^-\quad \psi_n(x) = \int_{s<x} \xi^+(s) dF_n(s) - \int_{s<x} \xi^-(s) dF_n(s),
\]

where \(\xi^+\) and \(\xi^-\) are the positive and negative part of \(\xi\). Define monotone functions:

\[
\psi_n^+(x) = \int_{s>x} \xi^+(s) dF_n(s), \quad \psi_n^-(x) = \int_{s>x} \xi^-(s) dF_n(s).
\]

The monotone \(\psi_n^+(x)\) converges to continuous \(\psi^+(x)\) almost surely for \(x \in [0, b_H]\) as shown by Stute and Wang (1993). The convergence is uniform on \([0, b_H]\). From these we conclude

\[
\sup_{0 \leq s \leq b_H} |\psi_n(s) - \psi(s)| = o(1), \quad \text{a.s.} \quad \text{(A.7)}
\]

and for \(b < b_H,

\[
\sup_{0 \leq s \leq b_H} \left( \frac{\psi_n(s)}{H_n(s^-)} - \frac{\psi(s)}{H(s)} \right)^2 \to 0, \quad \text{a.s.}
\]

Applying \((a + b)^2 \leq 2a^2 + 2b^2\) and Lemma A.1, we have

\[
\frac{1}{n} \sum_{i=1}^{n} \eta_i^2 = \int_b^{b_H} \left( \frac{\psi_n(s)}{H_n(s^-)} - \frac{\psi(s)}{H(s)} \right)^2 dH_n^b(s) 
= \int_b^{b_H} \left( \frac{\psi_n(s)}{H_n(s^-)} - \frac{\psi(s)}{H(s)} \right)^2 dH_n^b(s) 
+ 2T_n(b, b_H) + 2S_n(b, b_H) 
= O_P(1) + O_P(1) h_n(b) = o_P(1). \quad \text{(A.8)}
\]

(C) The sample mean of \(\gamma_i^2\) is \(O_P(1)\).

Proof: To split the integral that defines \(\sum \gamma_i^2/n\), let us define

\[
B_n(a, t) = \int_a^t \frac{\psi_n(s) dH_n^b(s)}{H_n^2(s^-)} \quad \tilde{B}_n(a, t) = \int_a^t \frac{\psi_n(s) dH_n^b(s)}{H_n^2(s^-)},
\]

\[
D(a, t) = \int_a^t \frac{\psi(s) dH_n^b(s)}{H_n^2(s^-)}.
\]

Then, for \(b < b_H\), we have

\[
\Delta_n^2 = \int_b^{b_H} B_n^2(b, b_H) dH_n(b) \leq \int_b^{b_H} \tilde{B}_n^2(b, b_H) d(-H_n(t)) \leq \frac{\psi_n(t)}{H_n(t^-)} dH_n(t) \leq 2 \int_b^{b_H} \tilde{B}_n^2(b, b_H) \left[ \frac{\psi_n(t)}{H_n(t^-)} \right] \frac{dH_n^b(t)}{H_n^b(t^-)} \leq 0 + 2 \left( \int_b^{b_H} \tilde{B}_n^2(b, b_H) dH_n(t) \right)^{1/2} \left( \int_b^{b_H} \frac{\psi_n(t)}{H_n(t^-)} \frac{dH_n^b(t)}{H_n^b(t^-)} \right)^{1/2} \quad \text{(A.9)}
\]

The second term in the first inequality is obtained using the Lebesgue–Stieltjes integration by parts and \(\tilde{B}_n(t^-) \leq \tilde{B}_n(t).\)

Applying (A.9), we get

\[
\int_b^{b_H} B_n^2(b, b_H) dH_n(t) = \Delta_n^2 \leq 4 \int_b^{b_H} \tilde{B}_n^2(b, b_H) dH_n(t) \leq 4 \int_b^{b_H} \tilde{B}_n^2(b, b_H) \left[ \frac{\psi_n(t)}{H_n(t^-)} \right] \frac{dH_n^b(t)}{H_n^b(t^-)} \leq 4 \int_b^{b_H} D^2(b, t) dH_n(t) \leq 4 \int_b^{b_H} D^2(b, t) dH_n(t) = 4 \int_b^{b_H} D^2(b, t) dH_n(t) + O_P(1) + O_P(1) h_n(b). \]

