EMPIRICAL LIKELIHOOD RATIO IN TERMS OF CUMULATIVE HAZARD FUNCTION FOR CENSORED DATA

Xiao-rong Pan and Mai Zhou

University of Kentucky and Searle

Abstract

It has been shown that (with complete data) empirical likelihood ratios can be used to form confidence intervals and test hypothesis about a linear functional of the *distribution function* just like the parametric case. We study here the empirical likelihood ratios for right censored data and with parameters that are *linear functionals* of the *cumulative hazard function.* Martingale techniques make the asymptotic analysis easier, even for random weighting functions. It is shown that the empirical likelihood ratio in this setting can be easily obtained by solving a one parameter monotone equation.

AMS 1991 subject classifications. Primary 62G10; Secondary 62G05.

Key Words and Phrases: Weighted hazard, One sample log rank test, Stochastic constraint, Median.

1. Introduction

Based on the likelihood function there are 3 different methods to produce confidence intervals: namely Wald's method, Rao's method and Wilks' method. Among the 3, the Wilks likelihood ratio (LR) method do not need the calculation of information or inverse of that. It *automatically* adjust the statistics $-2\log \mathrm{LR}$ to a pivotal. This can be a real advantage in the case where the information (or inverse of it) is difficult to estimate. Even when all 3 are easy to obtain, the LR method still holds some unique advantages. For example, the confidence intervals produced by the LR method is always range respecting (confidence bounds inside the parameter space), while the other two is not. Therefore, transformation on the parameter is often used in connection with the Wald's and Rao's method to overcome the range problem. However, the choice of the transformation is ad hoc. For new parameters it is often unclear what transformation to use. In this respect, the LR method can be described as to achieve the result comparable to the Wald's method with the *best* transformation, but without the need to explicitly find the best transformation.

Recently, Owen (1988, 1990) and many others showed that the likelihood ratio method can also be used to produce confidence intervals in the nonparametric settings after some modification. He term this empirical likelihood ratio method. The empirical likelihood (EL) of n i.i.d. observations X_i are just

$$EL(F) = \prod_{i=1}^{n} \Delta F(X_i) \ .$$

Without any restrictions, the empirical distribution function, $F = \hat{F}_n(t) = 1/n \sum I_{[X_i \le t]}$, will maximize the EL among all possible distribution functions, therefore it is referred to as nonparametric maximum likelihood estimator or NPMLE. With a linear constraint of the form

$$\int g(t) dF(t) = \mu \ , \tag{1.1}$$

Owen (1988, 1990) showed that the distribution function that maximize the EL subject to the constraint can be calculated using the Lagrange multiplier method. He showed that such a distribution function F has jump at X_i equals to

$$\Delta F(X_i) = \Delta \hat{F}_n(X_i) \times \frac{1}{1 + \lambda(g(X_i) - \mu)} \ ,$$

where λ is defined by the following equation

$$\sum_{i=1}^{n} \Delta \hat{F}_n(X_i) \frac{g(X_i)}{1 + \lambda(g(X_i) - \mu)} = \mu.$$

Once the constrained maximum is obtained, it can be shown that the empirical likelihood ratio statistic, $-2 \log ELR(\mu)$, converges in distribution to a chi-square distribution (Owen 1988). However a generalization of the above to the right censored data case is more complicated.

In the analysis of censored data, it is often more convenient to model the data in terms of the (cumulative) hazard function $\Lambda(t)$ which is defined by

$$\Lambda(t) = \int_{[0,\ t)} \frac{dF(s)}{1 - F(s-)} \ . \tag{1.2}$$

It gives rise to a martingale formulation of the observations. For example, regression model in terms of hazard leads to the Cox proportional hazards model, nonparametric estimation in terms of cumulative hazard leads to the Nelson-Aalen estimator which is much easier to analyze than the Kaplan-Meier estimator. Also, information in terms of hazard (Efron & Johnston 1990), Hellinger distance in terms of hazard (Ying 1992) all have been studied and proved to be informative.

Therefore it is natural to look at the Empirical Likelihood in terms of hazard and constraints in terms of hazard as in (2.6). It turns out that the theory for the EL in terms of hazard is much simpler for right censored data. Also, martingale formulation makes it easy to handle even stochastic (predictable) weight functions.

We obtained results for general parameters of the following types: (1) $\theta = \int g(t) d\Lambda(t)$ for arbitrary given $g(t)$. (2) $\theta_n = \int g_n(t) d\Lambda(t)$ where $g_n(t)$ is a random but predictable function and depend on sample size n. (3) θ is defined implicitly: $\int g(t, \theta) d\Lambda(t) = C$ for a constant C.

Parameters of the first type can arise in the context of a time-dependent covariate Cox model. In such a model the cumulative hazard for a person with a time-dependent multiplicative covariate $g(t)$ can be computed as $\Lambda_i(\tau) = \int_0^\tau g(t) d\Lambda_b(t)$, where Λ_b is the baseline cumulative hazard. As a specific example, suppose smoking cigarettes doubles a person's hazard and a person started smoking at age 15 and quit smoking at age 45. Then his hazard up to age 55 is estimate by $\Lambda_i(55) = \int_0^{55} g(t) d\hat{\Lambda}_0(t)$ based on the Nelson-Aalen estimator $\hat{\Lambda}_0(t)$ computed from a cohort of non-smokers, where $g(t) = 2$ for $15 < t \leq 45$ and $g(t) = 1$ elsewhere.

The parameter of the second type is prompted by the one sample log-rank type tests. See, for example, Andersen et. al. (1993) Section V.1 for details. The weight function of the one sample log-rank type of tests takes the form $g(t) = [1 - F_0(t-)]^\rho Y(t)/n$ where $Y(t)$ is the size of the risk set at time t. Usually Wald type normal approximation is used for the log-rank test statistic without transformation. As a further example for stochastic weight function g, we take the statistic of mean, which can be obtained from the integration of cumulative hazard with $g(t) = t[1 - F(t-)]$. Since F is unknown, we may use $g_n(t) = t[1 - \hat{F}_n(t-)]$.

The prime example for the implicit type parameters are the quantiles. For example the parameter θ of median may be defined implicitly as $\int I_{[t \leq \theta]} d\Lambda(t) = \log 2$.

