Ann Inst Stat Math (2008) 60:575–589
DOI 10.1007/s10463-007-0114-8

Empirical likelihood confidence intervals for hazard and density functions under right censorship

Junshan Shen · Shuyuan He

Received: 1 November 2005 / Revised: 19 October 2006 / Published online: 1 March 2007

Abstract In this paper, we use smoothed empirical likelihood methods to construct confidence intervals for hazard and density functions under right censorship. Some empirical log-likelihood ratios for the hazard and density functions are obtained and their asymptotic limits are derived. Approximate confidence intervals based on these methods are constructed. Simulation studies are used to compare the empirical likelihood methods and the normal approximation methods in terms of coverage accuracy. It is found that the empirical likelihood methods provide better inference.

Keywords Censored data · Density function · Empirical likelihood · Hazard function · Kernel smoothing

1 Introduction

Right-censored data appear naturally in biomedical research and industrial lifetime analysis. Let $\{T_i : \ i = 1, 2, \cdots\}$ be independent and identically distributed (i.i.d.) nonnegative failure times with common distribution function F_0, $\{C_i : \ i = 1, 2, \cdots\}$ be i.i.d. censoring times with distribution function G_0 and independent of the failure times. We assume throughout that F_0 and G_0 are continuous. Under right censorship, we observe the right censored vectors

$$(X_1, \delta_1), \ldots, (X_n, \delta_n), \tag{1}$$

where $X_i = \min(T_i, C_i), \delta_i = I(T_i \leq C_i)$, the indicator of $T_i \leq C_i$.

J. Shen (✉) · S. He
School of Mathematical Sciences, Peking University, Beijing, 100871, China
e-mail: shenjunshan@math.pku.edu.cn

In the analysis of lifetime data, researchers are often interested in estimating the density function f_0 of the failure times and the associated hazard function

$$h_0(t) := -\frac{\mathrm{d}}{\mathrm{d}t}\ln(S_0(t)) = \frac{f_0(t)}{1-F_0(t)},$$

where $S_0 = 1 - F_0$ is the survival function. In many approaches, the kernel method, studied extensively in the literature, is the simplest. For example, Ramlau-Hansen (1983) used this method to smooth counting process intensities, Tanner and Wang (1983) gave expressions for the bias and variance of the kernel estimates by direct calculations and proved its asymptotic normality. Other properties of the kernel estimates under right censorship can be found from Tanner and Wang (1983), Lo et al. (1989), Diehl and Stute (1990), Xiang (1994), and so on.

In this paper, we use the smoothed empirical likelihood (EL) method to obtain confidence intervals for $h_0(t)$ and $f_0(t)$. The EL method was introduced by Owen (1988, 1990) as a method for constructing nonparametric confidence intervals. The advantages and references about empirical likelihood can be seen from Owen (2001). For complete data, Chen and Hall (1993) developed the smoothed empirical likelihood confidence regions for quantiles and proved it is Bartlett-correctable. Hall and Owen (1993) used empirical likelihood to construct confidence bands for density function, then Chen (1996, 1997) showed that the empirical likelihood produces confidence intervals having theoretical coverage accuracy of the same order of magnitude as the bootstrap, and which are empirically more accurate. The applications of empirical likelihood in survival analysis can be dated back to Thomas and Grunkemeier (1975) who constructed confidence intervals for survival probability with censored data (see also Li, 1995, Murphy, 1995). Empirical likelihood based confidence bands for individual quantile functions and survival functions have been derived by Li et al. (1996) and Hollander et al. (1997), respectively. However, the results about the application of smoothed empirical likelihood to censored data are few.

The paper is organized as follows. In Sect. 2, some empirical log-likelihood ratios are derived and their asymptotic limits are obtained. Simulation studies are given in Sect. 3. Proofs of the main results are put in Sect. 4.

2 Main results

For any cumulative distribution function F, let $\bar{F} = 1 - F$ and (a_F, b_F) be the range of F defined by

$$a_F := \inf\{x : F(x) > 0\} \quad \text{and} \quad b_F := \sup\{x : F(x) < 1\}.$$

Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ be the order statistics of the sample $X_1, X_2, \ldots, X_n$ and $\delta_{(i)}$ the concomitant of $X_{(i)}$ for $i = 1, \ldots, n$. Let

$$r_i := \sum_{j=1}^{n} I(X_j \geq X_{(i)}) = n - i + 1$$

be the number of subjects that are "alive" before $X_{(i)}$.

Let Γ be the space of all the distribution functions defined on $[0, \infty)$. For any $F \in \Gamma$, the likelihood function based on the censored data (1) is defined by

$$L(F) := \prod_{i=1}^{n} \Big(F(X_i) - F(X_i-)\Big)^{\delta_i} \Big(1 - F(X_i)\Big)^{1-\delta_i}.$$

From Li (1995), we know that the likelihood $L(F)$ can be rewritten as

$$L(F) = \prod_{i=1}^{n} \lambda_i^{\delta_{(i)}} (1 - \lambda_i)^{(r_i - \delta_{(i)})}.$$

where $\lambda_1, \ldots, \lambda_n$ are the hazard values at $X_{(1)}, \ldots, X_{(n)}$ given by

$$\begin{aligned} \lambda_i &:= \frac{F(X_{(i)}) - F(X_{(i)}-)}{1 - F(X_{(i)}-)} \\ &= P(X = x | X \geq x)|_{x = X_{(i)}}, \quad i = 1, 2, \ldots, n. \end{aligned} \tag{2}$$

