International Journal of Statistics and Management System
Vol. 5 No. 1 (January-June, 2020)

Submitted: 1st March 2020   Revised: 11th April 2020   Accepted: 15th April 2020

The Lipschitz condition in the expansion of weighted empirical log-likelihood ratio *

JIAN-JIAN REN †

Abstract

So far, there has not been any work on likelihood-based interval estimates with partly interval-censored data. In this article, we derive the high-order expansion of the weighted empirical log-likelihood ratio for survival probabilities with right censored data, doubly censored data and partly interval-censored data, and we show that if the Lipschitz condition is satisfied for the distribution of the leading term(s) of this expansion, the *theoretical coverage accuracy equation* for the weighted empirical likelihood ratio confidence intervals (WELRCI) can be obtained. When there is no censoring, such Lipschitz condition is established in an example where smoothing based on the kernel density method is used. Simulation results show that WELRCI for survival probabilities compare well with those empirical likelihood-based methods and other alternative methods.

1. Introduction Since Owen (1988), the empirical likelihood method has been developed to construct tests and confidence sets based on the nonparametric likelihood ratio. For more references, see Owen (1990, 1991), DiCiccio, Hall and Romano (1991), Qin and Lawless (1994), Mykland (1995), among others. A recent book titled '*Empirical Likelihood*' by Owen (2001) provides a quite complete review of the developments on this topic.

*This research was partially supported by NSF Grants DMS-0204182 and DMS-0604488.

Key words and phrases: Bootstrap; doubly censored data; weighted empirical likelihood; interval censored data; partly interval-censored data; right censored data.

AMS 2000 subject classifications. Primary 62F25; secondary 60F10.

†*Mailing Address:* Department of Mathematics, University of Central Florida, Orlando, FL 32816. *E–mail:* jren@mail.ucf.edu.

Studies have shown that the empirical log-likelihood ratio usually has an asymptotic chi-squared distribution, and that the empirical likelihood ratio inference is of comparable accuracy to alternative methods. In particular, it is shown that the empirical likelihood is Bartlett-correctable for smooth function models (DiCiccio, Hall and Romano, 1991).

However, we no longer have a smooth function model for censored data, such as right censored data (Kaplan and Meier, 1958), doubly censored data (Turnbull, 1974; Gu and Zhang, 1993) and partly interval-censored data (Huang, 1999), while it is known that right censored data are frequently encountered in biomedical research or reliability studies, and that doubly censored data have recently been encountered in breast cancer research (Ren and Peer, 2000), and partly interval-censored data in heart disease (Odell, Anderson and D'Agostino; 1992) and diabetes studies (Enevoldsen et al., 1987). Generally, the empirical likelihood ratio does not have an analytic expression, which makes it difficult to study its asymptotic properties and the coverage accuracy of the empirical likelihood-based confidence intervals with censored data. So far, for censored data the available works based on empirical likelihood include Thomas and Grunkemeir (1975), Li, Hollander, McKeague and Yang (1996), Murphy and van der Vaart (1997), Banerjee and Wellner (2004), but none of these contains any coverage accuracy results, and none of these deals with likelihood inferences with partly interval-censored data.

Recently, Ren (2001) used a new likelihood function, called *weighted empirical likelihood function*, to construct confidence intervals for the mean with various types of censored data, including right censored data, doubly censored data, interval censored data and partly interval-censored data. For different types of censored data, this weighted empirical likelihood function is formulated in a unified form depending only upon the weights of the *nonparametric maximum likelihood estimator* (NPMLE) $\hat{F}_n$ for the underlying lifetime distribution F_0. Thus, the weighted empirical likelihood-based confidence intervals can be computed by the same algorithm for any given $\hat{F}_n$, and their asymptotic properties can be studied in a unified form through the weighted empirical log-likelihood ratio for different types of censored data.

In this article, we derive the high-order expansion of the weighted empirical log-likelihood ratio for survival probabilities with right censored data, doubly censored data and partly interval-censored data, and we show that if the Lipschitz condition is satisfied for the distribution of the leading term(s) of this expansion, the *theoretical coverage accuracy equation* for the weighted empirical likelihood ratio confidence intervals (WELRCI) can be obtained.

When there is no censoring, such Lipschitz condition is established in an example where a smoothed $\hat{F}_n$ based on the kernel density method is used. The implication of this example is that it is possible to have the Lipschitz condition for censored data with an appropriate smoothing of $\hat{F}_n$, thus the theoretical coverage accuracy equation above mentioned may be used as guidance for selecting the order of the expansion in practice. Our simulation results show that WELRCI for survival probabilities compare well with those empirical likelihood-based methods and other alternative methods.

The rest of the paper is organized as follows: Section 2 introduces the weighted empirical likelihood function with a review of the asymptotic properties of the NPMLE $\hat{F}_n$; Section 3 constructs *weighted empirical likelihood ratio confidence interval* (WELRCI) for survival probabilities, and discusses its theoretical coverage accuracy and the related Lipschitz condition aforementioned, while the proofs and the Lipschitz condition example are deferred to the Appendix; Section 4 gives some simulation results with comparison between WELRCI and those by alternative methods, and includes some concluding remarks.

2. Weighted empirical likelihood

In Owen (1988), the empirical likelihood function is given by

$$L(F) = \prod_{i=1}^{n} [F(X_i) - F(X_i-)], \tag{2.1}$$

and the empirical likelihood ratio function is given by

$$R(F) = L(F)/L(F_n), \tag{2.2}$$

where F is any distribution function (d.f.), $X_1, \ldots, X_n$ is a random sample from d.f. F_0, and the empirical d.f. F_n of sample $X_1, \ldots, X_n$ is the *nonparametric maximum likelihood estimator* (NPMLE) for F_0; that is F_n maximizes $L(F)$ over all distribution functions F.

For each type of censored data aforementioned, the likelihood function has been given, and the NPMLE $\hat{F}_n$ is the solution which maximizes the likelihood function. See Mykland and Ren (1996) for doubly censored data, which includes the right censored data as a special case; and Huang (1999) for partly interval-censored data. As a general notation for this paper, we let $\hat{F}_n$ be the NPMLE of F_0 based on the observed censored data, then from observed data points there exist m distinct points $W_1 < W_2 < \cdots < W_m$ along with

$\hat{p}_j > 0, 1 \le j \le m$, such that $\hat{F}_n$ can be expressed as

$$\hat{F}_n(x) = \sum_{i=1}^{m} \hat{p}_i \, I\{W_i \le x\}, \tag{2.3}$$

for right censored data (Kaplan and Meier, 1958), doubly censored data (Mykland and Ren, 1996), and partly interval-censored data (Huang, 1999). In Ren (2001), the *weighted empirical likelihood function* for censored data is given by

$$\hat{L}(F) = \prod_{i=1}^{m} [F(W_i) - F(W_i-)]^{n\hat{p}_i}, \tag{2.4}$$

where F is any d.f., and $W_i, \hat{p}_i$ are as in (2.3). It is easy to show that $\hat{L}(F)$ is maximized at $\hat{F}_n$. Thus the *weighted empirical likelihood ratio* is given by

$$\hat{R}(F) = \hat{L}(F)/\hat{L}(\hat{F}_n). \tag{2.5}$$

One may note that when there is no censoring, $W_1 < \cdots < W_m$ in (2.3) are all distinct observations in the random sample $X_1, \ldots, X_n$, thus the weighted empirical likelihood function (2.4) coincides with the empirical likelihood function (2.1) by Owen (1988).

Remark 1. *Asymptotic Results of NPMLE $\hat{F}_n$:* Letting $\|\cdot\|$ stand for the uniform norm, it is known that for $\hat{F}_n$ given by (2.3), we have $\|\hat{F}_n - F_0\| \xrightarrow{a.s.} 0$ for right censored data (Stute and Wang, 1993), doubly censored data (Gu and Zhang, 1993), and partly interval-censored data (Huang, 1999), respectively. Also, it is shown that for each type of these censored data, $\hat{F}_n$ is of $\sqrt{n}$ convergence rate; in fact, under certain conditions, $\sqrt{n}\,(\hat{F}_n - F_0)$ weakly converges to a centered Gaussian process (Gill, 1983; Gu and Zhang, 1993; Huang, 1999).

3. Confidence intervals for probabilities

In survival analysis, it is often of interest to construct confidence intervals for the survival probability $(1 - \theta_0)$, where for a constant $t_0 > 0$,

$$\theta_0 = F_0(t_0) \tag{3.1}$$

is the probability at t_0 for the underlying lifetime distribution F_0. In this section, we show that set $S_n = \{F(t_0) \,|\, \hat{R}(F) \ge c_n, F \ll \hat{F}_n\}$ may be used as confidence interval for θ_0 with right censored data, doubly censored data and partly interval-censored data, where constant $0 < c_n < 1$ is set by equation (3.9), and '$F \ll \hat{F}_n$' means that F is absolutely continuous with respect to $\hat{F}_n$.

First, note that the NPMLE $\hat{F}_n$ for censored data is not always a proper d.f. (Mykland and Ren, 1996), but in this work we always consider the adjusted version of the NPMLE, still denoted as $\hat{F}_n$. Precisely, for the rest of this paper, $\hat{F}_n$ in (2.3) denotes the proper d.f. obtained by setting 1 as the value of the NPMLE at the largest observation in the data set, which implies $\sum_{i=1}^m \hat{p}_i = 1$ in (2.3). This kind of adjustment of the NPMLE is a generally adopted convention for censored data (Efron, 1967; Miller, 1976). Although this $\hat{F}_n$ in (2.3) no longer necessarily maximizes the underlying likelihood function, the usual asymptotic properties of the NPMLE needed for this work still hold for this $\hat{F}_n$, because the work here only concerns its asymptotic behavior around point t_0.