Murphy (1995) also studied the empirical likelihood ratio using counting process formulations. She obtained the explicit result only when the constraint is the hazard function itself evaluated at a point, $\Lambda(t_0) = -\log[1 - F(t_0)]$. Li (1995), building on earlier work of Thomas and Grunkemeier (1975), studied the empirical likelihood method for censored data, but only for the parameters of the form $F(t_0)$. Murphy and Van der Vaart (1997) proved a very general result but in each specific case one still need to workout the often non-trivial conditions, also it is not clear how the empirical likelihood should be computed. Our result gives a more explicit way to compute such intervals. We need only to find the root of a monotone univariate function. Once the root is found the likelihood ratio is easily obtained (see 3.2 or 4.0). Besides, none of the above papers deals with stochastic constraints.

Due to the similarity of technical treatment between the three types of constraints we shall present the detailed proof only for the first type of constraint and omit the proofs for the other 2 types of constraints. The rest of the paper is organized as follows: section 2 defines the likelihood in terms of hazard and calculates the maximum of the likelihood under the constraint of type 1. Section 3 studies the asymptotic behavior of the likelihood ratio and shows that it converges to a chi-square distribution. Section 4 looks at the difference between 2 versions of the likelihood. Section 5 deals with the stochastic constraint and the implicit constraint. Section 6 contains some examples. Finally some technical proofs are collected in the appendix.

Obviously, 2 sample and k sample analogs of the results presented here are possible. We shall present these and other generalization in a forthcoming paper.

2. Likelihood in terms of hazard and its maximum under a constraint of type 1

Suppose that $X_1, \cdots, X_n$ are i.i.d. nonnegative random variables denoting the lifetimes with a continuous distribution function F_0. Independent of the lifetimes there are censoring times $C_1, \cdots, C_n$ which are i.i.d. with a distribution G_0. Only the censored observations are available to us:

$$T_i = \min(X_i, C_i) \ ; \quad \delta_i = I[X_i \le C_i] \quad \text{for} \ \ i = 1, 2, \cdots, n. \tag{2.1}$$

The empirical likelihood based on censored observations $(T_i, \ \delta_i)$ pertaining F is

$$EL(F) = \prod_{i=1}^{n} [\Delta F(T_i)]^{\delta_i} [1 - F(T_i)]^{1-\delta_i} \ . \tag{2.2}$$

Since the NPMLE of the distribution F and hazard Λ are both known to be purely discrete functions (i.e. Kaplan-Meier/Nelson-Aalen estimator), it is reasonable to restrict the analysis of the likelihood ratio to the purely discrete functions dominated by their NPMLE's. This is similar to the use of sieves in the likelihood analysis. See Owen (1988) for more discussion on this restriction.

Using the relation between hazard and distribution

$$1 - F(t) = \prod_{s \le t} (1 - \Delta\Lambda(s)) \qquad \text{and} \qquad \Delta\Lambda(t) = \frac{\Delta F(t)}{1 - F(t-)} \tag{2.3}$$

that are valid for purely discrete distributions we can rewrite (2.2) in terms of cumulative hazard function. The empirical likelihood (2.2) becomes

$$EL(\Lambda) = \prod_{i=1}^{n} [\Delta\Lambda(T_i)]^{\delta_i} [\prod_{j: T_j < T_i} (1 - \Delta\Lambda(T_j)) \ ]^{\delta_i} [\prod_{j: T_j \le T_i} (1 - \Delta\Lambda(T_j)) \ ]^{1-\delta_i}. \tag{2.4}$$

The hazard function that maximizes the likelihood $EL(\Lambda)$ without any constraint is the Nelson-Aalen estimator, see e.g. Andersen et. al. (1993). We shall denote the Nelson-Aalen estimator by $\hat{\Lambda}_{NA}(t)$.

On the other hand, a simpler version of the likelihood can be obtained if we merge the second and third factor in (2.4) and replace it by $\exp[-\Lambda(T_i)]$, which was called a Poisson extension of the likelihood by Murphy (1994):

$$AL(\Lambda) = \prod_{i=1}^{n} [\Delta\Lambda(T_i)]^{\delta_i} \exp\{-\Lambda(T_i)\} \ . \tag{2.5}$$

See also Gill (1989) for a detailed discussion of different extensions of likelihood function for discrete distributions. Notice we have used a formula that is only valid for continuous distribution in the case of a discrete distribution. But the difference is small and negligible for large n as we shall see later. On the other hand, the maximizer for $AL(\Lambda)$ for finite n is also the Nelson-Aalen estimator, giving AL some legitimacy. We shall use AL in our analysis first due to its simplicity and examine the difference between AL and EL later.

The first and crucial step in our analysis is to find a (discrete) cumulative hazard function that maximize $AL(\Lambda)$ under the constraint (of type 1)

$$\int g(t) d\Lambda(t) = \theta \tag{2.6}$$

where $g(t)$ is a given function that satisfy some moment conditions, θ is a given constant.

We point out before proceeding that the last jump of a (proper) discrete cumulative hazard function must be one. This is evident from the relation (2.3), second equation. This restriction is similar to the "jumps sum to one" restriction on the discrete distribution functions. The consequence is that any discrete cumulative hazard function dominated by the Nelson-Aalen estimator must, at the last observation, have the same jump as the Nelson-Aalen estimator.

In light of this we rewrite the constraint (2.6) in terms of jumps. For simplicity we shall assume there is no tie in the uncensored observations. Without loss of generality we assume $T_1 \leq T_2 \leq \cdots \leq T_n$ where only possible ties are between censored observations.

Let $w_i = \Delta\Lambda(T_i)$ for $i = 1, 2, \cdots, n$, where we notice $w_n = \delta_n$. The constraint (2.6) for any Λ, that is dominated by Nelson-Aalen estimator, can be written as

$$\sum_{i=1}^{n-1} \delta_i g(T_i) w_i + g(T_n)\delta_n = \theta \ . \tag{2.7}$$

Similarly, the likelihood AL at this Λ can be written in term of the jumps

$$AL = \prod_{i=1}^{n} [w_i]^{\delta_i} \exp\{-\sum_{j=1}^{i} w_j\} \ . \tag{2.8}$$

Another important issue is that the constraint equation may not always have a solution for certain values of θ. An obvious example is when $g(t) \leq 0$ and $\theta > 0$. Thus for each given $g(t)$ and sample, we shall only study in detail the *feasible constraints*, those θ values that have at least one set of solution to (2.7). For those that do not have a solution we define the value of the likelihood under this constraint to be zero. Note that to be qualified as a solution, we must have $0 \leq w_i < 1$ for $i = 1, 2, \cdots, n-1$.