Using (2), we can represent the survival function $S(t)$ and the cumulative hazard function $\Lambda(t) = -\ln S(t)$ as the following functions of $\lambda_1, \ldots, \lambda_n$:

$$S(t) = \prod_{i=1}^{n} (1 - \lambda_i) I(X_{(i)} \leq t), \quad \Lambda(t) = -\sum_{i=1}^{n} \ln(1 - \lambda_i) I(X_{(i)} \leq t).$$

Let $K(t)$ be a kernel function and $a = a_n$ be a smoothing bandwidth. Then the derivative $h_0(t) = \Lambda_0'(t)$ can be estimated by kernel smoothing method, such as

$$h_n(t|\lambda) = -\sum_{i=1}^{n} \ln(1 - \lambda_i) K_i(t),$$

where $K_i(t) = a^{-1} K(\frac{t - X_{(i)}}{a})$. Maximizing $L(F)$, we get $\hat{\lambda}_i = \delta_{(i)}/r_i, i = 1, \ldots, n$, and obtain

$$h_n(t) := h_n(t|\hat{\lambda}) = -\sum_{i=1}^{n} \ln\left(1 - \frac{\delta_{(i)}}{r_i}\right) K_i(t),$$

the familiar estimate of $h_0(t)$. Under constraint $h_n(t|\lambda) = h_0(t)$, we introduce the following empirical likelihood ratio

$$\mathcal{R}(h_0, t) := \frac{\sup\limits_{\lambda_1,\ldots,\lambda_n} \left\{L(F) : \sum_{i=1}^n \ln(1-\lambda_i)K_i(t) + h_0(t) = 0\right\}}{\sup\limits_{\lambda_1,\ldots,\lambda_n} L(F)}. \tag{3}$$

By Lagrange's method, we have

$$\ln \mathcal{R}(h_0, t) = \sum_{i=1}^n \left\{ (r_i - \delta_{(i)}) \ln \left(1 + \frac{\mu K_i(t)}{r_i - \delta_{(i)}}\right) - r_i \ln \left(1 + \frac{\mu K_i(t)}{r_i}\right) \right\}, \tag{4}$$

where the Lagrange multiplier μ satisfies

$$\sum_{i=1}^n \ln \left(1 - \frac{\delta_{(i)}}{r_i + \mu K_i(t)}\right) K_i(t) + h_0(t) = 0. \tag{5}$$

Let H_0 be the distribution of X, then $\bar{H}_0 = \bar{F}_0 \bar{G}_0$, $b_{H_0} = \min(b_{F_0}, b_{G_0})$. To study the asymptotic properties of $\mathcal{R}(h_0, t)$, we introduce the following conditions. Suppose

(C1) $K(t)$ is a bounded nonnegative function having compact support $[-c, c]$, such that

$$\int_{-\infty}^{\infty} u^i K(u) \mathrm{d}u = \begin{cases} 1, & i = 0, \\ 0, & 1 \le i \le r-1, \\ C_0, & i = r, \end{cases}$$

where C_0 is a nonzero constant. The derivative of $K(t)$ exists.

(C2) For $t \in (a_{F_0}, b_{H_0})$, suppose $h_0(t) > 0$. The derivative $h_0'(t)$ of $h_0(t)$ exists and is continuous.

(C3) As $n \to \infty$, we have $a \to 0$, $na \to \infty$, $n^{1/2} a^{r+\frac{1}{2}} \to 0$, $\ln a^{-1}/na \to 0$ and $\ln a^{-1}/\ln \ln n \to \infty$.

Theorem 1 *Assume the conditions* (C1)–(C3). *Then for each fixed* $t \in [a_{F_0}, b_{H_0}]$, *as* $n \to \infty$, *we have*

$$-2 \ln \mathcal{R}(h_0, t) \xrightarrow{\mathcal{D}} \chi_1^2. \tag{6}$$

By Theorem 1, a confidence interval for $h_0(t)$ with asymptotic coverage probability $1-\alpha$ each fixed $t \in [a_{F_0}, b_{H_0}]$ can be defined by

$$I_{n,\alpha}(h, t) := \{h : -2 \ln \mathcal{R}(h, t) \le C_\alpha\},$$

where C_α is given by $P(\chi_1^2 \le C_\alpha) = 1 - \alpha$.

Remark 1 To select an adapted bandwidth is very important as we use kernel smoothing method. Theoretically, the optimal bandwidth should be chosen as the value for which the coverage error is minimized. For complete data, Chen (1996) gave a explicit expression of coverage error and found that the optimal order of bandwidth is $n^{-1/3}$. We can not obtain the same result in our paper, but we believe that this result remain true for right censored data. In our simulation studies, we choose $a = cn^{-1/3}$ and select c by bootstrap method.