To state our main results in this section, we let

$$r(\theta) = \sup \left\{ \prod_{i=1}^m (p_i/\hat{p}_i)^{n\hat{p}_i} \;\middle|\; \sum_{i=1}^m p_i \, I\{W_i \le t_0\} = \theta, \, p_i \ge 0, \, \sum_{i=1}^m p_i = 1 \right\}, \tag{3.2}$$

where $p_i = F(W_i) - F(W_i-), 1 \le i \le m$. In the Appendix, it is shown that S_n is an interval satisfying $S_n = [X_L, X_U]$ and

$$X_L \le \theta_0 \le X_U \quad \text{if and only if} \quad r(\theta_0) \ge c_n, \tag{3.3}$$

where $0 < \theta_0 < 1$ and

$$\begin{aligned} X_L &= \inf \left\{ \textstyle\sum_{i=1}^m p_i I\{W_i \le t_0\} \;\middle|\; p_i \ge 0, \, \sum_{i=1}^m p_i = 1, \, \prod_{i=1}^m (p_i/\hat{p}_i)^{n\hat{p}_i} \ge c_n \right\} \\ X_U &= \sup \left\{ \textstyle\sum_{i=1}^m p_i I\{W_i \le t_0\} \;\middle|\; p_i \ge 0, \, \sum_{i=1}^m p_i = 1, \, \prod_{i=1}^m (p_i/\hat{p}_i)^{n\hat{p}_i} \ge c_n \right\}. \end{aligned} \tag{3.4}$$

We call $[X_L, X_U]$ *weighted empirical likelihood ratio confidence interval* (WELRCI) for θ_0. Since (3.3) implies

$$P\{X_L \le \theta_0 \le X_U\} = P\{-2\log r(\theta_0) \le -2\log c_n\}, \tag{3.5}$$

the asymptotic behavior of $[X_L, X_U]$ is studied through *weighted empirical log-likelihood ratio* $\log r(\theta_0)$ in the following theorem with proofs deferred to the Appendix.

Theorem 3.1. *Assume that $n \to \infty$,*

$$\sqrt{n}[\hat{F}_n(t_0) - F_0(t_0)] \overset{D}{\longrightarrow} \mathbb{Z}_0 \overset{D}{=} N(0, \sigma_0^2), \tag{AS1}$$

$$\hat{F}_n(t_0) \overset{a.s.}{\longrightarrow} F_0(t_0). \tag{AS2}$$

Then for $0 < \theta_0 = F_0(t_0) < 1$ *and* $\hat{\theta} = \hat{F}_n(t_0)$,

(i) We have that for $0 < \hat{\theta} < 1$,

$$-2 \log r(\theta_0) = B_n^{(k)} + n(\hat{\theta} - \theta_0)^{k+3} r_{n,k}, \tag{3.6}$$

where for fixed k, *there exists a constant* $1 \leq M_{r,k} < \infty$ *such that* $|r_{n,k}| \leq M_{r,k}$ *all but finitely often with probability 1, and*

$$B_n^{(k)} = \frac{n(\hat{\theta} - \theta_0)^2}{\hat{\theta}(1 - \hat{\theta})} \Big(1 + \textstyle\sum_{j=1}^{k} \hat{a}_j (\hat{\theta} - \theta_0)^j \Big), \quad k = 1, 2, 3, 4 \tag{3.7}$$

with

$$\hat{a}_1 = \frac{2[(1-\hat{\theta})^2 - \hat{\theta}^2]}{3\,\hat{\theta}\,(1-\hat{\theta})}, \ \hat{a}_2 = \frac{[(1-\hat{\theta})^3 + \hat{\theta}^3]}{2\,[\hat{\theta}\,(1-\hat{\theta})]^2}, \ \hat{a}_3 = \frac{2[(1-\hat{\theta})^4 - \hat{\theta}^4]}{5[\hat{\theta}\,(1-\hat{\theta})]^3}, \ \hat{a}_4 = \frac{[(1-\hat{\theta})^5 + \hat{\theta}^5]}{3[\hat{\theta}\,(1-\hat{\theta})]^4};$$

(ii) Assuming $c_n = O(1)$, *we have that*

$$P\{X_L \leq \theta_0 \leq X_U\} = P\{B_n^{(k)} \leq -2 \log\, c_n\} + O(\|G_{n,k} - G_0\|) + O(n^{-(k+1)/2}), \tag{3.8}$$

where $G_{n,k}$ *and* G_0 *are the d.f.'s of* $B_n^{(k)}$ *and* $\mathbb{Z}_0^2[\theta_0(1-\theta_0)]^{-1}$, *respectively.*

In practice, we let $\rho_{n,\alpha}^{(k)}$ be the $(1-\alpha)100th$ percentile of $B_n^{(k)}$ in (3.7) for $0 < \alpha < 1$, then $[X_L^{(k)}, X_U^{(k)}]$ computed by (3.4) based on constant c_n set by

$$-2 \log\, c_n = \rho_{n,\,\alpha}^{(k)}, \tag{3.9}$$

is called the *kth order weighted empirical likelihood ratio confidence interval* (k-WELRCI) for θ_0. Thus, from (3.8)-(3.9) we have

$$P\{X_L^{(k)} \leq \theta_0 \leq X_U^{(k)}\} = (1-\alpha) + O(\|G_{n,k} - G_0\|) + O(n^{-(k+1)/2}). \tag{3.10}$$

Remark 2. From Remark 1 in Section 2, we know that under certain conditions, (AS1)-(AS2) hold for censored data under consideration here. It should be noted that for chi-squared random variable (r.v.) χ_1^2, (3.6)-(3.7) imply $-2 \log\, r(\theta_0) \overset{D}{\approx} n(\hat{\theta} - \theta_0)^2[\hat{\theta}(1-\hat{\theta})]^{-1} \overset{D}{\approx} \sigma_0^2[\theta_0(1-\theta_0)]^{-1}\chi_1^2$, which is a *scaled* chi-squared distribution. Thus, $-2 \log\, r(\theta_0)$ does not have an asymptotic chi-squared distribution as in Thomas and Grunkemeir (1975) and Murphy and van der Vaart (1997). It is known that having an asymptotic chi-squared distribution indicates that the likelihood ratio (or the statistic of interest) is studentized (i.e., free of the underlying parameters of interest), in turn, usually the likelihood-based method better catches the skewness of the distribution of the statistics of interest than that based on asymptotic normality. Nonetheless, our simulation studies in Section 4 show that the k-WELRCI $[X_L^{(k)}, X_U^{(k)}]$ preserves this appealing feature of the empirical likelihood method. The reason for this is as follows. Letting $\chi^2_{1,\alpha}$ be the $(1-\alpha)100th$ percentile of χ^2_1, the fact of $\rho^{(k)}_{n,\alpha} \approx \sigma_0^2[\theta_0(1-\theta_0)]^{-1}\chi^2_{1,\alpha}$ suggests us to adjust the weights of the weighted empirical likelihood to studentize the likelihood ratio, i.e., replacing $n\hat{p}_i$ by $\sigma_0^{-2}[\theta_0(1-\theta_0)]n\hat{p}_i$ in (3.2), then the consequent $-2\log r(\theta_0)$ has an asymptotic chi-squared distribution, in turn, $-2\log c_n$ should be set as $\chi^2_{1,\alpha}$. To compute the confidence interval based on this studentized likelihood ratio, we use (3.4) with weights $n\hat{p}_i$ replaced by $\sigma_0^{-2}[\theta_0(1-\theta_0)]n\hat{p}_i$, and we have $\prod_{i=1}^m (p_i/\hat{p}_i)^{\sigma_0^{-2}\theta_0(1-\theta_0)n\hat{p}_i} \geq c_n = \exp\{-\frac{1}{2}\chi^2_{1,\alpha}\}$, which is equivalent to $-2n\sum_{i=1}^m \hat{p}_i \log(p_i/\hat{p}_i) \leq \sigma_0^2[\theta_0(1-\theta_0)]^{-1}\chi^2_{1,\alpha} \approx \rho^{(k)}_{n,\alpha}$. Thus, we know that the resulting confidence interval is asymptotically the same as the k-WELRCI $[X_L^{(k)}, X_U^{(k)}]$. This could be viewed as that the use of $\rho^{(k)}_{n,\alpha}$ in k-WELRCI procedure is actually studentizing the weighted empirical likelihood ratio.

The Lipschitz Condition in (3.10):

From the proof of Theorem 1 (ii) in the Appendix, it is easy to see that if $G_{n,k}$ satisfies the Lipschitz condition, term $O(\|G_{n,k} - G_0\|)$ in equation (3.8) or (3.10) disappears. However, for $\hat{\theta} = \hat{F}_n(t_0)$ the Lipschitz condition generally does not hold for $G_{n,k}$. For instance, if there is no censoring, the NPMLE for the complete i.i.d. sample $X_1, \ldots, X_n$ is the empirical d.f. F_n, and the d.f. of $\sqrt{n}\,(\hat{\theta} - \theta_0) = \sqrt{n}[F_n(t_0) - F_0(t_0)]$ is discrete. Thus, from (3.7) it is easy to see that $G_{n,k}$ does not satisfy the Lipschitz condition when there is no censoring. This leads us to consider using a smoothed version of the NPMLE $\hat{F}_n$ in Theorem 1 in order to eliminate term $O(\|G_{n,k} - G_0\|)$ in (3.8).

Specifically, let $\tilde{F}$ be a continuous d.f. which is the continuous version of a discrete d.f. F according to a given smoothing method. For instance, $\tilde{F}$ could be obtained by simply connecting the jump points of F through straight lines, or by the kernel density method which is discussed in the Appendix. Let $\tilde{F}_n$ be the smoothed version of $\hat{F}_n$, and let $\tilde{\theta} = \tilde{F}_n(t_0)$. Then, with minor modification of the proof of Theorem 1, we have (3.6) for using the smoothed $\tilde{F}$ in (3.2), where $B_n^{(k)}$ is replaced by

$$\tilde{B}_n^{(k)} = \frac{n(\tilde{\theta} - \theta_0)^2}{\tilde{\mu}_2} \left(1 + \sum_{j=1}^{k} \tilde{a}_j (\tilde{\theta} - \theta_0)^j \right), \quad k = 1, 2, 3, 4, \tag{3.11}$$

with $\tilde{a}_j$'s calculated according to smoothed $\tilde{F}$ in (3.2). Moreover, if $\tilde{G}_{n,k}$ is the d.f. of $\tilde{B}_n^{(k)}$ and satisfies the following Lipschitz condition:

$$|\tilde{G}_{n,k}(x) - \tilde{G}_{n,k}(y)| \leq M_k|x - y|, \tag{AS3}$$

for all x and y in some neighborhood of c_n, the proof of Theorem 1 (ii) gives

$$P\{\tilde{X}_L \leq \theta_0 \leq \tilde{X}_U\} = P\{\tilde{B}_n^{(k)} \leq -2\log c_n\} + O(n^{-(k+1)/2}), \tag{3.12}$$

where $[\tilde{X}_L^{(k)}, \tilde{X}_U^{(k)}]$ is the k-WELRCI based on above mentioned smoothing.