To find the maximizer of AL under constraint (2.7), we use Lagrange multiplier method. Once the constrained maximizer is found by Lagrange multiplier, (recall the un-constrained maximizer was known to be the Nelson-Aalen estimate), we can proceed to study the empirical likelihood ratio.

Theorem 1 *The feasible values of θ in the constraint (2.7) is given by the interval: $\mathcal{V}$ defined at the end of the proof.*

If the constraint (2.7) is feasible, then the maximum of AL under the constraint is obtained when

$$w_i = W_i = \frac{\delta_i}{(n-i+1)+n\lambda g(T_i)\delta_i} = \frac{\delta_i}{n-i+1} \times \frac{1}{1+\lambda \frac{\delta_i g(T_i)}{\frac{n-i+1}{n}}} , \tag{2.9}$$

where λ in turn, is the solution of the following equation

$$l(\lambda) = \theta \quad \textit{where} \quad l(\lambda) \equiv \sum_{i=1}^{n-1} g(T_i) \frac{\delta_i}{n-i+1} \times \frac{1}{1+\lambda \frac{\delta_i g(T_i)}{\frac{n-i+1}{n}}} + g(T_n)\delta_n. \tag{2.10}$$

Proof: To use Lagrange multiplier, we form the target function

$$G = \sum_{i=1}^{n} \delta_i \log w_i - \sum_{i=1}^{n} \sum_{j=1}^{i} w_j + n\lambda \left[\theta - \sum_{i=1}^{n-1} \delta_i g(T_i) w_i - \delta_n g(T_n) \right].$$

Taking partial derivative with respect to w_i, for $i = 1, \cdots, n-1$, and letting them equal to zero, we obtain

$$\frac{\partial G}{\partial w_i} = \frac{\delta_i}{w_i} - (n-i+1) - n\lambda g(T_i)\delta_i = 0, \qquad i = 1, 2, \cdots, n-1.$$

By solving this equation we get the explicit expression for w_i

$$\begin{aligned} W_i &= \frac{\delta_i}{(n-i+1)+n\lambda g(T_i)\delta_i} \\ &= \frac{\delta_i}{n-i+1} \times \frac{1}{1+\lambda \frac{\delta_i g(T_i)}{\frac{n-i+1}{n}}} \\ &= \Delta\hat{\Lambda}_{NA}(T_i) \frac{1}{1+\lambda \frac{\delta_i g(T_i)}{\frac{n-i+1}{n}}} \qquad \text{for } i = 1, 2, \cdots, n-1 \end{aligned}$$

where λ has to be chosen to satisfy the constraint (2.7). By plug W_i into (2.7) we see that λ can be obtained as a solution to the following equation

$$l(\lambda) \equiv \sum_{i=1}^{n-1} g(T_i) \frac{\delta_i}{n-i+1} \frac{1}{1+\lambda \frac{\delta_i g(T_i)}{\frac{n-i+1}{n}}} + g(T_n)\delta_n = \theta.$$

The function $l(\lambda)$ above is monotone decreasing and continuous in λ, a fact that can be verified by taking a derivative of $l(\lambda)$ with respect to λ. On the other hand, any choice of legitimate value λ

must result in w_i through (2.9) that are bona fide jumps of a discrete cumulative hazard function, which must be bounded between zero and one. This restriction leads to the following legitimate λ range $\mathcal{J}$:

All max and min in the following definitions are taken in the domain $\{i : 1 \le i \le n-1,\ \delta_i = 1,\ \text{and}\ g(T_i) \neq 0\}$, if there is any additional restriction then we specify in each individual case.

Case 1: when $\min g(T_i) > 0$

$$\mathcal{J} = \left(\max \frac{i-n}{n g(T_i)}, \quad \infty \right) := (\underline{\lambda}, \quad \infty),$$

Case 2: when $\max g(T_i) < 0$

$$\mathcal{J} = \left(-\infty, \quad \min \frac{i-n}{n g(T_i)} \right) := (-\infty, \quad \overline{\lambda}),$$

Case 3: when $\max g(T_i) > 0 > \min g(T_i)$

$$\mathcal{J} = \left(\max_{g(T_i)>0} \frac{i-n}{n g(T_i)}, \quad \min_{g(T_i)<0} \frac{i-n}{n g(T_i)} \right) := (\underline{\lambda}, \quad \overline{\lambda}).$$

Since the function $l(\cdot)$ is continuous and monotone, the corresponding range of θ value that make the equation (2.10) feasible (has a set of solution that is bona fide cumulative hazard function) are as follows. Notice these θ values also make the constraint (2.7) feasible.

Case 1:

$$\mathcal{V} = \left(g(T_n)\delta_n, \quad \sum_{i=1}^{n-1} \frac{\delta_i g(T_i)}{n-i+1+n\underline{\lambda} g(T_i)} + g(T_n)\delta_n \right),$$

Case 2:

$$\mathcal{V} = \left(\sum_{i=1}^{n-1} \frac{\delta_i g(T_i)}{n-i+1+n\overline{\lambda} g(T_i)} + g(T_n)\delta_n, \quad g(T_n)\delta_n \right),$$

Case 3:

$$\mathcal{V} = \left(\sum_{i=1}^{n-1} \frac{\delta_i g(T_i)}{n-i+1+n\overline{\lambda} g(T_i)} + g(T_n)\delta_n, \quad \sum_{i=1}^{n-1} \frac{\delta_i g(T_i)}{n-i+1+n\underline{\lambda} g(T_i)} + g(T_n)\delta_n \right).$$

□

3. Asymptotic properties

Now we study the large sample behavior of the empirical likelihood under constraint (2.6). First, we present a lemma about the large sample behavior of the solution λ of (2.10).

