Ann. Inst. Statist. Math.
Vol. 53, No. 3, 517–527 (2001)

EMPIRICAL LIKELIHOOD FOR A CLASS OF FUNCTIONALS OF SURVIVAL DISTRIBUTION WITH CENSORED DATA

QI-HUA WANG[1]* AND BING-YI JING[2]**

[1] *Institute of Applied Mathematics, Academy of Mathematics and System Science, Chinese Academy of Science, Beijing, 10080, China*
[2] *Department of Mathematics, Hong Kong University of Science and Technology, Kowloon, Hong Kong,* e-mail: majing@ust.hk

(Received May 24, 1999; revised March 6, 2000)

Abstract. The empirical likelihood was introduced by Owen, although its idea originated from survival analysis in the context of estimating the survival probabilities given by Thomas and Grunkemeier. In this paper, we investigate how to apply the empirical likelihood method to a class of functionals of survival function in the presence of censoring. We define an adjusted empirical likelihood and show that it follows a chi-square distribution. Some simulation studies are presented to compare the empirical likelihood method with the Studentized-t method. These results indicate that the empirical likelihood method works better than or equally to the Studentized-t method, depending on the situations.

Key words and phrases: Emprical likelihood, censoring, Kaplan-Meier estimate, survival probability, mean lifetime, Studentized-t.

1. Introduction

Many statistical experiments result in incomplete sample, even under well-controlled conditions. This is because individuals will experience some other competing events which cause them to be removed. In such cases, the event of interest is not observable and the situation is termed random censoring. Censored data often arise in the study of medical follow-up, survival analysis, biometry and reliability study. In the last two decades, statistical inference with censored data has been paid considerable attention and studied extensively.

Let $X_1, \ldots, X_n$ be nonnegative independent and identically distributed (i.i.d.) random variables (r.v.) denoting survival times with the unknown distribution function F. Let $Y_1, \ldots, Y_n$ be nonnegative i.i.d. r.v. denoting censoring times with the distribution function G. It is assumed that X_i's and Y_i's are independent. In the random censoring model, the true survival times $X_1, \ldots, X_n$ are not observable. Instead, one observes only $Z_i = \min(X_i, Y_i)$, and $\delta_i = I(X_i < Y_i)$, where $I(\cdot)$ denotes the indicator function. A well-known nonparametric estimator of F is the Kaplan-Meier (KM) product limit

*Research Supported by the National Natural Science Foundation of China.
**Research Supported by Hong Kong RGC CERG No. HKUST6162/97P.

estimator defined by

$$1-\hat{F}(t)=\prod_{i=1}^{n}\left[\frac{n-i}{n-i+1}\right]^{I(Z_{(i)}\le t,\delta_{(i)}=1)}, \quad \text{for} \quad t\le Z_{(n)},$$

where $Z_{(1)} \le \cdots \le Z_{(n)}$ are the ordered statistics of the Z-sample and $\delta_{(i)}$ is the δ associated with $Z_{(i)}$. Also define $1-\hat{F}(t)=0$ if $t>Z_{(n)}$.

Let us consider the functional of $F(\cdot)$ in the form of

$$\theta(F)=\int_0^\infty \xi(t)dF(t), \tag{1.1}$$

where $\xi(t)$ is some (nonnegative) measurable function satisfying $E\xi(X)<\infty$. Some special cases of $\theta(F)$ include the mean lifetime of F if $\xi(t)=t$, the survival probability $1-F(t_0)$ at a fixed time t_0 if $\xi(t)=I(t\ge t_0)$, and the cumulative hazard function $\int_0^{t_0}(1-F(t))^{-1}dF(t)$ at a fixed time t_0 if $\xi(t)=I(t\le t_0)/[1-F(t)]$. Clearly, an obvious estimator of $\theta(F)$ can be obtained by replacing $F(t)$ by its KM estimator $\hat{F}(t)$ resulting in