However, verifying (AS3) can be quite involved. In the Appendix, an example of smoothing based on the kernel density method is discussed, where (AS3) is established up to a remainder term converging to 0 in exponential rate when there is no censoring. The implication of this example is that it is possible to have (AS3) and (3.12) for censored data with an appropriate smoothing of $\hat{F}_n$, thus the *theoretical coverage accuracy equation* (3.12) may be used as guidance for the selection of k in practice.

Remark 3. *Choice of k:* If the complete random sample $X_1, \ldots, X_n$ is available, then with $k = 4$ in (3.12), the theoretical coverage accuracy is $O(n^{-5/2})$. We know that the coverage accuracy with Bartlett-correction is only $O(n^{-2})$ for smooth function models (DiCiccio, Hall and Romano, 1991). Thus, the use of the $4th$ order (not higher) expansion of the log-likelihood ratio in Theorem 1 or (3.12) for censored data is usually sufficient. Moreover, since the coverage accuracy of empirical likelihood-based confidence intervals is $O_p(n^{-1})$ for non-censored data (DiCiccio, Hall and Romano; 1991), in practice we suggest that the use of $k = 2$ in (3.12) or (3.8) should be sufficient.

4. Simulation Since the WELRCI based on smoothing through the kernel density method involves issues such as band-width selection, we do not consider them here. In this section, we present some simulation results on Theorem 1. Since this problem was considered by Thomas and Grunkemeir (1975) for right censored data, denoted as TGCI, we make comparisons between WELRCI and TGCI. Moreover, other procedures such as bootstrap-t confidence interval and bootstrap percentile confidence interval (BPCI) (Efron and Tibshirani, 1993), are also considered in our studies. We find that bootstrap-t performs erratically, thus the results using bootstrap-t are not included here. As follows, we discuss some computational issues on WELRCI, then present some simulation results.

Here, we compute k-WELRCI $[X_L^{(k)},\, X_U^{(k)}]$ by solving the corresponding optimization problems in (3.4), where $\rho_{n,\alpha}^{(k)}$ in (3.9) is estimated by the n out of n bootstrap method. Specifically, $\rho_{n,\alpha}^{(k)}$ is estimated by the $(1-\alpha)100th$ percentile of

$$B_n^{(k)*} = \frac{n(\hat{\theta}^* - \hat{\theta})^2}{\hat{\theta}(1-\hat{\theta})}\Big(1 + \sum_{j=1}^{k} \hat{a}_j (\hat{\theta}^* - \hat{\theta})^j\Big), \tag{4.1}$$

where $\hat{\theta}^* = \hat{F}_n^*(t_0)$ is based on a bootstrap sample of size n. Noting that the right censored sample is a special case of doubly censored sample, from Proposition 2.1 of Bickel and Ren (1996) we know that the distribution of $\sqrt{n}(\hat{\theta}-\theta_0) = \sqrt{n}[\hat{F}(t_0) - F_0(t_0)]$ can be consistently estimated by that of $\sqrt{n}[\hat{F}_n^*(t_0) - \hat{F}_n(t_0)]$ for right censored data and doubly censored data. This bootstrap consistency also holds for partly interval-censored data as shown in Huang (1999). Since $B_n^{(k)}$ in (3.7) is a 'polynomial' in $\sqrt{n}(\hat{\theta}-\theta_0)$, we know that $B_n^{(k)*}$ in (4.1) provides consistent estimate for $\rho_{n,\alpha}^{(k)}$.

Let Exp(μ) represent the exponential distribution function with mean μ. In Table 1, 1000 right censored samples of size $n = 100$ with $V_i = \min\{X_i, Y_i\}, \delta_i = I\{X_i \le Y_i\}, 1 \le i \le n$, are taken from exponential distributions, and for each sample, 90% k-WELRCI, TGCI and BPCI for θ_0 with different t_0 are computed, where 400 bootstrap samples of size $n = 100$ are used for BPCI and for estimating $\rho_{n,\alpha}^{(k)}$ in (3.9) to construct k-WELRCI. Simulation coverages are included in Table 1, and the simulation standard deviation of the length of C.I. is given in the parenthesis next to the average length of C.I.. The same studies in Table 1 are repeated in Table 2 with $n = 50$.

TABLES

Table 1. 90% Confidence Interval for $\theta_0 = F_0(t_0)$ with Exponential Right Censored Data

Sample Size $n = 100$ with $X \sim$ Exp(1), $Y \sim$ Exp(3); **Percentage of δ:** $\delta = 1:\ 75.0\%$; $\delta = 0:\ 25.0\%$

t_0 & θ_0	Method	% Coverage	% Miss Left	% Miss Right	Average Length (S.D.)
$t_0 = 0.500$	1-WELRCI	89.8	5.3	4.9	0.16634 (0.00819)
$\theta_0 = 0.393$	2-WELRCI	90.1	5.2	4.7	0.16781 (0.00827)
	TGCI	90.0	5.4	4.6	0.16652 (0.00468)
	BPCI	89.8	6.4	3.8	0.16798 (0.00819)
$t_0 = 1.000$	1-WELRCI	90.1	5.6	4.3	0.17164 (0.00962)
$\theta_0 = 0.632$	2-WELRCI	90.6	5.2	4.2	0.17367 (0.00968)
	TGCI	91.2	5.2	3.6	0.17269 (0.00628)
	BPCI	88.9	8.1	3.0	0.17349 (0.00984)
$t_0 = 1.500$	1-WELRCI	89.1	6.0	4.9	0.15465 (0.01530)
$\theta_0 = 0.777$	2-WELRCI	90.4	5.1	4.5	0.15981 (0.01448)
	TGCI	89.8	5.6	4.6	0.15749 (0.01240)
	BPCI	87.6	9.9	2.5	0.15811 (0.01507)
$t_0 = 2.000$	1-WELRCI	84.3	11.0	4.7	0.13223 (0.04557)
$\theta_0 = 0.865$	2-WELRCI	88.6	7.3	4.1	0.14415 (0.04459)
	TGCI	88.0	7.8	4.2	0.13700 (0.01965)
	BPCI	82.5	15.5	2.0	0.13658 (0.02521)

Table 2. 90% Confidence Interval for $\theta_0 = F_0(t_0)$ with Exponential Right Censored Data

Sample Size $n = 50$ with $X \sim \text{Exp}(1)$, $Y \sim \text{Exp}(3)$; **Percentage of δ:** $\delta = 1: 74.8\%$; $\delta = 0: 25.2\%$

t_0 & θ_0	Method	% Coverage	% Miss Left	% Miss Right	Average Length (S.D.)
$t_0 = 0.500$	1-WELRCI	90.2	3.9	5.9	0.23146 (0.01438)
$\theta_0 = 0.393$	2-WELRCI	91.5	3.6	4.9	0.23567 (0.01439)
	TGCI	90.9	3.9	5.2	0.23192 (0.00993)
	BPCI	90.6	5.4	4.0	0.23544 (0.01446)
$t_0 = 1.000$	1-WELRCI	88.0	6.5	5.5	0.23904 (0.01801)
$\theta_0 = 0.632$	2-WELRCI	89.8	5.2	5.0	0.24524 (0.01751)
	TGCI	89.0	5.7	5.3	0.24138 (0.01309)
	BPCI	87.0	9.9	3.1	0.24389 (0.01824)
$t_0 = 1.500$	1-WELRCI	89.4	6.1	4.5	0.21267 (0.03234)
$\theta_0 = 0.777$	2-WELRCI	92.1	4.1	3.7	0.22948 (0.03068)
	TGCI	91.6	4.1	4.3	0.22039 (0.02447)
	BPCI	87.5	11.0	1.5	0.22239 (0.03407)
$t_0 = 2.000$	1-WELRCI	92.3	3.6	4.1	0.19973 (0.14166)
$\theta_0 = 0.865$	2-WELRCI	95.4	1.5	3.1	0.22387 (0.13986)
	TGCI	90.5	5.8	3.7	0.19158 (0.03878)
	BPCI	80.6	18.5	0.9	0.17454 (0.05931)

From Tables 1-2, we see that for right censored data, WELRCI and TGCI have similar performances, while BPCI performs poorly when θ_0 is near either 0 or 1. Also, we see that 1-WELRCI and 2-WELRCI perform similarly, though 2-WELRCI seems preferred for larger t_0. Our extensive simulation studies for other distributions have similar performance.

Moreover, it is worth noting that simulation results show that the likelihood-based confidence intervals, WELRCI or TGCI, usually have miscoverage on both sides more evenly than the bootstrap percentile method. This indicates that the likelihood-based method better catches the skewness of the distribution of the statistics of interest, and the weighted empirical likelihood method, with its evident generality in formulation and computation, preserves this appealing feature of the empirical likelihood method.

Concluding Remarks:

It should be noted that (3.10) does not necessarily give the exact coverage accuracy due to the techniques we used in our proofs, but it generally guarantees the likelihood-based confidence intervals to have at least 'the first order' accuracy for different types of censored data as long as the convergence rate of $G_{n,k}$ is $\sqrt{n}$. From the example of smoothing $\hat{F}_n$ based on the kernel density method, we know that with smoothing of $\hat{F}_n$, the Lipschitz condition may hold for $\tilde{\theta}$, in turn, we have the stronger theoretical coverage accuracy equation (3.12), which obviously indicates that the practical coverage accuracy of WELRCI boils down to the accuracy of the bootstrap method, or the accuracy of the estimation for the percentiles of $B_n^{(k)}$ or $\tilde{B}_n^{(k)}$ by any other methods. Thus, a more sophisticated bootstrap procedure, such as in Hall (1992) or Efron and Tibshirani (1993), might be applicable to obtain better coverage accuracy than $O(n^{-1/2})$. This will be studied in a separate forthcoming paper. Finally, we note that the use of smoothing can give shorter confidence intervals; see Ren (2005).