Lemma 2 *Suppose $g(t)$ is a left continuous function and*

$$0 < \int \frac{|g(x)|^m d\Lambda_0(x)}{(1-F_0(x))(1-G_0(x))} < \infty, \quad m = 1, 2.$$

Then $\theta_0 = \int g(t) d\Lambda_0(t)$, is feasible with probability approaching 1 as $n \to \infty$, and the solution λ of (2.10) with $\theta = \theta_0$ satisfy

$$n\lambda^2 \xrightarrow{\mathcal{D}} \chi^2_{(1)} \left(\int \frac{g^2(x) d\Lambda_0(x)}{(1-F_0(x))(1-G_0(x))} \right)^{-1} \qquad as\ n \to \infty\ .$$

Proof: See appendix. □

Next we define the empirical likelihood ratio in terms of hazard for the constraint (2.7) as

$$\mathcal{ALR}(\theta) = \frac{\sup\{AL(\Lambda) | \Lambda \ll \hat\Lambda_{NA},\ \text{and}\ \Lambda\ \text{satisfy (2.7)}\}}{AL(\hat\Lambda_{NA})}\ .$$

By Theorem 1, $\mathcal{ALR}(\theta)$ can be computed, when the constraint is feasible, by using W_i defined there and the known property of $\hat\Lambda_{NA}$: $\Delta\hat\Lambda_{NA}(T_i) = \delta_i/(n-i+1)$.

Theorem 2 *Let $(T_1, \delta_1), \cdots, (T_n, \delta_n)$ be n pairs of random variables as defined in (2.1). Suppose g is a left continuous function and*

$$0 < \int \frac{|g(x)|^m}{(1-F_0(x))(1-G_0(x))} d\Lambda_0(x) < \infty, \qquad m = 1, 2.$$

Then, $\theta_0 = \int g(t) d\Lambda_0(t)$ will be a feasible value with probability approaching one as $n \to \infty$ and

$$-2 \log \mathcal{ALR}(\theta_0) \xrightarrow{\mathcal{D}} \chi^2_{(1)} \qquad as\ n \to \infty\ .$$

Proof: In view of Lemma 2, we need only to proof the last claim: $-2\log \mathcal{ALR}(\theta_0) \xrightarrow{\mathcal{D}} \chi^2_{(1)}$ as $n \to \infty$. To this end, define

$$Z_i = \delta_i g(T_i) \frac{1}{\frac{n-i+1}{n}} \quad \text{for } i = 1, 2, \cdots, n, \tag{3.1}$$

and consider

$$\begin{aligned} & -2 \log \mathcal{ALR}(\theta_0) \\ = \ & 2 \left[\sum_{i=1}^n \delta_i \log \Delta\Lambda_{NA}(T_i) - \sum_{i=1}^n (n-i+1) \Delta\Lambda_{NA}(T_i) \right] - 2 \left[\sum_{i=1}^n \delta_i \log \Delta\Lambda_{NA}(T_i) \right] \end{aligned}$$

$$
\begin{aligned}
&+2\left[\sum_{i=1}^{n-1}\delta_i\log(1+\lambda Z_i)+\sum_{i=1}^{n-1}\frac{(n-i+1)\Delta\Lambda_{NA}(T_i)}{1+\lambda Z_i}+\Delta\Lambda_{NA}(T_n)\right]\\
=\;& -2\sum_{i=1}^{n}\delta_i+2\sum_{i=1}^{n-1}\delta_i\log(1+\lambda Z_i)+2\sum_{i=1}^{n-1}\frac{\delta_i}{1+\lambda Z_i}+2\delta_n\\
=\;& -2\sum_{i=1}^{n}\delta_i+2\sum_{i=1}^{n-1}\delta_i\log(1+\lambda Z_i)+2\sum_{i=1}^{n-1}\delta_i-2\sum_{i=1}^{n-1}\frac{\delta_i\lambda Z_i}{1+\lambda Z_i}+2\delta_n\\
=\;& 2\sum_{i=1}^{n-1}\delta_i\log(1+\lambda Z_i)-2\sum_{i=1}^{n-1}\delta_i\lambda Z_i+2\sum_{i=1}^{n-1}\frac{\delta_i\lambda^2 Z_i^2}{1+\lambda Z_i}\;. \qquad (3.2)
\end{aligned}
$$

Notice $\max_{1\le i\le n}|\lambda Z_i| = O_p(n^{-1/2})\max_{1\le i\le n}|Z_i|$ by Lemma 2. Now use Lemma A2 with $h = g/\sqrt{(1-F)(1-G)}$ and Zhou (1991) we have

$$
\begin{aligned}
\max_{1\le i\le n}|Z_i| \;&\le\; \max_{1\le i\le n}\frac{\delta_i|g(T_i)|}{(1-F_0(T_i))(1-G_0(T_i))}\max_{1\le i\le n}\frac{(1-F_0(T_i))(1-G_0(T_i))}{(n-i+1)/n}\\
&=\; o_p(n^{1/2})O_p(1)=o_p(n^{1/2})\;. \qquad (3.3)
\end{aligned}
$$

Thus $\max_{1\le i\le n-1}|\lambda Z_i| = O_p(n^{-1/2})o_p(n^{1/2}) = o_p(1)$ and we may expand

$$
\log(1+\lambda Z_i)=\lambda Z_i-\frac{1}{2}\lambda^2Z_i^2+O_p(\lambda^3)Z_i^3. \qquad (3.4)
$$

Substituting (3.4) in the expression of $-2\log\mathcal{ALR}(\theta_0)$, we have

$$
\begin{aligned}
&-2\log\mathcal{ALR}(\theta_0)\\
=\;& 2\sum_{i=1}^{n-1}\delta_i\lambda Z_i-\sum_{i=1}^{n-1}\delta_i\lambda^2Z_i^2+O_p(\lambda^3)\sum_{i=1}^{n-1}Z_i^3-2\sum_{i=1}^{n-1}\delta_i\lambda Z_i+2\sum_{i=1}^{n-1}\delta_i\lambda^2Z_i^2-2\sum_{i=1}^{n-1}\frac{\delta_i\lambda^3Z_i^3}{1+\lambda Z_i}\\
=\;& \lambda^2\sum_{i=1}^{n-1}\delta_iZ_i^2+O_p(\lambda^3)\sum_{i=1}^{n-1}Z_i^3-2\lambda^3\sum_{i=1}^{n-1}\frac{\delta_iZ_i^3}{1+\lambda Z_i} \qquad (3.5)
\end{aligned}
$$

where, as $n\to\infty$

$$
\left|O_p(\lambda^3)\sum_{i=1}^{n-1}Z_i^3\right|\le|O_p(n^{-1/2})||o_p(n^{1/2})|\times\frac{1}{n}\sum_{i=1}^{n}Z_i^2,
$$

and, notice $\delta_iZ_i^3 = Z_i^3$,

$$
2\lambda^3\sum_{i=1}^{n-1}\frac{\delta_iZ_i^3}{1+\lambda Z_i}\le|O_p(n^{-1/2})||o_p(n^{1/2})|\times\frac{1}{n}\sum_{i=1}^{n}Z_i^2.
$$