Construction of empirical likelihood for density function $f_0(t)$ is more complex. Notice that $f_0(t) = h_0(t)S_0(t)$. For any fixed t, if $f_0(t)$ is known, adding another constraint $S_0(t) = p$, we have $h_0(t) = f_0(t)/p$. Introduce the empirical likelihood ratio

$$\mathcal{R}(f_0, p, t) := \frac{\sup\limits_{\lambda_1,\ldots,\lambda_n} \left\{L(F) : \sum_{i=1}^{n} \ln(1-\lambda_i) I(X_{(i)} \leq t) = \ln p, \sum_{i=1}^{n} \ln(1-\lambda_i) K_i(t) + f_0(t)/p = 0\right\}}{\sup\limits_{\lambda_1,\ldots,\lambda_n} L(F)}$$

and define

$$\mathcal{R}(f_0, t) := \sup_{p\in(0,1)} \mathcal{R}(f_0, p, t). \tag{7}$$

Let

$$W_{ni} := \left(I(X_{(i)} \leq t), K_i(t)\right)^{\mathrm{T}}.$$

By Lagrange method we have

$$\ln \mathcal{R}(f_0, t) = \sum_{i=1}^{n} \left\{ (r_i - \delta_{(i)}) \ln\left(1 + \frac{\gamma^T W_{ni}}{r_i - \delta_{(i)}}\right) - r_i \ln\left(1 + \frac{\gamma^T W_{ni}}{r_i}\right)\right\}, \tag{8}$$

where the Lagrange multiplier $\gamma = (\gamma_1, \gamma_2)^T$ and nuisance parameter p satisfy

$$\sum_{i=1}^{n} \ln\left(1 - \frac{\delta_{(i)}}{r_i + \gamma^T W_{ni}}\right) I(X_{(i)} \leq t) - \ln p = 0, \tag{9}$$

$$\sum_{i=1}^{n} \ln\left(1 - \frac{\delta_{(i)}}{r_i + \gamma^T W_{ni}}\right) K_i(t) + f_0(t)/p = 0, \tag{10}$$

$$\frac{\gamma_1}{n} p + \frac{\gamma_2}{n} f_0(t) = 0. \tag{11}$$

For $i = 1, 2, 3$, we define the left hand sides of (9)–(11) by $Q_{in}(p, \gamma_1, \gamma_2)$. For fixed p, assume the solutions of (9) and (10) are $(\gamma_1, \gamma_2) = \left(\gamma_1(p, t), \gamma_2(p, t)\right)$. Plugging them in (11), we get

$$\frac{\gamma_1(p, t)}{n} p + \frac{\gamma_2(p, t)}{n} f_0(t) = 0. \tag{12}$$

Theorem 2 *Assume the conditions* (C1)–(C3). *Then for each* $t \in [a_{F_0}, b_{H_0}]$, *with probability 1 for large n,* (12) *has a solution* $p_E = p_E(t)$, *such that* $\mathcal{R}(f_0, p, t)$ *attains its maximum value at* $p = p_E$, *and as* $n \to \infty$, *we have*

$$-2\ln \mathcal{R}(f_0, p_E, t) \xrightarrow{\mathcal{D}} \chi_1^2. \tag{13}$$

Similarly, the asymptotic $100(1-\alpha)\%$ confidence interval for $f_0(t)$ for each fixed t by

$$I_{n,\alpha}(f, t) := \{f : -2\ln \mathcal{R}(f, p_E, t) \le C_\alpha\}.$$

3 Simulation results

We use Monte Carlo simulation to compare the empirical likelihood and the normal approximation method in term of coverage accuracy. We use the Epanechnikov kernel

$$K(u) = \begin{cases} \frac{3}{4}(1-u^2), & \text{if } |u| \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

and the smoothing parameter $a = cn^{-1/3}$. We choose c at interval $[0.5, 2]$ such that the bootstrap coverage probability close to $1-\alpha$ (See Li and Van Keilegom, 2002. Let F_0, G_0 be the exponential distributions with mean 1 and θ, respectively. We choose $\theta = 5$ to get 10% censoring rate and $\theta = 4$ to get 20% censoring rate and choose nominal coverage of $\alpha = 0.90$ and $\alpha = 0.95$ to compare the performance between EL intervals and normal approximation (NA) intervals. For fixed t, by the central limit theorem of $h_n(t)$ (Lo et al., 1989), since

$$\sqrt{na}\big(h_n(t) - h_0(t)\big) \xrightarrow{\mathcal{D}} N\big(0, \sigma^2(t)\big),$$

where $\sigma^2(t) = h_0(t)\bar{H}_0^{-1}(t)\int_{-1}^{1} K^2(u)du$, the asymptotic $100(1-\alpha)\%$ confidence intervals for $h_0(t)$ and $f_0(t)$ are defined by

$$\left(h_n(t) - \frac{\hat{\sigma}(t)}{\sqrt{na}} q(1-\alpha/2), \quad h_n(t) + \frac{\hat{\sigma}(t)}{\sqrt{na}} q(1-\alpha/2)\right)$$

and

$$\left(f_n(t) - \frac{\hat{\sigma}(t) S_n(t)}{\sqrt{na}} q(1-\alpha/2), \quad f_n(t) + \frac{\hat{\sigma}(t) S_n(t)}{\sqrt{na}} q(1-\alpha/2)\right),$$

respectively, where $f_n(t) = a^{-1}\int K((t-s)/a)dS_n(t)$, $S_n(t)$ is the Kaplan-Meier estimator based on data (1) and $q(\alpha)$ is α-quantile of the standard normal distribution. For $n = 40$, 80, 120 and $t = 0.6$, 1,000 duplications were calculated,

Table 1 Coverage probability for $t = 0.6$

Censoring rate	Size n_1	$h(t)$				$f(t)$			
		$\alpha = 0.90$		$\alpha = 0.95$		$\alpha = 0.90$		$\alpha = 0.95$	
		EL	NA	EL	NA	EL	NA	EL	NA
	40	0.907	0.847	0.960	0.894	0.889	0.920	0.921	0.970
0.10	80	0.899	0.879	0.937	0.908	0.899	0.920	0.939	0.960
	120	0.890	0.878	0.943	0.925	0.898	0.918	0.947	0.962
	40	0.912	0.853	0.958	0.889	0.895	0.934	0.936	0.983
0.20	80	0.902	0.864	0.956	0.917	0.902	0.931	0.946	0.970
	120	0.895	0.873	0.944	0.918	0.909	0.925	0.957	0.957

respectively. The results are summarized in Table 1. The bandwidth a is selected by bootstrap method.