5. Appendix

Proof of (3.3). First, we notice that S_n can be expressed as

$$S_n = \Big\{\tau(\boldsymbol{p}) \;\Big|\; \boldsymbol{p} = (p_1, \cdots, p_m) \in E_c\Big\}, \tag{A.1}$$

where $E_c = \{\boldsymbol{p} \,|\, p_i \geq 0, \sum_{i=1}^m p_i = 1, \prod_{i=1}^m (p_i/\hat{p}_i)^{n\hat{p}_i} \geq c\}$ is a compact and convex set in $\mathbb{R}^m$, and $\tau(\boldsymbol{p}) = \sum_{i=1}^m p_i I\{W_i \leq t_0\}$ is a continuous function in $\boldsymbol{p}$. From Royden (1968, page 158-159), we know that S_n is a compact set in $\mathbb{R}$. Since convexity implies connectivity, from Royden (1968, page 152-153) we know that S_n is either an interval or a single point. Since S_n is compact, we know that S_n must be a closed interval $[X_L, X_U]$ with X_L and X_U given by (3.4). To show (3.3), we let $E_0 = \{\boldsymbol{p} \,|\, \tau(\boldsymbol{p}) = \theta_0,\, p_i \geq 0,\, \sum_{i=1}^m p_i = 1\}$, thus we have

$$r(\theta_0) = \sup\Big\{\prod_{i=1}^m (p_i/\hat{p}_i)^{n\hat{p}_i} \;\Big|\; \boldsymbol{p} \in E_0\Big\}. \tag{A.2}$$

Assume $r(\theta_0) \geq c$. Since τ is continuous and $\{\boldsymbol{p} \,|\, p_i \geq 0, \sum_{i=1}^m p_i = 1\}$ is a compact set, we know that E_0 is compact and E_0 is not empty. Thus, (A.2) and (3.4) imply $X_L \leq \theta_0 \leq X_U$.

Assume $X_L \leq \theta_0 \leq X_U$. Since τ is continuous with X_L and X_U as the lower and upper bound on compact set E_c, respectively, we know that from The Intermediate Value Theorem, there exists $\boldsymbol{q} \in E_c$ such that $\tau(\boldsymbol{q}) = \theta_0$. Thus, (A.2) implies $r(\theta_0) \geq c$. □

Proof of Theorem 1 (i). To get an expression for $r(\theta_0)$ in (3.2), using the Lagrange multiplier it is easy to obtain

$$\log r(\theta_0) = -n \sum_{i=1}^m \hat{p}_i \log[1 + \lambda_0(U_i - \theta_0)], \tag{A.3}$$

where $U_i = I\{W_i \leq t_0\}$, and λ_0 is the unique solution in $(-\frac{1}{1-\theta_0}, \frac{1}{\theta_0})$ of equation:

$$g(\lambda) \equiv \sum_{i=1}^m \frac{\hat{p}_i(U_i - \theta_0)}{1 + \lambda(U_i - \theta_0)} = 0. \tag{A.4}$$

To study the asymptotic behavior of λ_0, we notice that in (A.3)-(A.4), we have that $g(\lambda_0) = 0$ and $g(0) = \hat{\theta} - \theta_0$, which give

$$-(\hat{\theta} - \theta_0) = g(\lambda_0) - g(0) = g'(\xi)\lambda_0 = -\lambda_0 \sum_{i=1}^m \frac{\hat{p}_i(U_i - \theta_0)^2}{[1 + \xi(U_i - \theta_0)]^2}, \tag{A.5}$$

where $|\xi| \leq |\lambda_0|$. From $[1 + \xi(U_i - \theta_0)]^2 \leq (1 + |\lambda_0|)^2$, we have

$$|\hat{\theta} - \theta_0| = |\lambda_0| \sum_{i=1}^m \frac{\hat{p}_i(U_i - \theta_0)^2}{[1 + \xi(U_i - \theta_0)]^2} \geq |\lambda_0| \sum_{i=1}^n \frac{\hat{p}_i(U_i - \theta_0)^2}{(1 + |\lambda_0|)^2},$$

in turn, from $|\lambda_0| \leq \max\{\frac{1}{1-\theta_0}, \frac{1}{\theta_0}\}$ we have

$$|\lambda_0|\hat{\mu}_2 \leq |\hat{\theta} - \theta_0|(1 + |\lambda_0|)^2 \leq |\hat{\theta} - \theta_0| \Big(1 + \max\Big\{\frac{1}{1-\theta_0}, \frac{1}{\theta_0}\Big\}\Big)^2,$$

where for $\hat{\theta} = \hat{F}_n(t_0) = \sum_{i=1}^m \hat{p}_i U_i$,

$$\hat{\mu}_k = \textstyle\sum_{i=1}^m \hat{p}_i (U_i - \theta_0)^k \quad \text{and} \quad \tilde{\mu}_k = \sum_{i=1}^m \hat{p}_i (U_i - \hat{\theta})^k, \tag{A.6}$$

and straightforward calculation gives

$$\begin{aligned} &\tilde{\mu}_2 = \hat{\theta}(1 - \hat{\theta}), \quad \tilde{\mu}_3 = \hat{\theta}(1 - \hat{\theta})[(1 - \hat{\theta})^2 - \hat{\theta}^2], \quad \tilde{\mu}_4 = \hat{\theta}(1 - \hat{\theta})[(1 - \hat{\theta})^3 + \hat{\theta}^3], \\ &\tilde{\mu}_5 = \hat{\theta}(1 - \hat{\theta})[(1 - \hat{\theta})^4 - \hat{\theta}^4], \quad \tilde{\mu}_6 = \hat{\theta}(1 - \hat{\theta})[(1 - \hat{\theta})^5 + \hat{\theta}^5], \end{aligned} \tag{A.7}$$

and

$$
\begin{aligned}
&\hat\mu_2 = \tilde\mu_2 + (\hat\theta-\theta_0)^2, \qquad \hat\mu_3 = \tilde\mu_3 + 3(\hat\theta-\theta_0)\tilde\mu_2 + (\hat\theta-\theta_0)^3,\\
&\hat\mu_4 = \tilde\mu_4 + 4(\hat\theta-\theta_0)\tilde\mu_3 + 6(\hat\theta-\theta_0)^2\tilde\mu_2 + (\hat\theta-\theta_0)^4,\\
&\hat\mu_5 = \tilde\mu_5 + 5(\hat\theta-\theta_0)\tilde\mu_4 + 10(\hat\theta-\theta_0)^2\tilde\mu_3 + 10(\hat\theta-\theta_0)^3\tilde\mu_2 + (\hat\theta-\theta_0)^5,\\
&\hat\mu_6 = \tilde\mu_6 + 6(\hat\theta-\theta_0)\tilde\mu_5 + 15(\hat\theta-\theta_0)^2\tilde\mu_4 + 20(\hat\theta-\theta_0)^3\tilde\mu_3 + 15(\hat\theta-\theta_0)^4\tilde\mu_2 + (\hat\theta-\theta_0)^6.
\end{aligned}
\tag{A.8}
$$

Thus, from

$$0 < \min\{\theta_0^2, (1-\theta_0)^2\} \le \hat\mu_2 \le 1, \tag{A.9}$$

we know that (AS2) and Theorem 4.2.2 of Chung (1974) imply that with probability 1,

$$|\lambda_0| \le M_\lambda^{-1}|\hat\theta-\theta_0| \le \tfrac{1}{2} \tag{A.10}$$

all but finitely often, where $M_\lambda = \min\{\theta_0^2,(1-\theta_0)^2\}(1+\max\{\frac{1}{1-\theta_0},\frac{1}{\theta_0}\})^{-2}$. The rest of the proof is established under (A.10), which along with (AS2) gives

$$\lambda_0 \xrightarrow{a.s.} 0, \quad \text{as} \quad n\to\infty. \tag{A.11}$$

To avoid overly messy algebra, we will establish (3.6)-(3.7) for $k=1$, while the method can easily be used for the case of $k=4$.

Let $h=g^{-1}$, then $g(\lambda_0)=0$ and $g(0)=(\hat\theta-\theta_0)$ imply $h(0)=\lambda_0$ and $h(\hat\theta-\theta_0)=0$, respectively. Moreover, we have

$$
\begin{aligned}
&h'(\hat\theta-\theta_0) = \frac{1}{g'(0)} = -\frac{1}{\hat\mu_2}, \quad h''(\hat\theta-\theta_0) = -\frac{g''(0)}{[g'(0)]^3} = \frac{2\hat\mu_3}{\hat\mu_2^3},\\
&h'''(y) = \frac{3[g''(x)]^2 - g'(x)g'''(x)}{[g'(x)]^5}, \quad \text{where} \quad x = h(y)
\end{aligned}
\tag{A.12}
$$

and from Taylor's expansion,

$$
\begin{aligned}
\lambda_0 = h(0) &= h(\hat\theta-\theta_0) - h'(\hat\theta-\theta_0)(\hat\theta-\theta_0) + \tfrac{1}{2}h''(\hat\theta-\theta_0)(\hat\theta-\theta_0)^2 - \tfrac{1}{6}h'''(\xi)(\hat\theta-\theta_0)^3\\
&= \frac{(\hat\theta-\theta_0)}{\hat\mu_2} + \frac{\hat\mu_3}{\hat\mu_2^3}(\hat\theta-\theta_0)^2 + R_h,
\end{aligned}
\tag{A.13}
$$

where $R_h = -\frac{1}{6}h'''(\xi)(\hat{\theta} - \theta_0)^3$ with $|\xi| \le |\hat{\theta} - \theta_0|$ satisfying $\eta = h(\xi)$ and $|\eta| \le |\lambda_0|$. Since

$$\frac{\hat{\mu}_2}{(1+|\lambda_0|)^2} \le |g'(\eta)| = \sum_{i=1}^m \frac{\hat{p}_i(U_i - \theta_0)^2}{[1+\eta(U_i - \theta_0)]^2} \le \frac{1}{(1-|\lambda_0|)^2},$$

$$|g''(\eta)| = 2\Big|\sum_{i=1}^m \frac{\hat{p}_i(U_i - \theta_0)^3}{[1+\eta(U_i - \theta_0)]^3}\Big| \le \frac{2}{(1-|\lambda_0|)^3},$$

$$|g'''(\eta) = 6\sum_{i=1}^m \frac{\hat{p}_i(U_i - \theta_0)^4}{[1+\eta(U_i - \theta_0)]^4} \le \frac{6}{(1-|\lambda_0|)^4},$$

then from (A.9)-(A.10) there exists a constant M_h such that

$$|R_h| = \frac{1}{6}|\hat{\theta} - \theta_0|^3 \Big| \frac{3[g''(\eta)]^2 - g'(\eta)g'''(\eta)}{[g'(\eta)]^5} \Big| \le M_h|\hat{\theta} - \theta_0|^3. \tag{A.14}$$