$$\hat{\theta}=\theta(\hat{F})=\int_0^{Z_{(n)}}\xi(t)d\hat{F}(t). \tag{1.2}$$

Asymptotic properties of $\theta(\hat{F})$ have been studied by several authors. For instance, Stute (1995) shows that under certain optimal integrability conditions on the function $\xi(\cdot)$, the distribution of $\sqrt{n}(\theta(\hat{F})-\theta(F))$ is asymptotically normal with mean 0 and some variance σ^2, say. In order to compute confidence intervals for $\theta(F)$, we need to estimate the asymptotic variance σ^2. However, the expression for σ^2 takes very complicated form, see the Lemma in Section 4 of this paper. Thus, the usual plug-in method, where one replaces F and G by their KM estimates $\hat{F}$ and $\hat{G}$, will produce a very complicated estimator for σ^2 for which its consistency may even not be guaranteed. An alternative way is to estimate σ^2 by the jackknife and Stute (1996) shows that this indeed provides a consistent estimate of σ^2. This method of confidence region construction will be referred to as the Studentized-t method with jackknife variance estimate. Clearly, this method produces confidence regions for $\theta(F)$ with asymptotically accurate coverage probabilities.

Despite its usefulness, there are several drawbacks associated with the Studentized-t method, where the variance estimates are either obtained by the jackknife method or the plug-in method. First of all, this method does not always work well for small samples. Secondly, the method is not range-preserving in the sense that confidence intervals for the parameter $\theta(F)$ obtained by the Studentized-t method with jackknife variance estimator or the plug-in variance estimator may contain values outside its range. An example of this is the mean lifetime, which must be nonnegative. However, if the jackknife variance estimate is not very stable, then it is possible that the confidence interval for the mean lifetime will contain negative values.

In this paper we shall investigate how to use the empirical likelihood method to overcome the drawbacks encountered by the Studentized-t method. It is interesting to note that the empirical likelihood based method was first applied in survival analysis by Thomas and Grunkemeier (1975) for constructing confidence intervals for a survival probability with censored data. However, it is Owen (1988, 1990) who shows that the idea has wide applicability to complete data and introduces the method for very general statistical problems. Much research work has been carried out by various authors since then, including DiCiccio *et al.* (1991), Qin and Lawless (1994), Chen (1993, 1994). Hall and La Scala (1990) gave an excellent exposition of the empirical likelihood method and also outlined some of the advantages of the empirical likelihood method over other competitors such as the normal approximations. For instance, the empirical likelihood is range-preserving and transformation respecting and usually has good small sample performances.

To the best of our knowledge, studies on Owen's empirical likelihood method have mostly been restricted to the complete data case. A natural problem is whether we can extend Owen's empirical likelihood to randomly censored data to cover more general cases than the survival probability studied by Thomas and Grunkemeier (1975). The answer to this question is affirmative. To be more precise, we shall generalize Owen's method to a class of functionals $\theta(F)$ defined in (1.1), which include the survival probability and the mean lifetime with randomly censored data. It turns out that the empirical likelihood ratio in this case will tend to a weighted χ^2 distribution, where the weight depends on the unknown population and is a result of random censoring. (See Theorem 2.1 in Section 2.) To apply the method, one needs to estimate the variance (e.g, by the jackknife variance estimator). Fortunately, the method still retains other properties associated with the ususal empirical likelihood, e.g, it is still range-preserving, the confidence interval is determined automatically by the sample.

We should note that our empirical likelihood method is developed for functionals of survival functions with censored data. In this sense, it is different from the empirical likelihood methods studied by Thomas and Grunkemeier (1975), Li (1995) and Murphy (1995) in that those methods use the product type constraints and can not be easily applied to the inference of the functionals considered such as the mean survival time.

The paper is organized as follows. In Section 2, an adjusted empirical likelihood method for the mean lifetime is described. We define an adjusted empirical likelihood and we shall see that the adjusted empirical likelihood follows a chi-square distribution. Some simulation studies are given in Section 3 for both the empirical likelihood method and the Studentized-t with jackknife variance estimate and comparisons will be made between these two methods. Finally, we shall give proofs of our main results in Section 4.