Comparing the performance of empirical likelihood method and normal approximation method from the tables, it is seen that for different simple sizes and different censoring rate, the performance of EL method is better than that of the normal approximation, especially in case of small sample size n.

4 Proofs

We prove Theorem 2 first. The proof of Theorem 1 is similar and more easier. Define $\varepsilon_n = n^{-s}$, $1/3 < s < 1/2$. In what follows, we suppose the conditions of Theorem 2 are satisfied.

Lemma 1 *If p satisfies $|p - S_0(t)| \leq \varepsilon_n$, then for each fixed $t \in [a_{F_0}, b_{H_0}]$, the solution $\gamma = (\gamma_1(p), \gamma_2(p))^T$ of Eqs.* (9) *and* (10) *satisfies*

$$\frac{\gamma_1(p)}{n} = O(\varepsilon_n) \text{ a.s.,} \quad \frac{\gamma_2(p)}{n} = O(a^{1/2}\varepsilon_n) \text{ a.s.} \tag{14}$$

Proof Define

$$\tilde{\sigma}_1^2(t) := n \sum_{i=1}^{n} \frac{\delta_i I(X_{(i)} \leq t)}{r_i^2},$$
$$\tilde{\sigma}_2^2(t) := na \sum_{i=1}^{n} \frac{\delta_i K_i^2(t)}{r_i^2},$$
$$\tilde{\sigma}_{12}^2(t) := n \sum_{i=1}^{n} \frac{\delta_i K_i(t) I(X_{(i)} \leq t)}{r_i^2}.$$

Similar to the proof of Proposition 3.3.1 of Ramlau-Hansen (1983), it is easy to get

$$\begin{aligned}\tilde{\sigma}_1^2(t) &\to \sigma_1^2(t) \equiv \int_0^{\mathrm{t}} \frac{dF_0(s)}{\bar{F}_0(s)\bar{H}_0(s)} \text{ a.s.},\\ \tilde{\sigma}_2^2(t) &\to \sigma_2^2(t) \equiv \frac{h_0(t)}{\bar{H}_0(t)} \int_{-\mathrm{c}}^{\mathrm{c}} K^2(t)\mathrm{d}t \text{ a.s.},\\ \tilde{\sigma}_{12}^2(t) &\to \sigma_{12}^2(t) \equiv \frac{h_0(t)}{\bar{H}_0(t)} \int_0^{\mathrm{c}} K(t)\mathrm{d}t \text{ a.s.}\end{aligned} \tag{15}$$

Let $|\gamma| = \sqrt{\gamma_1^2 + \gamma_2^2}$. Define

$$A_n(p,t) := \Big(\ln p - \ln S_n(t), h_n(t) - f_0(t)/p\Big)^{\mathrm{T}}. \tag{16}$$

Because

$$\left[\ln\left(1 - \frac{\delta_{(i)}}{r_i + \gamma^T W_{ni}}\right) - \ln\left(1 - \frac{\delta_{(i)}}{r_i}\right)\right]\gamma^T W_{ni} \geq 0,$$

by (9), (10) and inequality $|\ln(1-x) - \ln(1-y)| \geq |x-y|$ for $x, y \in (0,1)$, we get

$$\begin{aligned}\gamma^T A_n(p,t) &= \sum_{i=1}^n \left(\ln\left(1 - \frac{\delta_{(i)}}{r_i + \gamma^T W_{ni}}\right) - \ln\left(1 - \frac{\delta_{(i)}}{r_i}\right)\right)\big(\gamma_1 I(X_{(i)} \leq t) + \gamma_2 K_i(t)\big)\\ &= \sum_{i=1}^n \left|\ln\left(1 - \frac{\delta_{(i)}}{r_i + \gamma^T W_{ni}}\right) - \ln\left(1 - \frac{\delta_{(i)}}{r_i}\right)\right| |\gamma^T W_{ni}|\\ &\geq \sum_{i=1}^n \left|\frac{\delta_{(i)}}{r_i + \gamma^T W_{ni}} - \frac{\delta_{(i)}}{r_i}\right| |\gamma^T W_{ni}|\\ &\geq \frac{1}{n + 2|\gamma| \max^*\{|n/r_i|\}}\Big(\gamma_1^2\hat{\sigma}_1^2(t) + 2\gamma_1\gamma_2\hat{\sigma}_{12}^2(t) + a^{-1}\gamma_2^2\hat{\sigma}_2^2(t)\Big),\end{aligned} \tag{17}$$

where $\max^*$ denotes $\max_{i:K_i(t)>0, X_{(i)} \leq t}$. Almost surely for large n,

$$\max^* \left|\frac{n}{r_i}\right| \leq \max_{i:X_{(i)} \leq t} \left|\frac{n}{r_i}\right| \leq \frac{2}{\bar{H}_0(t)},$$

we get

$$\gamma^T A_n(p,t) \geq \frac{1}{n + 4|\gamma|\bar{H}_0^{-1}(t)}\Big(\gamma_1^2\hat{\sigma}_1^2(t) + 2\gamma_1\gamma_2\hat{\sigma}_{12}^2(t) + a^{-1}\gamma_2^2\hat{\sigma}_2^2(t)\Big). \tag{18}$$