From Taylor's expansion, in (A.3) we have

$$\begin{aligned} -2\log r(\theta_0) &= 2n\sum_{i=1}^m \hat{p}_i \log(1+\lambda_0(U_i - \theta_0)) = 2n\sum_{i=1}^m \hat{p}_i \Big\{\lambda_0(U_i - \theta_0) \\ &\quad - \frac{1}{2}[\lambda_0(U_i - \theta_0)]^2 + \frac{1}{3}[\lambda_0(U_i - \theta_0)]^3 - \frac{1}{4}(1+\zeta_i)^{-4}[\lambda_0(U_i - \theta_0)]^4\Big\} \\ &= 2n\Big\{\lambda_0(\hat{\theta} - \theta_0) - \frac{1}{2}\lambda_0^2\hat{\mu}_2 + \frac{1}{3}\lambda_0^3\hat{\mu}_3\Big\} + R_1, \end{aligned} \tag{A.15}$$

where $|\zeta_i| \le |\lambda_0(U_i - \theta_0)| \le |\lambda_0|$ and $R_1 = -\frac{1}{2}n\sum_{i=1}^m \hat{p}_i(1+\zeta_i)^{-4}[\lambda_0(U_i - \theta_0)]^4$. Easily, from (A.10) we know that

$$|R_1| \le \frac{1}{2}n(1-|\lambda_0|)^{-4}|\hat{\theta} - \theta_0|^4 M_\lambda^{-4} \le 8n|\hat{\theta} - \theta_0|^4 M_\lambda^{-4}. \tag{A.16}$$

From (A.13), we have in (A.15)

$$\begin{aligned} -2\log r(\theta_0) - R_1 &= 2n\Big\{\lambda_0(\hat{\theta} - \theta_0) - \frac{1}{2}\lambda_0^2\hat{\mu}_2 + \frac{1}{3}\lambda_0^3\hat{\mu}_3\Big\} \\ &= 2n\Big\{(\hat{\theta} - \theta_0)\Big(\frac{(\hat{\theta} - \theta_0)}{\hat{\mu}_2} + \frac{\hat{\mu}_3}{\hat{\mu}_2^3}(\hat{\theta} - \theta_0)^2 + R_h\Big) \\ &\quad - \frac{1}{2}\hat{\mu}_2\Big(\frac{(\hat{\theta} - \theta_0)}{\hat{\mu}_2} + \frac{\hat{\mu}_3}{\hat{\mu}_2^3}(\hat{\theta} - \theta_0)^2 + R_h\Big)^2 \\ &\quad + \frac{1}{3}\hat{\mu}_3\Big(\frac{(\hat{\theta} - \theta_0)}{\hat{\mu}_2} + \frac{\hat{\mu}_3}{\hat{\mu}_2^3}(\hat{\theta} - \theta_0)^2 + R_h\Big)^3\Big\} \\ &= 2n\Big\{\frac{(\hat{\theta} - \theta_0)^2}{\hat{\mu}_2} + \frac{\hat{\mu}_3}{\hat{\mu}_2^3}(\hat{\theta} - \theta_0)^3 + (\hat{\theta} - \theta_0)R_h \end{aligned}$$

$$
\begin{aligned}
&- \tfrac{1}{2}\hat{\mu}_2\Big(\frac{(\hat{\theta}-\theta_0)^2}{\hat{\mu}_2^2} + \frac{\hat{\mu}_3^2}{\hat{\mu}_2^6}(\hat{\theta}-\theta_0)^4 + R_h^2 + 2\frac{\hat{\mu}_3}{\hat{\mu}_2^4}(\hat{\theta}-\theta_0)^3 \\
&\qquad + 2R_h\Big(\frac{(\hat{\theta}-\theta_0)}{\hat{\mu}_2} + \frac{\hat{\mu}_3}{\hat{\mu}_2^3}(\hat{\theta}-\theta_0)^2\Big)\Big) \\
&+ \tfrac{1}{3}\hat{\mu}_3\frac{(\hat{\theta}-\theta_0)^3}{\hat{\mu}_2^3} \\
&+ \tfrac{1}{3}\hat{\mu}_3\frac{(\hat{\theta}-\theta_0)}{\hat{\mu}_2}\Big(\frac{\hat{\mu}_3^2}{\hat{\mu}_2^6}(\hat{\theta}-\theta_0)^4 + R_h^2 + 2\frac{\hat{\mu}_3}{\hat{\mu}_2^4}(\hat{\theta}-\theta_0)^3 \\
&\qquad + 2R_h\Big(\frac{(\hat{\theta}-\theta_0)}{\hat{\mu}_2} + \frac{\hat{\mu}_3}{\hat{\mu}_2^3}(\hat{\theta}-\theta_0)^2\Big)\Big) \\
&+ \tfrac{1}{3}\hat{\mu}_3\Big(\frac{\hat{\mu}_3}{\hat{\mu}_2^3}(\hat{\theta}-\theta_0)^2 + R_h\Big)\Big(\frac{(\hat{\theta}-\theta_0)^2}{\hat{\mu}_2^2} + \frac{\hat{\mu}_3^2}{\hat{\mu}_2^6}(\hat{\theta}-\theta_0)^4 + R_h^2 \\
&\qquad + 2\frac{\hat{\mu}_3}{\hat{\mu}_2^4}(\hat{\theta}-\theta_0)^3 + 2R_h\Big(\frac{(\hat{\theta}-\theta_0)}{\hat{\mu}_2} + \frac{\hat{\mu}_3}{\hat{\mu}_2^3}(\hat{\theta}-\theta_0)^2\Big)\Big)\Big\} \\
&= 2n\Big\{\frac{(\hat{\theta}-\theta_0)^2}{\hat{\mu}_2} + \frac{\hat{\mu}_3}{\hat{\mu}_2^3}(\hat{\theta}-\theta_0)^3 - \frac{1}{2}\frac{(\hat{\theta}-\theta_0)^2}{\hat{\mu}_2} - \frac{\hat{\mu}_3}{\hat{\mu}_2^3}(\hat{\theta}-\theta_0)^3 + \tfrac{1}{3}\hat{\mu}_3\frac{(\hat{\theta}-\theta_0)^3}{\hat{\mu}_2^3}\Big\} + R_2 \\
&= \frac{n(\hat{\theta}-\theta_0)^2}{\hat{\mu}_2}\Big(1 + \frac{2\hat{\mu}_3}{3\hat{\mu}_2^2}(\hat{\theta}-\theta_0)\Big) + R_2, \qquad \text{(A.17)}
\end{aligned}
$$

where

$$
\begin{aligned}
R_2 = 2n\Big\{&(\hat{\theta}-\theta_0)R_h - \tfrac{1}{2}\hat{\mu}_2\Big(\frac{\hat{\mu}_3^2}{\hat{\mu}_2^6}(\hat{\theta}-\theta_0)^4 + R_h^2 + 2R_h\Big(\frac{(\hat{\theta}-\theta_0)}{\hat{\mu}_2} + \frac{\hat{\mu}_3}{\hat{\mu}_2^3}(\hat{\theta}-\theta_0)^2\Big)\Big) \\
&+ \tfrac{1}{3}\hat{\mu}_3\frac{(\hat{\theta}-\theta_0)}{\hat{\mu}_2}\Big(\frac{\hat{\mu}_3^2}{\hat{\mu}_2^6}(\hat{\theta}-\theta_0)^4 + R_h^2 + 2\frac{\hat{\mu}_3}{\hat{\mu}_2^4}(\hat{\theta}-\theta_0)^3 + 2R_h\Big(\frac{(\hat{\theta}-\theta_0)}{\hat{\mu}_2} + \frac{\hat{\mu}_3}{\hat{\mu}_2^3}(\hat{\theta}-\theta_0)^2\Big)\Big) \\
&+ \tfrac{1}{3}\hat{\mu}_3\Big(\frac{\hat{\mu}_3}{\hat{\mu}_2^3}(\hat{\theta}-\theta_0)^2 + R_h\Big)\Big(\frac{(\hat{\theta}-\theta_0)^2}{\hat{\mu}_2^2} + \frac{\hat{\mu}_3^2}{\hat{\mu}_2^6}(\hat{\theta}-\theta_0)^4 + R_h^2 + 2\frac{\hat{\mu}_3}{\hat{\mu}_2^4}(\hat{\theta}-\theta_0)^3 \\
&\qquad + 2R_h\Big(\frac{(\hat{\theta}-\theta_0)}{\hat{\mu}_2} + \frac{\hat{\mu}_3}{\hat{\mu}_2^3}(\hat{\theta}-\theta_0)^2\Big)\Big)\Big\}.
\end{aligned}
$$

Since (A.6) implies $|\hat{\mu}_3| \le 1$, thus (A.9) and (A.14) imply

$$|R_2| \le n|\hat{\theta}-\theta_0|^4 M_{R_2}, \qquad \text{where } M_{R_2} \text{ is a constant.} \tag{A.18}$$

Note that Taylor's expansion and (A.8) give

$$\frac{1}{\hat{\mu}_2} = \frac{1}{\tilde{\mu}_2}\Big\{1 - \frac{(\hat{\theta}-\theta_0)^2}{\tilde{\mu}_2(1+\eta)^2}\Big\}, \quad \text{where} \quad |\eta| \le \tilde{\mu}_2^{-1}|\hat{\theta}-\theta_0|^2. \tag{A.19}$$