2. Methodology and main results

Let us first give some motivations for our definition of the empirical likelihood for $\theta(F)$. First notice that

$$
\begin{aligned}
E\left(\frac{\xi(Z_i)\delta_i}{1-G(Z_i)}\right) &= E\left[\frac{\xi\left(\min(X_i,Y_i)\right) I(X_i<Y_i)}{1-G\left(\min(X_i,Y_i)\right)}\right] \\
&= \iint_{x<y} \frac{\xi(x)}{1-G(x)} dF(x)dG(y) \\
&= \int_0^\infty \frac{\xi(x)}{1-G(x)} \int_x^\infty dG(y)dF(x) \\
&= \int_0^\infty \frac{\xi(x)}{1-G(x)} \left(1-G(x)\right) dF(x) \\
&= \int_0^\infty \xi(x)dF(x)
\end{aligned}
$$

$$= \theta(F), \qquad i = 1, \ldots, n.$$

Hence, the problem of testing whether θ_0 is the true parameter of $\theta(F)$ is equivalent to testing whether $E(\frac{\xi(Z_i)\delta_i}{1-G(Z_i)}) = \theta_0$, $i = 1, \ldots, n$. This can be done using Owen's empirical likelihood method (1990). Let $p = (p_1, \ldots, p_n)$ be a probability vector, i.e., $\sum_{i=1}^n p_i = 1$ and $p_i \geq 0$ for $1 \leq i \leq n$. Let F_p be the distribution function which assigns probability p_i at point $\frac{\xi(Z_i)\delta_i}{1-G(Z_i)}$. Hence,

$$\theta(F_p) = \sum_{i=1}^n p_i \left(\frac{\xi(Z_i)\delta_i}{1 - G(Z_i)} \right).$$

Then, the empirical likelihood, evaluated at true parameter value θ_0, is defined by

$$\tilde{L}(\theta_0) = \max_{\theta(F_p)=\theta_0, \sum p_i = 1} \prod_{i=1}^n p_i.$$

Since $G(\cdot)$ in the definition of $\theta(F_p)$ is assumed to be unknown, we can replace it by its KM estimator $\hat{G}(t)$ defined by

$$1 - \hat{G}_n(t) = \prod_{i=1}^n \left[\frac{n-i}{n-i+1} \right]^{I(Z_{(i)} \leq t, \delta_{(i)} = 0)}.$$

Then, we can defined an estimated empirical likelihood, evaluated at the true value θ_0 of $\theta(F)$, by

$$L(\theta_0) = \max \prod_{i=1}^n p_i$$

subject to

$$\sum_{i=1}^n p_i \left(\frac{\xi(Z_i)\delta_i}{1 - \hat{G}(Z_i)} \right) = \theta_0, \qquad \text{and} \qquad \sum_{i=1}^n p_i = 1.$$

For simplicity, let us write

$$V_{ni} = \frac{\xi(Z_i)\delta_i}{1 - \hat{G}_n(Z_i)}, \qquad \bar{V}_n = \frac{1}{n} \sum_{i=1}^n V_{ni}.$$

Then, by the Lagrange multiplier, we can easily get

$$p_i = \frac{1}{n} \{1 + \lambda(V_{ni} - \theta_0)\}^{-1}, \qquad i = 1, \ldots, n,$$

where λ is the solution of

$$\frac{1}{n} \sum_{i=1}^n \frac{(V_{ni} - \theta_0)}{1 + \lambda(V_{ni} - \theta_0)} = 0. \tag{2.1}$$

Note that $\prod_{i=1}^n p_i$, subject to $\sum_{i=1}^n p_i = 1$, attains its maximum n^{-n} at $p_i = n^{-1}$. So we define the empirical likelihood ratio at θ_0 by

$$R(\theta_0) = \prod_{i=1}^n (np_i) = \prod_{i=1}^n \{1 + \lambda(V_{ni} - \theta_0)\}^{-1},$$

and the corresponding empirical log-likelihood ratio is defined as

$$l(\theta_0) = -2\log R = 2\sum_{i=1}^{n}\log\{1+\lambda(V_{ni}-\theta_0)\}, \tag{2.2}$$

where λ is the solution of (2.1).