Let $\zeta_1 = \gamma_1(p), \zeta_2 = a^{-1/2}\gamma_2(p)$, $|\zeta| = \sqrt{\zeta_1^2 + \zeta_2^2}$. From the LIL of $S_n(t)$ (Csörgő and Horváth, 1983) and $h_n(t)$ (Diehl and Stute, 1988), for each fixed $t \in [a_{F_0}, b_{H_0}]$, we know that

$$\begin{aligned}
|\gamma^T A_n(p,t)| &\le |\gamma_1|\big|\ln p - \ln S_0(t)\big| + |\gamma_1|\big|\ln S_0(t) - \ln S_n(t)\big| \\
&\quad + |\gamma_2|\big|h_n(t) - h_0(t)\big| + |\gamma_2|\big|h_0(t) - \ln \frac{f_0(t)}{p}\big| \\
&= |\gamma_1| O(\varepsilon_n) + a^{-1/2}|\gamma_2| O(\varepsilon_n) \\
&= |\zeta| O(\varepsilon_n) \text{ a.s.} \qquad (19)
\end{aligned}$$

On the other hand, because $\hat{\sigma}_{12}^2(t) \ge \hat{\sigma}_1(t)\hat{\sigma}_2(t)$, we get $2\zeta_1\zeta_2\hat{\sigma}_{12}^2(t) \ge -\zeta_1^2\hat{\sigma}_1^2(t) - \zeta_2^2\hat{\sigma}_2^2(t)$. Thus almost surely for large n,

$$\begin{aligned}
\gamma_1^2\hat{\sigma}_1^2(t) + 2\gamma_1\gamma_2\hat{\sigma}_{12}^2(t) + a^{-1}\gamma_2^2\hat{\sigma}_2^2(t) &= \zeta_1^2\hat{\sigma}_1^2(t) + 2a^{1/2}\zeta_1\zeta_2\hat{\sigma}_{12}^2(t) + \zeta_2^2\hat{\sigma}_2^2(t) \\
&\ge (1 - a^{1/2})\left(\zeta_1^2\hat{\sigma}_1^2(t) + \zeta_2^2\hat{\sigma}_2^2(t)\right) \\
&\ge \frac{\zeta^2}{2} \min\left(\hat{\sigma}_1^2(t), \hat{\sigma}_2^2(t)\right). \qquad (20)
\end{aligned}$$

By (18)–(20), we get $|\zeta|/n = O(\varepsilon_n)$ a.s. That completes the proof. □

Lemma 2 *With probability one, for large n, there exists a solution p_E of* (11) *such that $\mathcal{R}(f_0, p, t)$ attain its maximum value at $p = p_E$.*

Proof For $X_{(i)} < t + hc < b_H$, by Taylor expansion, we get

$$\begin{aligned}
\ln\left(1 - \frac{\delta_{(i)}}{r_i + \gamma^T W_{ni}}\right) &= \ln\left\{1 - \frac{\delta_{(i)}}{r_i}\left(1 + \frac{\gamma^T W_{ni}}{r_i}\right)^{-1}\right\} \\
&= \ln\left\{1 - \frac{\delta_{(i)}}{r_i}\left(1 - \frac{\gamma^T W_{ni}}{r_i} + O\left(\frac{(\gamma^T W_{ni})^2}{r_i^2}\right)\right)\right\} \\
&= \ln\left(1 - \frac{\delta_{(i)}}{r_i}\right) + \ln\left\{1 + \left(1 - \frac{\delta_{(i)}}{r_i}\right)^{-1} \cdot \left(\frac{\gamma^T W_{ni}\delta_{(i)}}{r_i^2} + O\left(\frac{(\gamma^T W_{ni})^2\delta_{(i)}}{r_i^3}\right)\right)\right\} \\
&= \ln\left(1 - \frac{\delta_{(i)}}{r_i}\right) + \frac{\delta_{(i)}}{r_i(r_i - \delta_{(i)})}\gamma^T W_{ni} + O\left(\frac{\delta_{(i)}(\gamma^T W_{ni})^2}{r_i^3}\right) \text{ a.s.} \qquad (21)
\end{aligned}$$

Let $p_0 = S_0(t)$ and $p = p_0 + \varepsilon_n$. Assertion (14) leads to

$$\gamma^T W_{ni} = O(n\varepsilon_n) \text{ a.s.}$$

Thus

$$\begin{aligned} A_n(p,t) &= (\ln p - \ln S_n(t), h_n(t) - f_0(t)/p)^{\mathrm{T}} \\ &= \sum_{i=1}^{n} \left\{ \ln\left(1 - \frac{\delta_{(i)}}{r_i + \gamma^T W_{ni}}\right) - \ln\left(1 - \frac{\delta_{(i)}}{r_i}\right) \right\} W_{ni}^T \\ &= \frac{1}{n}\gamma^T \hat{\Sigma} + O\left(\varepsilon_n^2\right) \text{ a.s.,} \end{aligned} \tag{22}$$