Similarly, for \(S(b, b_H) = ES_n(b, b_H),\) we have

\[
\int_b^{b_H} D^2(b, t) dH_n(t) \leq 4S(b, b_H) + O_P(1) + O_P(1) h_n(b). \]
Applying (A.7), we conclude that for any \( b < b_H \), with probability 1, 
\( B_n(0, t) \rightarrow D(0, t) \) uniformly on \([0, b] \). It follows that 
\( B_n(0, b) \rightarrow D(0, b) \), and
\[
\frac{1}{n} \sum_{i=1}^{n} w_i^2 = o_p(1) + \int_b^{b_H} \left( B_n(0, b) - D(0, b) + B_n(b, t) \right) \left( D(b, t) \right)^2 dH_n(t)
\]
\[
\leq o_p(1) + 4 \int_b^{b_H} B_n(b, t) dH_n(t) + 4 \int_b^{b_H} D^2(b, t) dH_n(t)
\]
\[
= o_p(1) + o_p(1) h_n(b) + o_p(1) h_n(b) = o_p(1).
\]

Finally, we prove the second assertion of Lemma 3.1. Since \( W_i \) are iid random variables with zero mean and finite variance \( \sigma^2 \), hence, \( \max_{1 \leq i \leq n} |W_i| = o_p(\sqrt{n}) \). It follows from assertion (1) of the lemma that
\[
\max_{1 \leq i \leq n} |W_i| \leq \left( \max_{1 \leq i \leq n} |W_m - W_i|^2 \right)^{1/2} + \max_{1 \leq i \leq n} |W_i| = o_p(\sqrt{n}).
\]

Note that \( W_m^2 \) is bounded by \( W^2 + (W - W_m)^2 \leq 2|W(W - W_m)| \). By assertion (1) of the lemma, we get
\[
\frac{1}{n} \sum_{i=1}^{n} W_m^2 = \sigma^2 + o_p(1).
\]

**Proof of Theorem 3.2.** The proof is similar to that of Theorem 3.2 in Chapter 3 of Owen (2001). Setting
\[
X_i = \lambda_n W_m, \quad W_n = \frac{1}{n} \sum_{i=1}^{n} W_i, \quad S_n^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2.
\]

We have \( S_n^2 = \sigma^2 + o_p(1), \) \( W_n = O_p(n^{-1/2}) \),
\[
\frac{1}{n} \sum_{i=1}^{n} X_i^2 = X - \lambda_n h(\lambda_n) = \lambda_n W_n.
\]

and
\[
\lambda_n^2 S_n^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 \leq \frac{1}{n} \sum_{i=1}^{n} X_i^2 (1 + \max_{1 \leq i \leq n} |X_j|)
\]
\[
= \lambda_n W_n + \lambda_n^2 o_p(1).
\]

It follows that
\[
\lambda_n = \frac{W_n}{\sigma^2 + o_p(1)} = O_p(n^{-1/2})
\]
and
\[
W_n = \lambda_n \sigma^2 + o_p(n^{-1/2}).
\]

Applying Lemma 3.1, we have
\[
\sum_{i=1}^{n} |X_i|^2 \leq \lambda_n \sum_{i=1}^{n} |W_i|^2 \max_{1 \leq j \leq n} |W_n| = o_p(1).
\]

Now, using Lemma 3.1, Taylor’s expansion and Theorem 3.1, we can write
\[
-2 \log R(\theta_n) = 2 \sum_{i=1}^{n} \ln(1 + \lambda_n W_n)
\]
\[
= \sum_{i=1}^{n} 2 \left( \lambda_n W_n - \frac{1}{2} \lambda_n^2 W_n^2 + \eta_i \right) \quad \text{(here } |\eta_i| \leq |X_i|^3 \text{)}
\]
\[
= 2n \frac{W_n^2}{\sigma^2} - \frac{1}{\sigma^4} n W_n^2 (\sigma^2 + o_p(1)) + o_p(1)
\]
\[
= n \frac{W_n^2}{\sigma^2} + o_p(1) \rightarrow \chi^2_1, \quad \text{in distribution.}
\]

**Acknowledgments**

The authors are grateful to the associate editor and referees for their valuable comments and providing missing references which led to an improvement of the presentation of our work and a simplification of the proof of Theorem 3.1.