Under certain conditions, it can be shown that $l(\theta_0)$, multiplied by some population quantity, follows a chi-square distribution with 1 degree of freedom. In other words, $rl(\theta_0) \sim \chi_1^2$ asymptotically. In order to use this to construct a confidence interval for θ_0, one must first estimate the coefficient r. To that end, let $n\widehat{\text{Var}}^*(jack)$ be the modified jackknife estimate of the asymptotic variance of $\hat{\theta}$ due to Stute (1996). Define

$$\hat{l}(\theta) = \hat{r}l(\theta), \tag{2.3}$$

where

$$\hat{r} = \frac{n^{-1}\sum_{i=1}^{n}\left(V_{ni}-\bar{V}_n\right)^2}{n\widehat{\text{Var}}^*(jack)}.$$

The following theorem gives a nonparametric version of Wilks' theorem for the adjusted empirical likelihood in the censored case. The proof of the theorem is given in Section 4. Before stating the theorem, we define

$$\bar{H}(s) = P(Z_1 > s), \qquad \tilde{H}_0(s) = P(Z_1 > s, \delta_1 = 0),$$

$$\tilde{H}_1(s) = P(Z_1 > s, \delta_1 = 1), \qquad \gamma_0(x) = \exp\left\{\int_0^{x-}\frac{d\tilde{H}_0(s)}{\bar{H}(s)}\right\},$$

$$C(x) = \int_0^{x-}\frac{dG(s)}{(1-H(s))(1-G(s))}, \qquad \tau_H = \inf\{t : H(t) = 1\}.$$

THEOREM 2.1. *Assume the conditions*

(C1) $\displaystyle\int_0^{\tau_H}\xi^2(x)\gamma_0^2(x)d\tilde{H}_1(x) < \infty,$

(C2) $\displaystyle\int_0^{\tau_H}\xi(x)C^{1/2}(x)dF(x) < \infty,$

(C3) $\displaystyle\int_0^{\tau_H}\frac{\xi^2(x)dF(x)}{1-G(x)} < \infty,$

(C4) $\tau_F = \tau_H$ *and* $F(\tau_F) = F(\tau_F-)$.

Let θ_0 *be the true value of* θ, *then* $\hat{l}(\theta_0)$ *has an asymptotic chi-square distribution with 1 degree of freedom, that is,*

$$\hat{l}(\theta_0) \xrightarrow{\mathcal{L}} \chi_1^2. \tag{2.4}$$

A simple approach to construct an α-level confidence interval for θ, based on (2.4), is $I_\alpha = \{\theta : \hat{l}(\theta) \le c_\alpha\}$ with $P(\chi_1^2 \le c_\alpha) \le 1-\alpha$. Then by Theorem 2.1, I_α will have the correct coverage probability $1-\alpha$ asymptotically, i.e. $P(\theta_0 \in I_\alpha) = 1-\alpha+o(1)$.

Remark 2.1. The definition of the empirical log-likelihood ratio $l(\theta_0)$, defined by (2.2) and (2.1), looks rather similar to that for the mean in the absence of censoring. However, the difference here is that V_{ni}'s, $i = 1, \ldots, n$, are no longer independent and identically distributed random variables, due to the estimation of $G(z)$ by its KM estimator $\hat{G}_n(z)$.

3. Some simulation studies

We shall conduct some simulation studies to compare the performances between the empirical likelihood method and the Studentized-t method with the variance estimated by the jackknife proposed by Stute (1996). In particular, we shall investigate two specific functionals of the survival functions: the first being the mean lifetime $\mu = \int_0^\infty x dF(x)$ and the second being the survival probability at a fixed point t_0, namely $1 - F(t_0)$.

In both situations, we generate the lifetimes and censoring times from exponential distributions with parameters 1 and c, respectively. That is, $F(x) = 1 - \exp(-x)$ and $G(x) = 1 - \exp(-cx)$ for $x \geq 0$. Here, $c > 0$ is chosen to accommodate certain preselected censoring proportions. The sample sizes have been chosen to be 10, 20, 50 and 100. The coverage probabilities are calculated for the empirical likelihood and Studentized-t methods based on 1000 pairs of simulated data generated from the lifetime and censoring distributions. The nominal levels are taken to be 0.90 and 0.95. The results are presented in Tables 1 and 2.

For the mean lifetime, we make the following observations from Table 1.

(1) For fixed censoring proportions, the coverage accuracies for both the empirical likelihood and the Studentized-t method generally increase as the sample size n increases, as can be expected.

(2) The performances of both methods critically depend on the censoring proportions. For a fixed sample size n, the coverage accuracies for both the empirical likelihood and the Studentized-t method generally decrease as the censoring proportion increases.