Thus, to express $\hat{\mu}_j$'s of the leading term of (A.17) in terms of $\tilde{\mu}_j$'s, (A.8) gives

$$
\begin{aligned}
-2\log r(\theta_0) - R_1 - R_2 &= \frac{n(\hat\theta-\theta_0)^2}{\hat\mu_2}\Big(1+\frac{2\hat\mu_3}{3\hat\mu_2^2}(\hat\theta-\theta_0)\Big)\\
&= \frac{n(\hat\theta-\theta_0)^2}{\tilde\mu_2}\Big\{1-\frac{(\hat\theta-\theta_0)^2}{\tilde\mu_2(1+\eta)^2}\Big\}\\
&\qquad \times\Big\{1+\frac{2}{3\tilde\mu_2^2}(\hat\theta-\theta_0)[\tilde\mu_3+3(\hat\theta-\theta_0)\tilde\mu_2+(\hat\theta-\theta_0)^3]\Big(1-\frac{(\hat\theta-\theta_0)^2}{\tilde\mu_2(1+\eta)^2}\Big)^2\Big\}\\
&= \frac{n(\hat\theta-\theta_0)^2}{\tilde\mu_2}\Big\{1-\frac{(\hat\theta-\theta_0)^2}{\tilde\mu_2(1+\eta)^2}\Big\}\Big\{1+\frac{2\tilde\mu_3}{3\tilde\mu_2^2}(\hat\theta-\theta_0)+\frac{2}{3\tilde\mu_2^2}[3(\hat\theta-\theta_0)^2\tilde\mu_2+(\hat\theta-\theta_0)^4]\\
&\qquad +\frac{2}{3\tilde\mu_2^2}(\hat\theta-\theta_0)[\tilde\mu_3+3(\hat\theta-\theta_0)\tilde\mu_2+(\hat\theta-\theta_0)^3]\Big(-\frac{2(\hat\theta-\theta_0)^2}{\tilde\mu_2(1+\eta)^2}+\frac{(\hat\theta-\theta_0)^4}{\tilde\mu_2^2(1+\eta)^4}\Big)\Big\}\\
&= \frac{n(\hat\theta-\theta_0)^2}{\tilde\mu_2}\Big\{1+\frac{2\tilde\mu_3}{3\tilde\mu_2^2}(\hat\theta-\theta_0)+\frac{2}{3\tilde\mu_2^2}[3(\hat\theta-\theta_0)^2\tilde\mu_2+(\hat\theta-\theta_0)^4]\\
&\qquad +\frac{2}{3\tilde\mu_2^2}(\hat\theta-\theta_0)[\tilde\mu_3+3(\hat\theta-\theta_0)\tilde\mu_2+(\hat\theta-\theta_0)^3]\Big(-\frac{2(\hat\theta-\theta_0)^2}{\tilde\mu_2(1+\eta)^2}+\frac{(\hat\theta-\theta_0)^4}{\tilde\mu_2^2(1+\eta)^4}\Big)\\
&\qquad -\frac{(\hat\theta-\theta_0)^2}{\tilde\mu_2(1+\eta)^2}\Big\{1+\frac{2\tilde\mu_3}{3\tilde\mu_2^2}(\hat\theta-\theta_0)+\frac{2}{3\tilde\mu_2^2}[3(\hat\theta-\theta_0)^2\tilde\mu_2+(\hat\theta-\theta_0)^4] \qquad \text{(A.20)}\\
&\qquad +\frac{2}{3\tilde\mu_2^2}(\hat\theta-\theta_0)[\tilde\mu_3+3(\hat\theta-\theta_0)\tilde\mu_2+(\hat\theta-\theta_0)^3]\Big(-\frac{2(\hat\theta-\theta_0)^2}{\tilde\mu_2(1+\eta)^2}+\frac{(\hat\theta-\theta_0)^4}{\tilde\mu_2^2(1+\eta)^4}\Big)\Big\}\Big\}\\
&= \frac{n(\hat\theta-\theta_0)^2}{\tilde\mu_2}\Big(1+\frac{2\tilde\mu_3}{3\tilde\mu_2^2}(\hat\theta-\theta_0)\Big)+R_3,
\end{aligned}
$$

where

$$
\begin{aligned}
R_3 = \frac{n(\hat\theta-\theta_0)^2}{\tilde\mu_2}\Big\{&\frac{2}{3\tilde\mu_2^2}[3(\hat\theta-\theta_0)^2\tilde\mu_2+(\hat\theta-\theta_0)^4]\\
&+\frac{2}{3\tilde\mu_2^2}(\hat\theta-\theta_0)[\tilde\mu_3+3(\hat\theta-\theta_0)\tilde\mu_2+(\hat\theta-\theta_0)^3]\Big(-\frac{2(\hat\theta-\theta_0)^2}{\tilde\mu_2(1+\eta)^2}+\frac{(\hat\theta-\theta_0)^4}{\tilde\mu_2^2(1+\eta)^4}\Big)\\
&-\frac{(\hat\theta-\theta_0)^2}{\tilde\mu_2(1+\eta)^2}\Big\{1+\frac{2\tilde\mu_3}{3\tilde\mu_2^2}(\hat\theta-\theta_0)+\frac{2}{3\tilde\mu_2^2}[3(\hat\theta-\theta_0)^2\tilde\mu_2+(\hat\theta-\theta_0)^4]\\
&+\frac{2}{3\tilde\mu_2^2}(\hat\theta-\theta_0)[\tilde\mu_3+3(\hat\theta-\theta_0)\tilde\mu_2+(\hat\theta-\theta_0)^3]\Big(-\frac{2(\hat\theta-\theta_0)^2}{\tilde\mu_2(1+\eta)^2}+\frac{(\hat\theta-\theta_0)^4}{\tilde\mu_2^2(1+\eta)^4}\Big)\Big\}\Big\}.
\end{aligned}
$$

Note that (A.10) implies $0 < (\theta_0 - \frac{1}{2}M_\lambda) \le \hat\theta \le (\theta_0 + \frac{1}{2}M_\lambda) < 1$. Thus, (A.7) gives

$$\tilde\mu_2 = \hat\theta(1-\hat\theta) \ge \Big(\theta_0 - \frac{1}{2}M_\lambda\Big)\Big(1-\theta_0-\frac{1}{2}M_\lambda\Big) > 0 \qquad \text{(A.21)}$$

and $|\tilde{\mu}_3| \le 1$. Hence, from $|\eta| \le \tilde{\mu}_2^{-1}|\hat{\theta}-\theta_0|^2 \le \frac{1}{4}M_\lambda^2[(\theta_0-\frac{1}{2}M_\lambda)(1-\theta_0-\frac{1}{2}M_\lambda)]^{-1} < 1$, we know that there exists a constant M_{R_3} such that

$$|R_3| \le n|\hat{\theta}-\theta_0|^4 M_{R_3}. \tag{A.22}$$

Note that from (A.7), we have $\hat{a}_1 = (2\tilde{\mu}_3)/(3\tilde{\mu}_2^2) = 2[(1-\hat{\theta})^2-\hat{\theta}^2]/[3\hat{\theta}(1-\hat{\theta})]$. Since (A.16), (A.18) and (A.22) are established under (A.10), hence (A.20) implies (3.6)-(3.7) for case $k=1$, where under (A.10) we have that for some constant $M_{r,k} \ge 1$,

$$|r_{n,k}| \le M_{r,k}. \qquad \square \tag{A.23}$$

Proof of Theorem 1 (ii). Noting that (AS2) implies $0 < \hat{\theta} < 1$ almost surly, from (3.3) and (3.6) we have that for $\tilde{c}_n = -2\log c_n$,

$$\begin{aligned} P\{X_L \le \theta_0 \le X_U\} &= P\{X_L \le \theta_0 \le X_U, 0<\hat{\theta}<1\} \\ &= P\{-2\log r(\theta_0) \le -2\log c_n, 0<\hat{\theta}<1\} \\ &= [P\{B_n^{(k)} + n(\hat{\theta}-\theta_0)^{k+3} r_{n,k} \le \tilde{c}_n\} - P\{B_n^{(k)} \le \tilde{c}_n\}] + P\{B_n^{(k)} \le -2\log c_n\}. \end{aligned} \tag{A.24}$$

If we let $\hat{U} = \sqrt{n}(\hat{\theta}-\theta_0)[\hat{\theta}(1-\hat{\theta})]^{-1/2}$, then straightforward algebra gives

$$B_n^{(k)} = \hat{U}^2 + q_k(\hat{U},\hat{\theta}), \quad \text{for } k=1,2,3,4 \tag{A.25}$$

with $q_k(\hat{U},\hat{\theta}) = \hat{U}^2\sum_{i=1}^k \hat{a}_i(\hat{\theta}-\theta_0)^i$. Thus, in (A.24) we have

$$\begin{aligned} &|P\{B_n^{(k)} + n(\hat{\theta}-\theta_0)^{k+3} r_{n,k} \le \tilde{c}_n\} - P\{B_n^{(k)} \le \tilde{c}_n\}| \\ &\quad = |P\{B_n^{(k)} + \hat{U}^2(\hat{\theta}-\theta_0)^{k+1}\tilde{\mu}_2 r_{n,k} \le \tilde{c}_n\} - P\{B_n^{(k)} \le \tilde{c}_n\}| \\ &\quad = |P\{\hat{U}^2 + q_k(\hat{U},\hat{\theta}) + \hat{U}^2(\hat{\theta}-\theta_0)^{k+1}\tilde{\mu}_2 r_{n,k} \le \tilde{c}_n\} - P\{\hat{U}^2 + q_k(\hat{U},\hat{\theta}) \le \tilde{c}_n\}| \\ &\quad \le P\{\tilde{c}_n - |\hat{\theta}-\theta_0|^{k+1}|\bar{r}_{n,k}|\hat{U}^2 \le \hat{U}^2 + q_k(\hat{U},\hat{\theta}) \le \tilde{c}_n + |\hat{\theta}-\theta_0|^{k+1}|\bar{r}_{n,k}|\hat{U}^2\}, \end{aligned} \tag{A.26}$$

where $\bar{r}_{n,k} = \tilde{\mu}_2\, r_{n,k}$. Note that for $M_{r,k} \ge 1$ in (3.6) and M_λ in (A.10), Theorem 4.2.2 of Chung (1974) and (AS2) imply that with probability 1,