**Funding**

This work was done when Yang was visiting the School of Mathematical Sciences, Xiamen University. Research of He, Liang, and Shen were supported by National Natural Science Foundation of China(11171230,11231010). Research of Yang was supported in part by NSF grant DMS-1209111.

**References**

Akritas, M. G. (2000), "The Central Limit Theorem Under Censoring," *Bernoulli*, 6, 1109–1120. [649,652]
Andrews, D. F., and Herzberg, A. M. (1985), *DATA: A Collection of Problems from Many Fields for the Student and Research Work,* Berlin: Springer, pp. 45–49. [648,652]
Chen, S. X. (1994), "Empirical Likelihood Confidence Intervals for Linear Regression Coefficients," *Journal of Multivariate Analysis*, 49, 24–40. [646]
DiCiccio, T. J., Hall, P., and Romano, J. P. (1991), "Empirical Likelihood is Bartlett-Callable," *Annals of Statistics*, 19, 1053–1061. [646]
He, S. Y., and Huang, X. (2003), "Central Limit Theorem of Linear Regression Model Under Right Censorship," *Science in China*, 46, 600–610. [649]
Hjort, N. L., McKeage, I. W., and van Keilegom, I. (2009), "Extending the Scope of Empirical Likelihood," *Annals of Statistics*, 37, 1079–1111. [647]
Li, G., and Wang, Q. H. (2003), "Empirical Likelihood Methods for Linear Regression Analysis of Right Censored Data," *Statistica Sinica*, 13, 51–68. [647]
Miller, R. G., and Halpern, J. W. (1982), "Regression With Censored Data," *Biometrika*, 69, 521–531. [652]
Owen, A. B. (1988), "Empirical Likelihood Ratio Confidence Intervals for Single Functional," *Biometrika*, 75, 237–249. [646]
Owen, A. B. (2001), *Empirical Likelihood*, London: Chapman and Hall. [646,647,654]
Pan, X. R., and Zhou, M. (2002), "Empirical Likelihood Ratio in Terms of Cumulative Hazard Function for Right Censored Data," *Journal of Multivariate Analysis*, 80, 166–188. [647]
Qin, G. S., and Zhao, Y. C. (2007), "Empirical Likelihood Inference for the Mean Residual Life Under Random Censorship," *Annals of Statistics*, 35, 646–647, 652.
Stute, W. (1996), "The Jackknife Estimate of Variance of a Kaplan-Meier Integral," *Annals of Statistics*, 24, 2679–2704. [646,650]
Stute, W., and Wang, J. L. (1993), "The Strong Law Under Random Censorship," *Annals of Statistics*, 21, 1591–1607. [653]
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. [646]

Vardi, Y. (1982), “Nonparametric Estimation in the Presence of Length Bias,” Annals of Statistics, 10, 616–620. [648]

Wang, Q. H., and Jing, B. Y. (2001), “Empirical Likelihood for a Class of Functions of Survival Distribution With Censored Data,” Annals of the Institute of Statistical Mathematics, 53, 517–527. [647, 648, 650, 652]

Yang, S. (1994), “A Central Limit Theorem for Functionals of the Kaplan-Meier Estimator,” Statistics and Probability Letters, 21, 337–345. [646]

Yang, Y., and He, X. M. (2012), “Bayesian Empirical Likelihood for Quantile Regression,” Annals of Statistics, 40, 1102–1131. [647]

Zheng, M., Zhao, A. Q., and Yu, W. (2012), “Empirical Likelihood Methods Based on Influence Functions,” Statistics and Its Interface, 5, 355–266. [646, 648]

Zhou, M. (1992), “Asymptotic Normality of the ’Synthetic Data’ Regression Estimator for Censored Survival Data,” Annals of Statistics, 20, 1002–1021. [653]

——— (2005), “Empirical Likelihood Ratio With Arbitrarily Censored/Truncated Data by EM Algorithm,” Journal of Computational and Graphical Statistics, 14, 643–656. [652]

Zhou, M., and Jeong, J. H. (2011), “Empirical Likelihood Ratio for Median and Mean Residual Lifetime,” Statistics in Medicine, 30, 152–159. [647]

Zhou, M., Kim, M. O., and Bathke, A. C. (2012), “Empirical Likelihood Analysis for The Heteroscedastic Accelerated Failure Time Model,” Statistica Sinica, 22, 295–316. [647]