(3) The empirical likelihood works uniformly better than the Studentized-t method. The contrast in their performances is made even more transparent when the sample size n is small (say $n = 10, 20$) and the censoring proportion is relatively large (see the case where censoring proportion is 0.4).

For the survival probability example, the simulation results given in Table 2 are not as clear cut as for those of the mean lifetime. But first, let us continue to make several remarks from Table 2.

(4) Both the empirical likelihood method and Studentized-t method work very well for moderate to large sample sizes (e.g., $n \geq 20$). In fact, in these cases, it is very difficult to assess which method is better than the other since neither of the two methods has better coverage accuracies all the time.

(5) When $n = 10$, some of the coverage probabilities for the empirical likelihood method could not be calculated. (These are marked as "NA" in Table 2.) To understand why this happens, we recall that $V_{ni} = \xi(Z_i)\delta_i/(1 - \hat{G}(Z_i))$ and $\xi(Z_i) = I(Z_i \geq t_0)$. So when n is small, our simulated data (Z_i, δ_i)'s could sometimes result in $V_{ni} = 0$ for all $i = 1, \ldots, n$, under which one could not solve for λ from (2.1) and hence could not calculate the log-likelihood ratio $l(\theta_0)$.

We close this section with some general remarks regarding the two methods. Our simulation results reveal that, in terms of coverage accuracy, the empirical likelihood method works well for both the mean lifetime and the survival probability at time t_0.

Table 1. Coverage probabilities for the mean lifetime.

		Nominal level is 0.90		Nominal level is 0.95	
Censoring proportion	n	Studentize-t	Empirical likelihood	Studentize-t	Empirical likelihood
	10	0.770	0.860	0.810	0.880
0.10	20	0.830	0.870	0.870	0.920
	50	0.866	0.900	0.904	0.936
	100	0.892	0.936	0.936	0.960
	10	0.710	0.750	0.730	0.780
0.25	20	0.790	0.870	0.800	0.890
	50	0.786	0.908	0.848	0.930
	100	0.804	0.890	0.856	0.940
	10	0.510	0.630	0.520	0.680
0.40	20	0.600	0.760	0.700	0.800
	50	0.632	0.768	0.674	0.798
	100	0.676	0.810	0.736	0.854

Table 2. Coverage probabilities for the survival probability $1 - F(0.5)$.

		Nominal level is 0.90		Nominal level is 0.95	
Censoring proportion	n	Studentize-t	Empirical likelihood	Studentize-t	Empirical likelihood
	10	0.886	NA	0.891	NA
0.10	20	0.904	0.920	0.950	0.970
	50	0.884	0.886	0.940	0.948
	100	0.914	0.918	0.952	0.956
	10	0.851	NA	0.886	NA
0.25	20	0.888	0.923	0.930	0.958
	50	0.906	0.916	0.940	0.950
	100	0.900	0.918	0.960	0.974
	10	0.754	NA	0.794	NA
0.40	20	0.862	0.924	0.902	0.952
	50	0.882	0.936	0.942	0.968
	100	0.894	0.938	0.936	0.974

On the other hand, the Studentized-t method also works well in the survival probability example, but less well in the mean lifetime example. This could perhaps be explained as follows. As mentioned in the Introduction, the performance for the Studentized-t method relies heavily on whether one can obtain a stable variance estimator. In the first example, the variance of $\hat{\mu} = \int_0^{Z_{(n)}} x d\hat{F}(x)$ and its estimates (e.g., by the plug-in method or the jackknife) relies on the properties of $\hat{F}(x)$ and $\hat{G}(x)$ on the whole interval $[0, \tau_H)$, which may not be stable near the end point τ_H. This subsequently affects the performance of the Studentized-t method. On the other hand, in the second example, the variance of $1-\hat{F}(t_0)$ and its estimates only rely on the properties of the KM estimates $\hat{F}(\cdot)$ and $\hat{G}(\cdot)$ on the interval $[0, t_0]$, which are known to be good. Therefore, it is not surprising that the Studentized-t method works well in the second example but not so well in the first one.