where

$$\hat{\Sigma} := \begin{pmatrix} \hat{\sigma}_1^2(t) & \hat{\sigma}_{12}^2(t) \\ \hat{\sigma}_{12}^2(t) & a^{-1}\hat{\sigma}_2^2(t) \end{pmatrix},$$

with

$$\hat{\sigma}_1^2(t) := n \sum_{i=1}^{n} \frac{\delta_i I(X_{(i)} \le t)}{r_i(r_i - \delta_{(i)})},$$

$$\hat{\sigma}_{12}^2(t) := n \sum_{i=1}^{n} \frac{\delta_i K_i(t) I(X_{(i)} \le t)}{r_i(r_i - \delta_{(i)})},$$

$$\hat{\sigma}_2^2(t) := na \sum_{i=1}^{n} \frac{\delta_i K_i^2(t)}{r_i(r_i - \delta_{(i)})}.$$

Proceeding similarly as in the proof of (15), we get

$$\hat{\sigma}_1^2(t) \to \sigma_1^2(t), \hat{\sigma}_2^2(t) \to \sigma_2^2(t), \hat{\sigma}_{12}^2(t) \to \sigma_{12}^2(t) \text{ a.s.} \tag{23}$$

By (22) and (23) we get

$$\frac{1}{n}\gamma^T(p,t) = A_n(p,t)\hat{\Sigma}^{-1} + O(\varepsilon_n^2) \text{ a.s.} \tag{24}$$

By (4) and (24), we get

$$\begin{aligned}
-2\ln\mathcal{R}(f_0,p,t) &= -2\sum_{i=1}^{n}\left\{(r_i-\delta_{(i)})\ln\left(1+\frac{\gamma^T W_{ni}}{r_i-\delta_{(i)}}\right)-r_i\ln\left(1+\frac{\gamma^T W_{ni}}{r_i}\right)\right\}\\
&= \sum_{i=1}^{n}\frac{\delta_{(i)}}{r_i(r_i-\delta_{(i)})}(\gamma^T W_{ni})^2+O(n\varepsilon_n^3)\\
&= \frac{1}{n}\gamma^T(p,t)\hat{\Sigma}\gamma(p,t)+O(n\varepsilon_n^3)\\
&= nA_n^T(p,t)\hat{\Sigma}^{-1}A_n(p,t)+O(n\varepsilon_n^3)\\
&= n\left(A_n(p_0,t)+A_n'(p_1,t)\varepsilon_n\right)^T\hat{\Sigma}^{-1}\left(A_n(p_0,t)+A_n'(p_1,t)\varepsilon_n\right)\\
&\quad +O(n\varepsilon_n^3)\ \text{ a.s.,} \qquad (25)
\end{aligned}$$

where $p_1 \in (p_0, p_0+\varepsilon_n)$ and $A_n'(p_1,t) = (1/p_1, f_0(t)/p_1^2)^T$. Notice that

$$\left|\ln p_0 - \ln S_n(t)\right| = o(\varepsilon_n), \left|h_n(t)-f_0(t)/p_0\right| = o(a^{-1/2}\varepsilon_n)\ \text{a.s.,}$$

we get

$$-2\ln\mathcal{R}(f_0,p,t) \geq Cn\varepsilon_n^2\ \text{ a.s.,}$$

where C is a constant, and

$$\begin{aligned}
-2\ln\mathcal{R}(f_0,p_0,t) &= nA_n(p_0,t)^T\hat{\Sigma}^{-1}A_n(p_0,t)+O\left(n\varepsilon_n^3\right)\\
&= o\left(n\varepsilon_n^2\right)\ \text{ a.s.}
\end{aligned}$$

Hence when n is large enough, we have

$$-2\ln\mathcal{R}(f_0,p_0+\varepsilon_n,t) > -2\ln\mathcal{R}(f_0,p_0,t)\ \text{ a.s.}$$

Similarly we obtain, ultimately as $n\to\infty$,

$$-2\ln\mathcal{R}(f_0,p_0-\varepsilon_n,t) > -2\ln\mathcal{R}(f_0,p_0,t)\ \text{ a.s.}$$

So $-2\ln\mathcal{R}(f_0,p,t)$ attains its minimum in the region $(p_0-\varepsilon_n, p_0+\varepsilon_n)$, say at p_E. Then we have

$$\begin{aligned}
\frac{\partial\ln\mathcal{R}(f_0,p,t)}{\partial p}\Big|_{p=p_E} &= -\sum_{i=1}^{n}\frac{\delta_{(i)}\gamma^T W_{ni}}{(r_i+\gamma^T W_{ni})(r_i+\gamma^T W_{ni}-\delta_{(i)})}\frac{\partial\gamma^T W_{ni}}{\partial p}\Big|_{p=p_E}\\
&= 0. \qquad (26)
\end{aligned}$$