By Lemma A3 and (3.3) we have

$$
\mathrm{Plim}\,\frac{1}{n}\sum_{i=1}^{n}Z_i^2=\mathrm{Plim}\,\frac{1}{n}\sum_{i=1}^{n-1}\delta_iZ_i^2=\mathrm{Plim}\,\frac{1}{n}\sum_{i=1}^{n-1}Z_i^2=\int\frac{g^2(x)\;d\Lambda_0(x)}{(1-F_0(x))(1-G_0(x))}<\infty,
$$

where Plim denotes the limit in probability as $n \to \infty$. Therefore the last two terms in (3.5) are negligible. As for the first term there, we see that it converges to a $\chi^2_{(1)}$ distribution in view of Lemma 2, Lemma A.3 and Slutsky theorem. Thus we have as $n \to \infty$

$$-2 \log \mathcal{ALR}(\theta_0) \xrightarrow{\mathcal{D}} \chi^2_{(1)}. \qquad \Box$$

4. Comparison of two versions of likelihood

In this section we examine the difference between the 2 versions of the likelihood EL and AL as defined in (2.4) and (2.5). We shall proof that if we replace AL in the Theorem 2 by EL and everything else remain the same, the likelihood ratio statistic $-2\log \mathcal{ELR}(\theta_0)$, still converges to $\chi^2_{(1)}$ as $n \to \infty$.

Define

$$\mathcal{ELR}(\theta) = \frac{EL(\Lambda^\star)}{EL(\hat\Lambda_{NA})}$$

where $\Lambda^\star$ is given by the jumps W_i defined in Theorem 1.

Theorem 3 *Suppose all the conditions of Theorem 2 holds, then*

$$-2 \log \mathcal{ELR}(\theta_0) \xrightarrow{\mathcal{D}} \chi^2_{(1)} \quad \text{as} \quad n \to \infty.$$

Proof: We shall proof that the 2 likelihood ratio statistics are asymptotically equivalent in the sense that their difference goes to zero in probability.

By (3.2) we have

$$-2 \log \mathcal{ALR}(\theta_0) = 2\sum_{i=1}^{n-1} \delta_i \log(1+\lambda Z_i) - 2\sum_{i=1}^{n-1} \frac{\delta_i \lambda Z_i}{1+\lambda Z_i},$$

where Z_i is defined as in (3.1). On the other hand, we also have

$$\begin{aligned} -2 \log \mathcal{ELR}(\theta_0) &= 2\sum_{i=1}^{n-1} \delta_i \log(1+\lambda Z_i) + 2\sum_{i=1}^{n-1}(n-i+1-\delta_i)\log(1-\Delta\hat\Lambda_{NA}(T_i)) \\ &\quad -2\sum_{i=1}^{n-1}(n-i+1-\delta_i)\log\left(1-\Delta\hat\Lambda_{NA}(T_i)\frac{1}{1+\lambda Z_i}\right). \end{aligned}$$

Observe

$$\log\left(1-\Delta\hat\Lambda_{NA}(T_i)\frac{1}{1+\lambda Z_i}\right) = \log\left(1-\Delta\hat\Lambda_{NA}(T_i)+\Delta\hat\Lambda_{NA}(T_i)\frac{\lambda Z_i}{1+\lambda Z_i}\right).$$

By the same reason as in (3.3), (3.4) we may expand

$$\log\left(1-\Delta\hat\Lambda_{NA}(T_i)\frac{1}{1+\lambda Z_i}\right) = \log\left(1-\Delta\hat\Lambda_{NA}(T_i)+\Delta\hat\Lambda_{NA}(T_i)\frac{\lambda Z_i}{1+\lambda Z_i}\right)$$

$$= \log(1 - \Delta\hat{\Lambda}_{NA}(T_i)) + \frac{\Delta\hat{\Lambda}_{NA}(T_i)}{1 - \Delta\hat{\Lambda}_{NA}(T_i)} \times \frac{\lambda Z_i}{1 + \lambda Z_i} - \left(\frac{\Delta\hat{\Lambda}_{NA}(T_i)}{1 - \Delta\hat{\Lambda}_{NA}(T_i)}\right)^2 \eta_i^2$$

$$= \log(1 - \Delta\hat{\Lambda}_{NA}(T_i)) + \frac{\delta_i}{n - i + 1 - \delta_i} \times \frac{\lambda Z_i}{1 + \lambda Z_i} - \left(\frac{\delta_i}{n - i + 1 - \delta_i}\right)^2 \eta_i^2 \qquad (4.1)$$

where $|\eta_i| \le |\frac{\lambda Z_i}{1 + \lambda Z_i}|$.

Substituting (4.1) in the expression of $-2\log \mathcal{ELR}(\theta_0)$, we obtain

$$-2\log \mathcal{ELR}(\theta_0) = 2\sum_{i=1}^{n-1} \delta_i \log(1 + \lambda Z_i) - 2\sum_{i=1}^{n-1} \frac{\delta_i \lambda Z_i}{1 + \lambda Z_i} + 2\sum_{i=1}^{n-1} \eta_i^2 \frac{1}{n - i + 1 - \delta_i}.$$

Therefore

$$-2\log \mathcal{ELR}(\theta_0) + 2\log \mathcal{ALR}(\theta_0) = 2\sum_{i=1}^{n-1} \eta_i^2 \frac{1}{n - i + 1 - \delta_i},$$

where

$$0 \le \sum_{i=1}^{n-1} \eta_i^2 \frac{1}{n - i + 1 - \delta_i} \quad \le \quad \lambda^2 \sum_{i=1}^{n-1} \frac{Z_i^2}{n - i + 1 - \delta_i}.$$

By Lemma 2 and Lemma A3 we have

$$n\lambda^2 \frac{1}{n} \sum_{i=1}^{n-1} \frac{Z_i^2}{n - i + 1 - \delta_i} = O_p(1) o_p(1) = o_p(1).$$

Therefore

$$-2\log \mathcal{ELR}(\theta_0) + 2\log \mathcal{ALR}(\theta_0) \xrightarrow{\mathrm{P}} 0 \quad \text{as} \quad n \to \infty.$$