$$|\hat{\theta}-\theta_0| \le \frac{1}{2}M_\lambda M_{r,k}^{-1}\Big(\theta_0 - \frac{1}{2}M_\lambda\Big)\Big(1-\theta_0-\frac{1}{2}M_\lambda\Big) \tag{A.27}$$

all but finitely often. Thus, under (A.27) and (A.21) we can show

$$\hat{U}^2 + q_k(\hat{U},\hat{\theta}) \ge \hat{U}^2 A_\lambda, \quad k=0,1,\ldots,4, \tag{A.28}$$

where $A_\lambda > 0$ is a constant and $q_0(\hat{U}, \hat{\theta}) = 0$. Since (A.27) implies (A.10), thus from $\tilde{\mu}_2 \le 1$ and (A.23) we know that $\hat{U}^2 + q_k(\hat{U}, \hat{\theta}) \le \tilde{c}_n + |\hat{\theta} - \theta_0|^{k+1} |\bar{r}_{n,k}| \hat{U}^2$ implies

$$\hat{U}^2 A_\lambda \le \hat{U}^2 + q_k(\hat{U}, \hat{\theta}) \le \tilde{c}_n + |\hat{\theta} - \theta_0|^{k+1} \hat{U}^2 M_{r,k}, \tag{A.29}$$

in turn, from (A.27) we have

$$\Big(A_\lambda - \tfrac{1}{2} M_\lambda\Big) \hat{U}^2 = B_\lambda^2 \hat{U}^2 \le \tilde{c}_n, \tag{A.30}$$

where $B_\lambda = \sqrt{1 - \frac{53}{60} M_\lambda}$. Since $\hat{U}^2 \le B_\lambda^{-2} \tilde{c}_n$ implies

$$|\hat{U}| \le \sqrt{\tilde{c}_n}/B_\lambda \qquad \Rightarrow \qquad |\sqrt{n}(\hat{\theta} - \theta_0)| \le \sqrt{\tilde{c}_n}/B_\lambda, \tag{A.31}$$

then (3.8) follows from (A.24), (A.26), (A.23), (A.29)-(A.31), (AS3), $\tilde{c}_n = O(1)$ and

$$\begin{aligned}
&|P\{B_n^{(k)} + n(\hat{\theta} - \theta_0)^{k+3} r_{n,k} \le \tilde{c}_n\} - P\{B_n^{(k)} \le \tilde{c}_n\}| \\
&\quad \le P\{\tilde{c}_n - B_\lambda^{-2} \tilde{c}_n M_{r,k} |\hat{\theta} - \theta_0|^{k+1} \le \hat{U}^2 + q_k(\hat{U}, \hat{\theta}) \le \tilde{c}_n + B_\lambda^{-2} \tilde{c}_n M_{r,k} |\hat{\theta} - \theta_0|^{k+1}\} \\
&\quad \le P\{\tilde{c}_n - B_\lambda^{-2} \tilde{c}_n M_{r,k} (\sqrt{\tilde{c}_n}/B_\lambda)^{k+1} n^{-(k+1)/2} \le \hat{U}^2 + q_k(\hat{U}, \hat{\theta}) \\
&\qquad\qquad \le \tilde{c}_n + B_\lambda^{-2} \tilde{c}_n M_{r,k} (\sqrt{\tilde{c}_n}/B_\lambda)^{k+1} n^{-(k+1)/2}\} \\
&\quad = G_{n,k}(\tilde{c}_n + B_\lambda^{-2} \tilde{c}_n M_{r,k} (\sqrt{\tilde{c}_n}/B_\lambda)^{k+1} n^{-(k+1)/2}) \\
&\qquad - G_{n,k}(\tilde{c}_n - B_\lambda^{-2} \tilde{c}_n M_{r,k} (\sqrt{\tilde{c}_n}/B_\lambda)^{k+1} n^{-(k+1)/2}) \\
&\quad \le 2\|G_{n,k} - G_0\| + 2\|G_0'\| B_\lambda^{-2} \tilde{c}_n M_{r,k} (\sqrt{\tilde{c}_n}/B_\lambda)^{k+1} n^{-(k+1)/2}. \qquad \square
\end{aligned}$$

Smoothing by the Kernel Density Method. Here we study the assumptions required for the theoretical coverage accuracy equation (3.12) in this example. Let $h_n > 0$ satisfy

$$h_n \to 0, \ n h_n \to \infty, \ \sqrt{n}\, h_n = O(1), \quad \text{as } n \to \infty. \tag{A.32}$$

For the NPMLE $\hat{F}_n$ with censored data or complete i.i.d. sample, the kernel density estimator for the bounded density f_0 of F_0 is given by

$$\hat{f}_n(x) = \frac{1}{h_n} \sum_{i=1}^m \hat{p}_i K\Big(\frac{x - W_i}{h_n}\Big) = \frac{1}{h_n} \int_{-\infty}^{\infty} K\Big(\frac{x - t}{h_n}\Big)\, d\hat{F}_n(t), \tag{A.33}$$

where K is a density function. Thus, we have

$$\int_{-\infty}^{\infty} \hat{f}_n(x)\, dx = \frac{1}{h_n} \sum_{i=1}^m \hat{p}_i \int_{-\infty}^{\infty} K\Big(\frac{x - W_i}{h_n}\Big)\, dx = \frac{1}{h_n} \sum_{i=1}^m \hat{p}_i \int_{-\infty}^{\infty} K(y) h_n\, dy = 1,$$

and the smoothed versions of $\hat{F}_n$ and $F_p(x) = \sum_{i=1}^m p_i I\{W_i \leq x\}$ in (3.11) are given by

$$\tilde{F}_n(t) = \int_{-\infty}^t \hat{f}_n(x)\, dx = \int_{-\infty}^\infty \hat{F}_n(t - h_n u)K(u)\, du, \tag{A.34}$$

and

$$\tilde{F}_p(t) = \frac{1}{h_n}\int_{-\infty}^t \int_{-\infty}^\infty K\Big(\frac{x-u}{h_n}\Big)\, dF_p(u)\, dx = \sum_{i=1}^m p_i F_K\Big(\frac{t - W_i}{h_n}\Big) \tag{A.35}$$

respectively, where F_K is the d.f. for p.d.f. K. In turn, we have

$$\begin{aligned} &\tilde{\theta} = \tilde{F}_n(t_0) = \textstyle\int_{-\infty}^\infty \hat{F}_n(t_0 - h_n u)K(u)du = \sum_{i=1}^m \hat{p}_i U_i, \\ &\tau(p) \equiv \tilde{F}_p(t_0) = \textstyle\sum_{i=1}^m p_i U_i, \end{aligned} \tag{A.36}$$

where $\boldsymbol{p} = (p_1, \ldots, p_m)$, and here we have $U_i = F_K((t_0 - W_i)/h_n)$ satisfying $0 \leq U_i \leq 1$.

In this study, we consider kernel $K(x) = \frac{1}{2}I\{|x| \leq 1\}$. Thus, F_K is strictly increasing on interval $(-1, 1)$, and $F_K(x) = 0$ if $x \leq -1$; 1, if $x \geq 1$. This gives

$$\begin{aligned} U_{(m)} &= F_K((t_0 - W_1)/h_n) = 1 \quad \Leftrightarrow \quad W_1 \leq t_0 - h_n, \\ U_{(1)} &= F_K((t_0 - W_m)/h_n) = 0 \quad \Leftrightarrow \quad W_m \geq t_0 + h_n. \end{aligned} \tag{A.37}$$

Assuming the support of W_i is $(0, \infty)$, we know that $W_1 \xrightarrow{a.s.} 0$ and $W_m \xrightarrow{a.s.} \infty$, as $n \to \infty$, in turn, we have that $U_{(1)} = 0$ and $U_{(m)} = 1$ all but finitely often with probability 1. Also, note that (A.36), the Dominated Convergence Theorem (DCT) and $\|\hat{F}_n - F_0\| \xrightarrow{a.s.} 0$ (see Remark 1) give

$$\begin{aligned} \tilde{\theta} &= \int_{-\infty}^\infty [\hat{F}_n(t_0 - h_n u) - F_0(t_0 - h_n u)]K(u)\, du + \int_{-\infty}^\infty F_0(t_0 - h_n u)K(u)\, du \\ &= o_{a.s.}(1) + \int_{-\infty}^\infty F_0(t_0)K(u)du = o_{a.s.}(1) + \theta_0, \end{aligned} \tag{A.38}$$

where $o_{a.s.}(1)$ converges to 0 with probability 1, and similarly by (A.36), here we have

$$\begin{aligned} \tilde{\mu}_2 &= \sum_{i=1}^m \hat{p}_i (U_i - \tilde{\theta})^2 = \sum_{i=1}^m \hat{p}_i U_i^2 - \tilde{\theta}^2 \\ &= \int_{-\infty}^\infty \int_{-\infty}^\infty \hat{F}_n((t_0 - h_n u) \wedge (t_0 - h_n v))K(u)K(v)\, du\, dv - \tilde{\theta}^2 \\ &= o_{a.s.}(1) + \int_{-\infty}^\infty \int_{-\infty}^\infty F_0(t_0)K(u)K(v)\, du\, dv - \theta_0^2 = o_{a.s.}(1) + \theta_0(1 - \theta_0). \end{aligned} \tag{A.39}$$

Hence, we know that (AS2) holds for $\tilde{F}_n$.

Moreover, we have that from (A.32) and (A.36),

$$\begin{aligned}\sqrt{n}(\tilde{\theta}-\theta_0) &= \sqrt{n}\Big(\int_{-\infty}^{\infty}\hat{F}_n(t_0-h_nu)K(u)\,du-\int_{-\infty}^{\infty}F_0(t_0)K(u)\,du\Big)\\ &= \sqrt{n}\int_{-\infty}^{\infty}[\hat{F}_n(t_0-h_nu)-F_0(t_0-h_nu)]K(u)\,du\\ &\qquad\qquad + \sqrt{n}\int_{-\infty}^{\infty}[F_0(t_0-h_nu)-F_0(t_0)]K(u)\,du\\ &= O_p(1)+\sqrt{n}\int_{-\infty}^{\infty}f_0(\xi)h_nuK(u)\,du = O_p(1), \end{aligned} \tag{A.40}$$

where ξ is between (t_0-h_nu) and t_0, because $\sqrt{n}(\hat{F}_n-F_0)$ converges to a centered Gaussian process (see Remark 1). Thus, (AS1) holds.