4. Proofs

Let $W_{ni} = V_{ni} - \theta_0$ and $\bar{W}_n = n^{-1}\sum_{i=1}^n W_{ni}$. Let

$$\gamma_1(x) = \frac{1}{\bar{H}(x)} \int I[x < s]\xi(s)\gamma_0(s)d\tilde{H}_1(s),$$
$$\gamma_2(x) = \iint \frac{I[s < x, s < t]\xi(t)\gamma_0(t)}{\bar{H}^2(s)} d\tilde{H}_0(s)d\tilde{H}_1(t).$$

To prove Theorem 2.1, we need the following lemma.

LEMMA. *Under the same conditions as in Theorem* 2.1, *we have*

$$\sqrt{n}\,\bar{W}_n \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

where

$$\sigma^2 = \text{Var}\left(\xi(Z_1)\gamma_0(Z_1)\delta_1 + \gamma_1(Z_1)(1-\delta_1) - \gamma_2(Z_1)\right).$$

PROOF OF THE LEMMA. Notice that

$$\bar{W}_n = \int_0^\infty \xi(z)d(\hat{F}_n(z) - F(z)). \tag{4.1}$$

Under the condition (C4), $\tilde{F}$ defined by Stute (1995) is just F. Hence, (4.1) and Corollary 1.2 of Stute (1995) together prove the lemma.

PROOF OF THE THEOREM. Under the condition (C3), we have

$$\frac{1}{n}\sum_{i=1}^n \left(\frac{\delta_i\xi(Z_i)}{1-G(Z_i)}\right)^2 = O_p(1).$$

From this and the following result due to Zhou (1992)

$$\sup_{0\le z\le Z_{(n)}} \left(\frac{\hat{G}_n(z) - G(z)}{1-\hat{G}_n(z)}\right) = O_p(1), \tag{4.2}$$

we have $n^{-1}\sum_{i=1}^n V_{ni}^2 = O_p(1)$ and

$$\frac{1}{n}\sum_{i=1}^n W_{ni}^2 = O_p(1). \tag{4.3}$$

From (4.3) and the fact that $n^{-1}\sum_{i=1}^{n} V_{ni} = O_p\left(n^{-1/2}\right)$ (implied by the lemma), we can show that, similarly to Owen (1990),

$$\lambda = O_p(n^{-1/2}). \tag{4.4}$$

Let $V_{Gi} = \frac{\delta_i \xi(Z_i)}{1-G(Z_i)}$. By the condition (C3), following the proof of lemma 3 in Owen (1990) line by line, it follows that $\max_{1\le i\le n}|V_{Gi}| = o_p(n^{1/2})$. This together with (4.2) proves that

$$\begin{aligned}\max_{1\le i\le n}|V_{ni}| &\le \max_{1\le i\le n}|V_{Gi}| + \max_{1\le i\le n}\left|\frac{\delta_i\xi(Z_i)\left(\hat{G}_n(Z_i)-G(Z_i)\right)}{(1-G(Z_i))(1-\hat{G}_n(Z_i))}\right| \\ &\le o_p(n^{1/2}) + \sup_{0\le z\le Z_{(n)}}\left|\frac{\hat{G}_n(z)-G(z))}{1-\hat{G}_n(z)}\right| \max_{1\le i\le n}|V_{Gi}| = o_p(n^{1/2}).\end{aligned}$$

Hence

$$\max_{1\le i\le n}|W_{ni}| = o_p(n^{1/2}). \tag{4.5}$$

From (2.1), we have

$$\frac{1}{n}\sum_{i=1}^{n}\frac{W_{ni}}{1+\lambda W_{ni}} = \frac{1}{n}\sum_{i=1}^{n} W_{ni}\left(1-\lambda W_{ni}+\frac{\lambda^2 W_{ni}^2}{1+\lambda W_{ni}}\right) = 0.$$

Hence,

$$\lambda = \frac{\sum_{i=1}^{n} W_{ni}}{\sum_{i=1}^{n} W_{ni}^2} + \gamma_n, \tag{4.6}$$

where

$$\gamma_n = \lambda^2 \frac{n^{-1}\sum_{i=1}^{n}\frac{W_{ni}^3}{1+\lambda W_{ni}}}{n^{-1}\sum_{i=1}^{n} W_{ni}^2}.$$

Now (4.3), (4.4) and (4.5) together prove

$$\gamma_n \le O_p(n^{-1}) \max_{1\le i\le n}|W_{ni}| = o_p(n^{-1/2}) \tag{4.7}$$

by the fact that $n^{-1}\sum_{i=1}^{n}\frac{W_{ni}^3}{1+\lambda W_{ni}} \le 2n^{-1}\left(\max_{1\le i\le n}|W_{ni}|\right)\sum_{i=1}^{n} W_{ni}^2$ in probability.