In view of Theorem 2, we have

$$-2\log \mathcal{ELR}(\theta_0) \xrightarrow{\mathcal{D}} \chi^2_{(1)} \quad \text{as} \quad n \to \infty.$$

□

5. Stochastic constraints and implicit constraints

5.1 Stochastic constraints

Some applications, specifically one sample log rank type tests, (cf. Andersen et. al. 1993 p. 334), mandate a random weight function $g(t) = g_n(t)$ in the constraint. Also, in order to obtain mean from the integration of cumulative hazard, we need to let $g(t) = g_n(t) = t[1 - \hat{F}_n(t-)]$, again a random function. To accommodate this, we allow the function g to depend on the sample (of size n) but require that it is a predictable function with respect to the filtration that makes $\hat{\Lambda}_{NA}(t) - \Lambda(t)$ a martingale. For example the filtration

$$\mathcal{F}_t = \sigma\left\{T_k I_{[T_k \le t]};\ \delta_k I_{[T_k \le t]};\ k = 1, 2, \ldots, n\right\} . \qquad (5.1)$$

Furthermore we require that for some nonrandom left continuous function $g(t)$, we have

$$\sup_{t \le T_n} |g_n(t) - g(t)| = o_p(1) \quad \text{and} \quad \sup_{1 \le i \le n} \left| \frac{g_n(T_i)}{g(T_i)} \right| = O_p(1) \quad \text{as} \quad n \to \infty \ . \tag{5.2}$$

The weight functions for the one sample log rank test and mean can be shown to satisfy these requirements. The stochastic version of the constraint is therefore

$$\int g_n(t) d\Lambda(t) = \theta_n \ . \tag{5.3}$$

The θ value may also depend on n. For example if we are testing the hypothesis $H_0 : \Lambda \equiv \Lambda_0$ then we should take $\theta_n = \int g_n(t) d\Lambda_0(t)$.

The empirical likelihood ratio statistics for the stochastic constraint is defined as

$$-2 \log \mathcal{ALR}_s(\theta_n) = \frac{\sup\{AL(\Lambda) | \Lambda \ll \hat{\Lambda}_{NA} \ \text{ and } \ \Lambda \ \text{ satisfy } (5.3)\}}{AL(\hat{\Lambda}_{NA})}$$

where the numerator of the ratio can be computed similarly as in Theorem 1 with $g_n(t)$ and θ_n replacing $g(t)$ and θ there.

Theorem 4 *Let $(T_1, \delta_1), \cdots, (T_n, \delta_n)$ be n pairs of random variables as defined in (2.1). Suppose $g_n(t)$ is a sequence of predictable functions with respect to the filtration (5.1) and satisfy (5.2). Also assume*

$$0 < \int \frac{|g(x)|^m}{(1 - F_0(x))(1 - G_0(x))} d\Lambda_0(x) < \infty, \qquad m = 1, 2.$$

Then, $\theta_n^0 = \int g_n(t) d\Lambda_0(t)$ will be a feasible value with probability approaching one as $n \to \infty$ and

$$-2 \log \mathcal{ALR}_s(\theta_n^0) \xrightarrow{\mathcal{D}} \chi^2_{(1)} \qquad \text{as } n \to \infty \ .$$

5.2 Implicit constraints

For the implicit functional constraint, we require that (i)

$$\int g(t, \theta) d\Lambda(t) \tag{5.4}$$

is monotone in θ for any given cumulative hazard function Λ, and (ii)

$$\int g(t, \theta) d\Lambda_0(t) = C \tag{5.5}$$

uniquely defines the parameter θ_0.

The likelihood ratio in this case is formed similarly. For given θ we first solve the the following equation to get λ.

$$\sum_{i=1}^{n-1} g(T_i, \theta) \frac{\delta_i}{n - i + 1} \times \frac{1}{1 + \lambda \frac{\delta_i g(T_i, \theta)}{(n-i+1)/n}} + g(T_n, \theta)\delta_n = C \tag{5.6}$$

where C is a given constant. Then $\mathcal{ALR}_i(\theta)$ is defined as the ratio of two AL's with the numerator computed as (2.8) with

$$w_i = \frac{\delta_i}{n-i+1} \times \frac{1}{1+\lambda \frac{\delta_i g(T_i,\theta)}{(n-i+1)/n}}$$

and the denominator computed via (2.8) with $w_i = \delta_i/(n-i+1)$ as before.

Theorem 5 *Let $(T_1, \delta_1), \cdots, (T_n, \delta_n)$ be n pairs of random variables as defined in (2.1). Suppose $g(t,\theta)$ is a function satisfy (5.4) and (5.5). Also assume*

$$0 < \int \frac{|g(x,\theta)|^m}{(1-F_0(x))(1-G_0(x))} d\Lambda_0(x) < \infty, \qquad m = 1, 2.$$

Then,

$$-2\log \mathcal{ALR}_i(\theta_0) \xrightarrow{\mathcal{D}} \chi^2_{(1)} \qquad \text{as } n \to \infty \ .$$

6. Simulations and examples

Notice our results in section 2 reduces the computation of the maximization to a single parameter λ. All we need to solve is the constraint equation for λ and it is monotone decreasing in λ. An Splus function that computes the empirical likelihood ratio described in this paper is available from the second author.

Example 1: For a small sample simulation, we generate the censored survival data from the following setting:

Survival Time Distribution:	$F_0(t) = 1 - e^{-t}$
Censoring Distribution:	$G_0(t) = 1 - e^{-0.35t}$
Cumulative Hazard Function:	$\Lambda_0(t) = t$
Sample Size:	$n = 20$
g :	$g(t) = e^{-t}$
parameter θ_0 :	$\theta_0 = \int_0^\infty g(t) d\Lambda_0(t) = 1$

The 95% confidence interval for θ_0 can be constructed as

$$\{\theta | -2\log \mathcal{ALR}(\theta) \le 3.84\} \ .$$

Each time we compute $-2\log \mathcal{ALR}(\theta = 1)$ and check to see if it is less then 3.84 (inside the interval). In 1000 independent such runs we recorded 947 coverage for intervals that suppose to have an asymptotical nominal coverage probability of 95%. For the same data the Wald confidence interval based on Nelson-Aalen type estimator results 920 coverage out of the 1000 runs.

Example 2: For a concrete example we took the data of Remission Times for Solid Tumor Patients ($n = 10$). These are slightly modified (break tie) version of Lee (1992, example 4.2): 3, 6.5, 6.51, 10, 12, 15, 8.4+, 4+, 5.7+, and 10+ .