By (9) and (10), since $(\gamma_1, \gamma_2) = \big(\gamma_1(p,t), \gamma_2(p,t)\big)$ satisfies

$$Q_{1n}\Big(p, \gamma_1(p,t), \gamma_2(p,t)\Big) \equiv 0, \quad Q_{2n}\Big(p, \gamma_1(p,t), \gamma_2(p,t)\Big) \equiv 0,$$

we get

$$\begin{aligned}
&\gamma_1(p,t)\frac{\partial Q_{1n}\big(p,\gamma_1(p,t),\gamma_2(p,t)\big)}{\partial p} + \gamma_2(p,t)\frac{\partial Q_{2n}\big(p,\gamma_1(p,t),\gamma_2(p,t)\big)}{\partial p} \\
&\quad = \sum_{i=1}^{n} \frac{\delta_{(i)}\gamma^T W_{ni}}{(r_i+\gamma^T W_{ni})(r_i+\gamma^T W_{ni}-\delta_{(i)})}\frac{\partial \gamma^T W_{ni}}{\partial p} - \frac{1}{p^2}\Big(\gamma_1(p,t)p + \gamma_2(p,t)f_0(t)\Big) \\
&\quad \equiv 0.
\end{aligned}$$

So Eq. (26) is simplified to

$$\gamma_1(p_E,t)p_E + \gamma_2(p_E,t)f_0(t) = 0.$$

It means (12) has a solution at $p = p_E$. □

Proof of Theorem 2 Recall that the left hand sides of (9), (10), (11) are denoted by $Q_{1n}(p,\gamma_1,\gamma_2)$, $Q_{2n}(p,\gamma_1,\gamma_2)$ and $Q_{3n}(p,\gamma_1,\gamma_2)$, respectively. For $i = 1, 2$, let $\eta_i = \gamma_i/n$. Define

$$\begin{aligned}
\hat{S}_n(p) &\equiv \frac{\partial(Q_{1n}, Q_{2n}, Q_{3n})}{\partial(p, \eta_1, \eta_2)}\bigg| (p_0, 0, 0) \\
&= \begin{pmatrix} -1/p & \hat{\sigma}_1^2(t) & \hat{\sigma}_{12}^2(t) \\ -f_0(t)/p & \hat{\sigma}_{12}^2(t) & a^{-1}\hat{\sigma}_2^2(t) \\ 0 & p & f_0(t) \end{pmatrix}.
\end{aligned}$$

Let $\eta_{1E} = \eta_1(p_E)$ and $\eta_{2E} = \eta_2(p_E)$. Recall ${\varepsilon_n}^2 = o(n^{-1/2})$, by Taylor expansion we get

$$\begin{aligned}
\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} &= \begin{pmatrix} Q_{1n}(p_E, \eta_{1E}, \eta_{2E}) \\ Q_{2n}(p_E, \eta_{1E}, \eta_{2E}) \\ Q_{3n}(p_E, \eta_{1E}, \eta_{2E}) \end{pmatrix} \\
&= \begin{pmatrix} Q_{1n}(p_0, 0, 0) \\ Q_{2n}(p_0, 0, 0) \\ Q_{3n}(p_0, 0, 0) \end{pmatrix} + \hat{S}_n(p_0)\begin{pmatrix} p_E - p_0 \\ \eta_{1E} \\ \eta_{2E} \end{pmatrix} + O(\varepsilon_n^2) \ \text{ a.s.}
\end{aligned}$$

By (23), since

$$a \det\big(\hat{S}_n(p_0)\big) = \hat{\sigma}_2^2(t) + a\hat{\sigma}_1^2(t)f_0^2(t)/p_0^2 - 2a\hat{\sigma}_{12}^2(t)f_0(t)/p_0 \to \sigma_2^2(t) \text{ a.s.}$$

and $\sigma_2^2(t) > 0$, so $\hat{S}_n(p_0)$ has full rank and its inversion exists almost surely as $n \to \infty$. We get

$$\begin{pmatrix} p_E - p_0 \\ \eta_{1E} \\ \eta_{2E} \end{pmatrix} = -\hat{S}_n^{-1}(p_0) \begin{pmatrix} Q_{1n}(p_0,0,0) \\ Q_{2n}(p_0,0,0) \\ Q_{3n}(p_0,0,0) \end{pmatrix} + o(n^{-\frac{1}{2}}), \text{ a.s.}$$
$$= \frac{Q_{1n}(p_0,0,0)}{\det\left(\hat{S}_n(p_0)\right)} \begin{pmatrix} f_0(t)\hat{\sigma}_{12}^2(t) - a^{-1}p_0\hat{\sigma}_2^2(t) \\ f_0^2(t)/p^2 \\ -f_0(t)/p \end{pmatrix}$$
$$+ \frac{Q_{2n}(p_0,0,0)}{\det\left(\hat{S}_n(p_0)\right)} \begin{pmatrix} -f_0(t)\hat{\sigma}_1^2(t) + p_0\hat{\sigma}_{12}^2(t) \\ -f_0(t)/p \\ 1 \end{pmatrix} + o(n^{-\frac{1}{2}}) \text{ a.s.} \tag{27}$$

Thus

$$\eta_{2E} = \left(\det\left(\hat{S}_n(p_0)\right)\right)^{-1}\left(Q_{2n} - f_0(t)Q_{1n}/p_0\right) + o(n^{-\frac{1}{2}}). \tag{28}$$

By (25) we have

$$-2\ln\mathcal{R}(f_0, p_E, t) = \frac{1}{n}\gamma^T(p_E,t)\hat{\Sigma}\gamma(p_E,t) + O(n\varepsilon_n^3)$$
$$= n\gamma_2^2 \det\left(\hat{S}_n(p_E)\right) + O(n\varepsilon_n^3)$$
$$= n\left(\det\left(\hat{S}_n(p_0)\right)\right)^{-2}\det\left(\hat{S}_n(p_E)\right)\left(Q_{2n} - f_0(t)Q_{1n}/p_0\right)^2 + o_p(1).$$

Notice $p_E = p_0 + o_p(1)$, $Q_{1n} = o_p(Q_{2n})$, $\hat{\sigma}_1^2 = o_p(a^{-1}\hat{\sigma}_2^2)$ and $\hat{\sigma}_{12}^2 = o_p(a^{-1}\hat{\sigma}_2^2)$, we get

$$-2\ln\mathcal{R}(f, p_E, t) = \frac{na\left(h_n(t) - h_0(t)\right)^2}{\hat{\sigma}_2^2(t)} + o_p(1).$$

Because $\hat{\sigma}_2^2(t) = \hat{\sigma}^2(t) = \sigma^2(t) + o_p(1)$, we complete the proof by the central limit theorem of $h_n(t)$ (Lo et al., 1989).