By (2.1), (4.3), (4.4) and (4.5), we have

$$0 = \lambda\sum_{i=1}^{n}\frac{W_{ni}}{1+\lambda W_{ni}} = \sum_{i=1}^{n}\frac{\lambda W_{ni}}{1+\lambda W_{ni}} = \sum_{i=1}^{n}(\lambda W_{ni}) - \sum_{i=1}^{n}(\lambda W_{ni})^2 + o_p(1).$$

That is

$$\sum_{i=1}^{n}(\lambda W_{ni}) = \sum_{i=1}^{n}(\lambda W_{ni})^2 + o_p(1). \tag{4.8}$$

Hence from (2.2) and (4.8), we have

$$l(\theta_0) = 2\sum_{i=1}^{n} \log\{1+\lambda W_{ni}\} \tag{4.9}$$

$$= 2\sum_{i=1}^{n} \left(\lambda W_{ni} - \frac{1}{2}(\lambda W_{ni})^2\right) + r_n \tag{4.10}$$

$$= \lambda^2 \sum_{i=1}^{n} W_{ni}^2 + r_n, \tag{4.11}$$

where

$$r_n \le \lambda^3 \sum_{i=1}^{n} W_{ni}^3 \le O_p(n^{-3/2}) \max_{1\le i\le n} |W_{ni}| \sum_{i=1}^{n} W_{ni}^2 = o_p(1).$$

Hence (4.6), (4.7) and (4.11) together gives

$$l(\theta_0) = \frac{(\sqrt{n}\bar{W}_n)^2}{n^{-1}\sum_{i=1}^{n} W_{ni}^2} + o_p(1). \tag{4.12}$$

By Stute and Wang (1993), we have with probability 1

$$\bar{V}_n = \frac{1}{n}\sum_{i=1}^{n} V_{ni} \to \theta_0. \tag{4.13}$$

This concludes that

$$n^{-1}\sum_{i=1}^{n} \left(V_{ni} - \bar{V}_n\right)^2 - \frac{1}{n}\sum_{i=1}^{n} W_{ni}^2 \xrightarrow{a.s.} 0. \tag{4.14}$$

It is known from Theorem 1.2 of Stute (1996) that

$$n\widehat{\mathrm{Var}}^{*}(jack) \xrightarrow{a.s.} \sigma^2. \tag{4.15}$$

Clearly, it follows from (4.12), (4.14) and (4.15) that

$$\hat{l}(\theta_0) = \left(\frac{\sqrt{n}\bar{W}_n}{\sigma}\right)^2 + o_p(1),$$

where $\hat{l}(\theta)$ is defined in (2.3). Therefore, Theorem 2.1 follows straightaway from the lemma.

References

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

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

DiCiccio, T. J., Hall, P. and Romano, J. P. (1991). Bartlett adjustment for empirical likelihood, *Ann. Statist.*, **19**, 1053–1061.

Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood, *International Statistical Review*, **58** (2), 109–127.

Li, G. (1995). On nonparametric likelihood ratio estimation of survival probabilities for censored data, *Statist. Probab. Lett.*, **90**, 95–104.

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

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

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

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

Stute, W. (1995). The central limit theorem under random censorship, *Ann. Statist.*, **23**, 422–439.

Stute, W. (1996). The jackknife estimate of variance of a Kaplan-Meier integral, *Ann. Statist.*, **24**, 2679–2704.

Stute, W. and Wang, J.-L. (1993). The strong law under random censorship, *Ann. Statist.*, **21**, 1591–1607.

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

Zhou, M. (1992). Asymptotic normality of the synthetic estimator for censored survival data, *Ann. Statist.*, **20**, 1002–1021.