Suppose we are interested in getting a 95% confidence interval for the cumulative hazard at the time $t = 9.8$, $\Lambda_0(9.8)$. Hence $\theta_0 = \Lambda_0(9.8)$. In this case the function g is an indicator function: $g(t) = I_{[t \leq 9.8]}$.

The 95% confidence interval using empirical likelihood ratio, $-2 \log \mathcal{ALR}$, for $\Lambda_0(9.8)$ is $(0.10024, 1.0917)$. On the other hand, the Wald confidence interval based on the Nelson-Aalen estimator and Greenwood's formula is $(-0.063, 0.882)$. Since the cumulative hazard function is nonnegative, this shows that the empirical likelihood ratio based confidence interval inherit some of the advantage from its parametric cousin.

Example 3: For the implicit function example we shall look at the data of Australian AIDS patients. The description of the data and some analysis can be found in Venables and Ripley (1994). We shall took the 1780 cases from the State of New South Wales and ignore other covariates, i.e. treat the 1780 cases as i.i.d. observations from one population.

The implicit function we illustrate here is the median. Since the median may not be uniquely defined for discrete distribution like the empirical distributions, some smoothing or other modification may be needed, particularly for small sample sizes. However, those modification will become negaligible for large samples. We shall discuss the discrete distribution in another paper and ignore the discreteness here in this example in view of its sample size.

Another aspect of the AIDS data is that it has a lot of ties in the observations. Since our formular developed in this paper assumes no ties in the data, we shall break the ties by subtracting a small amount (0.00001) to the successive observations. This is equivalent to assume that the survival time of AIDS patient is a continuous random variable, and ties in the data are due to rounding (to the nearest day). We therefore suppose the distribution F_0 is continuous and median is uniquely defined for F_0. We shall took $g(t, \theta) = I_{[t \leq \theta]}$ and constraint $\int g(t, \theta) d\Lambda(t) = \log 2$.

The 95% confidence interval (434.8, 492.8) for the median of AIDS survival data is obtained as

$$\{\theta | -2 \log \mathcal{ALR}_i(\theta) < 3.84\}$$

with the constraint $\int g(t, \theta) d\Lambda(t) = \log 2$. The .8 in the confidence interval is due to the addition of 0.9 to the original data by Venables and Ripley and my subtraction of a small amount to break ties.

7. Appendix

Lemma A1: *For any random variable Y, if $E|Y|^k < \infty$ then for an i.i.d. sample $Y_1, Y_2, \cdots, Y_n$*

that have same distribution as Y, we have

$$\max_{1\le i\le n} |Y_i| = o(n^{1/k}) \quad a.s.$$

Proof: See Chow and Teicher 1980 p131, problem # 8. □

Lemma A2: *Let $(T_1,\delta_1),\cdots,(T_n,\delta_n)$ be n i.i.d. pairs of random variables, where each (T_i,δ_i) is defined by (2.1). Let also $T_n^\star = \max_{1\le i\le n} T_i$. If $\int h^2(x)d\Lambda_0(x) < \infty$, then*

$$\max_{1\le i\le n} \frac{\delta_i |h(T_i)|}{\sqrt{(1-F_0(T_i))(1-G_0(T_i))}} = o(n^{1/2}) \quad a.s. \qquad and \qquad \delta_n^\star h(T_n^\star) = o_p(1),$$

where $\delta_n^\star$ is the indicator function corresponding to $T_n^\star$.

Proof: Since $\int h^2(x)d\Lambda_0(x) < \infty$, we have

$$\mathrm{E}_{F_0,G_0} \frac{\delta_i h^2(T_i)}{(1-F_0(T_i))(1-G_0(T_i))} = \int h^2(x)d\Lambda_0(x) < \infty.$$

Therefore, by Lemma A1, we have

$$\max_{1\le i\le n} \frac{\delta_i |h(T_i)|}{\sqrt{(1-F_0(T_i))(1-G_0(T_i))}} = o(n^{1/2}), \tag{A.1}$$

with probability 1 as $n\to\infty$.

The fact that

$$\frac{\delta_n^\star |h(T_n^\star)|}{\sqrt{(1-F_0(T_n^\star))(1-G_0(T_n^\star))}} \le \max_{1\le i\le n} \frac{\delta_i |h(T_i)|}{\sqrt{(1-F_0(T_i))(1-G_0(T_i))}}$$

implies

$$\frac{\delta_n^\star |h(T_n^\star)|}{\sqrt{(1-F_0(T_n^\star))(1-G_0(T_n^\star))}} = o(n^{1/2}), \tag{A.2}$$

with probability 1 as $n\to\infty$.

Let $H_0(t)$ be the distribution function of T_i, where $T_i = \min(X_i, C_i)$, then $1-H_0(t) = (1-F_0(t))(1-G_0(t))$. If we can show

$$1 - H_0(T_n^\star) = O_p(n^{-1}), \tag{A.3}$$

or

$$\sqrt{(1-F_0(T_n^\star))(1-G_0(T_n^\star))} = O_p(n^{-1/2}),$$

then it follows from (A.2) that $\delta_n^\star h(T_n^\star) = o_p(1)$.

Now we show $1-H_0(T_n^\star)=O_p(n^{-1})$. For any $\epsilon>0$, there exists $M_0>0$ such that $\exp(-M_0)<\epsilon$. For $M>M_0$ consider

$$\begin{aligned}
P\left(\frac{1-H_0(T_n^\star)}{n^{-1}}>M\right) &= P\left(\frac{1-\max\limits_{1\le i\le n} H_0(T_i)}{n^{-1}}>M\right)\\
&= P\left(\max_{1\le i\le n} H_0(T_i)<(1-n^{-1}\times M)\right)\\
&= \left(1-\frac{M}{n}\right)^n \le \exp(-M)<\epsilon.
\end{aligned}$$

Therefore $1-H_0(T_n^\star)=O_p(n^{-1})$. □

Lemma A3 *Under the assumptions of Theorem 2, we have, for Z_i defined in (3.1),*

$$\frac{1}{n}\sum_{i=1}^n Z_i^2=\sum_{i=1}^n\frac{\delta_i g^2(T_i)n}{(n-i+1)^2}=\int\frac{g^2(t)}{Y(t)/n}d\hat\Lambda_{NA}(t)\xrightarrow{\mathrm{P}}\int\frac{g^2 d\Lambda(t)}{(1-F)(1-G)} \tag{A.4}$$

and

$$\frac{1}{n}\sum_{i=1}^{n-1}\frac{Z_i^2}{n-i}=\int\frac{I_{[Y(t)>1]}g^2(t)}{(Y(t)-1)Y(t)/n}d\hat\Lambda_{NA}(t)\xrightarrow{\mathrm{P}}0 \qquad as\ \ n\to\infty\ , \tag{A.5}$$

where $Y(t)=\sum I_{[T_i\ge t]}$.