Proof of Theorem 1 The proof is similar to that of Theorem 2. We only give the outlines. Similar to the proof of (17), almost surely for large n, we get

$$\mu(h_n(t) - h_0(t)) = \sum_{i=1}^{n}\left|\ln\left(1 - \frac{\delta_{(i)}}{r_i + \mu K_i(t)}\right) - \ln\left(1 - \frac{\delta_{(i)}}{r_i}\right)\right||\mu K_i(t)|$$
$$\geq \frac{a^{-1}\mu^2\hat{\sigma}^2(t)}{n + 2|\mu| \max\limits_{i:K_i(t)>0}\{|n/r_i|\}}$$
$$\geq \frac{\mu^2\sigma^2(t)}{2a\left(n + 4|\mu|\bar{H}_0^{-1}(t)\right)}.$$

By $|h_n(t) - h_0(t)| = O_p((na)^{-1/2})$, we get

$$\mu/n = O_p(a^{1/2}n^{-1/2}).$$

Similar to (24), we calculate

$$\frac{\mu}{n} = \frac{a\big(h_n(t) - h_0(t)\big)}{\hat{\sigma}^2} + O_p(an^{-1}).$$

Using Taylor expression to Eq. (4), we get

$$\begin{aligned}
-2\ln\mathcal{R}(h_0,t) &= \frac{\mu^2\hat{\sigma}^2(t)}{nh} + O(an^{-1/2}) \\
&= \frac{na\big(h_n(t) - h_0(t)\big)^2}{\hat{\sigma}^2(t)} + o_p(1).
\end{aligned}$$

Thus

$$-2\ln\mathcal{R}(h_0,t) \xrightarrow{\mathcal{D}} \chi_1^2.$$

Acknowledgments Research Supported by NSFC (10231030) and RFDP. We are grateful to referees for their valuable comments and suggestions.

References

Chen, S.X. (1996). Empirical likelihood confidence intervals for nonparametric density estimation. *Biometrika*, *83*, 329–341.

Chen, S.X. (1997). Empirical likelihood-based kernel density estimation. *The Australian Journal of Statistics*, *39*, 47–56.

Chen, S.X., Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Annals of Statistics*, *21*, 1166–1181.

Csörgő, S., Horváth, P. (1983). The rate of strong uniform consistency for the product-limit estimator. *Probability Theory and Related Fields*, *62*, 411–426.

Diehl, S., Stute, W. (1988). Kernel density and hazard function estimation in the presence of censoring. *Journal of Multivariate Analysis*, *25*, 299–310.

Hall, P., Owen A.B. (1993). Empirical likelihood confidence bands in density estimation. *Journal of Computational Graphical Statistics*, *2*, 273–289.

Hollander, M., McKeague, I.W., Yang, J. (1997). Likelihood ratio-based confidence bands for survival functions. *Journal of the American Statistical Association*, *92*, 215–227.

Li, G. (1995). On nonparametric likelihood ratio estimation of survival probabilities for censored data. *Statistics and Probability Letters*, *25*, 95–104.

Li, G., Hollander, M., McKeague, I.W., Yang, J. (1996). Nonparametric likelihood ratio confidence bands for quantile functions from incomplete survival data. *Annals of Statistics*, *24*, 628–640.

Li, G., Van Keilegom, I. (2002) Likelihood ratio confidence bands in nonparametric regression with censored data. *Scandinavian Journal of Statistics*, *29*, 547–562.

Lo, S.H., Mack, Y.P., Wang, J.L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan–Meier estimator. *Probability Theory and Related Fields*, *80*, 461–473.

Murphy, S.A. (1995). Likelihood ratio-based confidence intervals in survival analysis. *Journal of the American Statistical Association*, *90*, 1399–1406.

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

Owen, A.B. (1990). Empirical likelihood ratio confidence regions. *Annals of Statistics*, *18*, 90–120.

Owen, A.B. (2001). *Empirical likelihood*. London: Chapman and Hall.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Annals of Statistics*, *11*, 453–466.

Tanner, M.A. (1983). A note on the variable kernel estimator of the hazard function from randomly censored data. *Annals of Statistics*, *11*, 994–998.

Tanner, M.A., Wang, W.H. (1983). The estimation of the hazard function from randomly censored data by the kernel method. *Annals of Statistics*, *11*, 989–993.

Thomas, D.R., Grunkemeier, G.L. (1975). Confidence interval estimation of survival probabilities for censored data. *Journal of the American Statistical Association*, *70*, 865–871.

Xiang, X. (1994). Law of the logarithm for density and hazard rate estimation for censored data. *Journal of Multivariate Analysis*, *49*, 278–286.