For this example, (AS3) may hold, but the proof in general case is very difficult. Here we show that it holds for the complete i.i.d. sample: $X_1, X_2, \ldots, X_n$. Without loss of the generality, we assume that X_i's are all distinct. First, we study the d.f. F_Y of

$$Y = \sqrt{n}(\tilde{\theta}-\theta_0), \tag{A.41}$$

then we establish (AS3) with a remainder term converging to 0 in exponential rate. It should be noted that this, along with a proof similar to that of Theorem 1 (ii), gives (3.12) for case $k=0$, and the general case of (3.12) can be established similarly.

Note that from (A.36) and (A.41), here we have

$$Y = \sqrt{n}\Big(n^{-1}\sum_{i=1}^{n}F_K\big(\frac{t_0-X_i}{h_n}\big)-\theta_0\Big) = \sqrt{n}\Big(n^{-1}\sum_{i=1}^{n}X_{ni}-\theta_0\Big), \tag{A.42}$$

where $X_{ni}=F_K((t_0-X_i)/h_n)$. It is easy to show that the d.f. G_n of X_{ni} is given by

$$G_n(x) = \begin{cases} 0 & \text{if } x<0\\ 1-F_0(t_0-h_nF_K^{-1}(x)) & \text{if } 0<x<1\\ 1 & \text{if } x\geq 1. \end{cases} \tag{A.43}$$

Thus, G_n is continuous on its support $(0,1)$ with jump sizes at $x=0$ and $x=1$ as

$$c_0 = 1-F_0(t_0+h_n) \quad \text{and} \quad c_1 = F_0(t_0-h_n), \tag{A.44}$$

respectively, because $F_K^{-1}(y) = 2y - 1$ for $0 < y < 1$. Moreover, the derivative g_n of G_n is

$$g_n(x) = \begin{cases} 2h_n f_0(t_0 + h_n - 2h_n x) & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere,} \end{cases} \tag{A.45}$$

which means $\|g_n\| \le 2h_n\|f_0\|$.

To study the d.f. F_Y, we let H_n be the d.f. of $\sum_{i=1}^n X_{ni}$. Then, for $n = 2$,

$$H_n(x) = P\{X_{n1} + X_{n2} \le x\} = \int_{-\infty}^{\infty} G_n(x-u)\,dG_n(u). \tag{A.46}$$

Thus, for any a by DCT we have

$$\begin{aligned} \lim_{x\to a-}[H_n(a) - H_n(x)] &= \int_{-\infty}^{\infty} \lim_{x\to a-}[G_n(a-u) - G_n(x-u)]\,dG_n(u) \\ &= \int_{-\infty}^{\infty} I\{a-u=0 \text{ or } 1\}[G_n(a-u) - G_n((a-u)-)]\,dG_n(u) \\ &= \int_{-\infty}^{\infty} [I\{a-u=0\}c_0 + I\{a-u=1\}c_1]\,dG_n(u) \\ &= c_0[G_n(a) - G_n(a-)] + c_1[G_n(a-1) - G_n((a-1)-)] \\ &= c_0^2 I\{a=0\} + 2c_0c_1 I\{a=1\} + c_1^2 I\{a=2\}. \end{aligned}$$

By induction, we can show that for the general case n,

$$H_n(a) - H_n(a-) = \sum_{i=0}^{n} \frac{n!}{i!(n-i)!} c_0^{n-i} c_1^i I\{a=i\}. \tag{A.47}$$

Hence,

$$\begin{aligned} &[\text{total jump size of } H_n] = (c_0 + c_1)^n = [1 - F_0(t_0 + h_n) + F_0(t_0 - h_n)]^n \\ &\quad = [1 - 2h_n f_0(\xi)]^n \sim e^{-2f_0(t_0)h_n n} = (e^{-2f_0(t_0)})^{h_n n} = c^{h_n n}, \end{aligned} \tag{A.48}$$

where ξ is between $(t_0 - h_n)$ and $(t_n + h_n)$, and for $f_0(t_0) > 0$,

$$0 < c = e^{-2f_0(t_0)} < 1. \tag{A.49}$$

Now consider the derivative of H_n: $h_n = H_n'$. Let $c_2 = \int_{-\infty}^{\infty} g_n(x)\,dx = \int_0^1 g_n(x)\,dx$, then $c_0 + c_1 + c_2 = 1$. For case $n = 2$, we have that from (A.46),

$$H_n(x) = \int_{-\infty}^{\infty} G_n(x-u)\,dG_n(u) = c_0 G_n(x) + c_1 G_n(x-1) + \int_0^1 G_n(x-u) g_n(u)\,du.$$

Since G_n is piecewise differentiable with bounded derivative, we know that H_n is piecewise differentiable with derivative

$$h_n(x) = c_0 g_n(x) + c_1 g_n(x-1) + \int_0^1 g_n(x-u) g_n(u)\, du \le (c_0 + c_1)\|g_n\| + c_2\|g_n\| = \|g_n\|$$

for any x. By induction, we can show that for the general case n,

$$\|h_n\| \le \|g_n\|. \tag{A.50}$$

Since (A.42) gives $F_Y(y) = P\{Y \le y\} = P\{\sum_{i=1}^n X_{ni} \le \sqrt{n}\, y + \theta_0 n\} = H_n(\sqrt{n}\, y + \theta_0 n)$, then from (A.45), (A.48) and (A.50) we know that F_Y is piecewise continuous with

$$[\text{total jump size of } F_Y] = [\text{total jump size of } G_n] = (c_0 + c_1)^n \sim c^{h_n n},$$

where $0 < c < 1$ as in (A.49), and F_Y is piecewise differentiable with derivative uniformly bounded, because by (A.32),

$$F_Y'(y) = \sqrt{n} h_n(\sqrt{n} y + n\theta_0) \le \sqrt{n}\|h_n\| \le \sqrt{n}\|g_n\| \le 2h_n\sqrt{n}\|f_0\| = O(1).$$

Therefore, F_Y satisfies the Lipschitz condition up to a remainder term converging to 0 at rate of $c^{h_n n}$; that is there exists a constant $0 < M_0 < \infty$ such that for any x and y,

$$|F_Y(x) - F_Y(y)| \le M_0|x-y| + M_0 c^{h_n n}, \quad \text{for sufficiently large } n. \qquad \square$$

Acknowledgment: The author thanks two referees for their comments and suggestions on the earlier version of the manuscript.

REFERENCES

[1] Banerjee, M. and Wellner, J.A. (2004). Confidence intervals for current status data. *Scan. J. of Statist.* (To appear)

[2] Bickel, P.J. and Ren, J. (1996). The m out of n bootstrap and goodness of fit tests with doubly censored data. *Lecture Notes in Statistics, Springer Verlag* **109** 35-47.

[3] Chung, K.L. (1974). *A Course in Probability Theory.* Academic Press, INC, New York.

[4] De Gruttola, V and Lagakos, S.W. (1989). Analysis of doubly-censored survival data, with application to AIDS. *Biometrics* **45** 1-11.

[5] DiCiccio, T.J., Hall, P.J. and Romano, J. (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist.* **19** 1053-1061.

[6] Efron, B. (1967). The two sample problem with censored data. Proc. Fifth Berkeley *Symp. Math. Stat. Prob.* **4** 831-853.
[7] Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap.* Chapman & Hall, New York.
[8] Enevoldsen, A.K., Borch-Johnson, K., Kreiner, S., Nerup, J. and Deckert, T. (1987). Declining incidence of persistent proteinuria in type I (insulin-dependent) diabetic patient in Denmark. *Diabetes* **36** 205-209.
[9] Gill, R.D. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.*, **11**, 49-58.
[10] Gu, M.G. and Zhang, C.H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.* **21** 611-624.
[11] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer-Verlag, New York.
[12] Huang, J. (1999). Asymptotic properties of nonparametric estimation based on partly interval-censored data. *Statistica Sinica* **9** 501-519.
[13] Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457-481.
[14] Li, G., Hollander, M., McKeague, I.W., and Yang, J. (1996). Nonparametric likelihood ratio confidence bands for quantile functions from incomplete survival data. *Ann. Statist.* **24** 628-640.
[15] Miller, R.G. (1976). Least squared regression with censored data. *Biometrika* Vol. 63, 449-464.
[16] Murphy, S.A. and van der Vaart, A.W. (1997). Semiparametric likelihood ratio inference. *Ann. Statist.* **25** 1471-1509.
[17] Mykland, P.A. (1995). Dual likelihood. *Ann. Statist.* **23**, 396-421.
[18] Mykland, P.A. and Ren, J. (1996). Self-consistent and maximum likelihood estimation for doubly censored data. *Ann. Statist.* **24** 1740-1764.
[19] Odell, P. M., Anderson, K.M., and D'Agostino, R.B. (1992). Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model. *Biometrics* **48** 951-959.
[20] Owen, A.B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237-249.
[21] Owen, A.B. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18** 90-120.
[22] Owen, A.B. (1991). Empirical likelihood for linear models. *Ann. Statist.* **19** 1725-1747.
[23] Owen, A.B. (2001). *Empirical Likelihood.* Chapman & Hall, New York.
[24] Qin, J. and Lawless, J.F. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22** 300-325.
[25] Ren, J. (2001). Weighted empirical likelihood ratio confidence intervals for the mean with censored data. *Ann. Inst. Statist. Math.* **53** 498-516.
[26] Ren, J. (2005). Weighted empirical likelihood ratio confidence intervals for quantiles. (revised for *Statistica Sinica*)
[27] Ren, J. and Peer, P.G. (2000). A study on effectiveness of screening mammograms. *International Journal of Epidemiology* **29** 803-806.
[28] Royden, H.L. (1968). *Real Analysis.* MacMillan Publishing Co., INC, New York.
[29] Stute, W. and Wang, J.L. (1993). The strong law under random censorship. *Ann. Statist.* **21**, 1591-1607.
[30] Thomas, D.R. and Grunkemeir, G.L. (1975). Confidence interval estimation of survival probabilities for censored data. *J. Amer. Statist. Assoc.* **70** 865-871.
[31] Turnbull, B.W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.* **69** 169-173.