Proof: For (A.5), use Lenglart's inequality on the integral to switch to a similar integral except with respect to $\Lambda(t)$, and then use uniform convergence of empirical distributions to finish the proof. The proof of (A.4) is similar. □

Lemma A4 *Under the assumptions of Theorem 2, we have, for Z_i defined in (3.1),*

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n Z_i-\theta_0\right)=\sqrt{n}\left(\sum_{i=1}^n g(T_i)\Delta\hat\Lambda_{NA}(T_i)-\theta_0\right)\xrightarrow{\mathcal{D}}N(0,\sigma_\Lambda^2(g))\ ,$$

where $\sigma_\Lambda^2(g)=\int\frac{g^2(x)d\Lambda_0(x)}{(1-F_0(x))(1-G_0(x))}$ *and* $\theta_0=\int g(t)d\Lambda_0(t)$.

Proof: Notice the summation can be written as an integral

$$\sum_{i=1}^n g(T_i)\Delta\hat\Lambda_{NA}(T_i)-\theta_0=\int g(t)d[\hat\Lambda_{NA}(t)-\Lambda_0(t)]\ .$$

Now counting process and martingale argument similar to Andersen et. al. (1993) chapter 4 can be used to analyze the integral (since $g(\cdot)$ is left continuous, it is predictable). An application of martingale central limit theorem will finish the proof. □

Proof of Lemma 2: First we notice that if we set $\lambda=0$ in the constraint equation (2.10), the jumps W_i reduce to those of Nelson-Aalen estimator, implying that $\theta=\hat\theta_n=\int g(t)d\hat\Lambda_{NA}(t)$ is always a feasible value, i.e. $\hat\theta_n\in\mathcal{V}$.

On the other hand, notice that the derivative

$$\frac{\partial l(\lambda)}{\partial \lambda} = -\sum_{i=1}^{n-1} \frac{\delta_i g(T_i)}{n-i+1} \times \frac{Z_i}{[1+\lambda Z_i]^2} \ ,$$

when evaluated at $\lambda = 0$ we have

$$\frac{\partial l(\lambda)}{\partial \lambda}\Big|_{\lambda=0} = -\frac{1}{n}\sum_{i=1}^{n-1} Z_i^2 \ .$$

By Lemma A2 and A3 it converges (in fact almost surely) to

$$-\int \frac{g^2(x) d\Lambda_0(x)}{(1-F_0(x))(1-G_0(x))} \ .$$

The integral is positive by assumption. Therefore the derivative of $l(\lambda)$ at $\lambda = 0$ will be bounded away from zero, in fact $l'(0) \leq \eta < 0$ at least for large n.

This implies that if the legitimate value of λ, $\mathcal{J}$, covers at least an open interval of length $\frac{1}{o_p(n^{1/2})}$ for all large n centered at 0, then the feasible value of θ, $\mathcal{V}$, will also contain an open interval of length $\frac{1}{o_p(n^{1/2})}$ centered at $\hat{\theta}_n$. Since $\hat{\theta}_n - \theta_0 = O_p(n^{-1/2})$, this will ensure that θ_0 will be in $\mathcal{V}$, i.e. a feasible value, for large n .

The fact that the legitimate value of λ, $\mathcal{J}$, covers at least an open interval of length $\frac{1}{o_p(n^{1/2})}$ for all large n centered at zero can easily be seen from the definition of $\mathcal{J}$ by noticing that

$$\frac{1}{|\underline{\lambda}|} = o_p(n^{1/2})$$

which can be proved similar to (3.3). The argument for $\overline{\lambda}$ is the same.

Now we turn to the asymptotic distribution of the solution λ when $\theta = \theta_0$. The first step is to show that $\lambda = O_p(n^{-1/2})$ where λ is the solution of (2.10) so that we can use expansion later.

Recall the definition of Z_i in (3.1) and its bound (3.3)

$$\max|Z_i| = \max_{1 \leq i \leq n} |Z_i| = o_p(n^{1/2}) \ .$$

We rewrite (2.10) in terms of Z_i's as follows

$$\begin{aligned} 0 &= |l(\lambda)| \\ &= \left|\theta_0 - \frac{1}{n}\sum_{i=1}^{n-1} \frac{Z_i}{1+\lambda Z_i} - \frac{1}{n} Z_n\right| \\ &= \left|\theta_0 - \frac{1}{n}\sum_{i=1}^{n-1} Z_i + \frac{\lambda}{n}\sum_{i=1}^{n-1} \frac{Z_i^2}{1+\lambda Z_i} - \frac{1}{n} Z_n\right| \\ &= \left|(\theta_0 - \frac{1}{n}\sum_{i=1}^{n} Z_i) + \frac{\lambda}{n}\sum_{i=1}^{n-1} \frac{Z_i^2}{1+\lambda Z_i}\right| \\ &\geq \frac{|\lambda|}{1+|\lambda| \max|Z_i|} \frac{1}{n}\sum_{i=1}^{n-1} Z_i^2 - \left|\theta_0 - \frac{1}{n}\sum_{i=1}^{n} Z_i\right| . \end{aligned} \tag{A.6}$$

The second term of (A.6) is $O_p(n^{-1/2})$ by Lemma A4. Now we consider the first term of (A.6). Since

$$\frac{1}{n}\sum_{i=1}^{n-1} Z_i^2 = \frac{1}{n}\sum_{i=1}^{n} Z_i^2 - \frac{1}{n}Z_n^2 \ ,$$

and by (3.3) we have $\frac{1}{n}Z_n^2 = o_p(1)$. Hence by Lemma A3

$$\frac{1}{n}\sum_{i=1}^{n-1} Z_i^2 \xrightarrow{\mathrm{P}} \int \frac{g^2(x)}{(1-F_0(x))(1-G_0(x))} d\Lambda_0(x), \tag{A.7}$$

and it follows that

$$\frac{|\lambda|}{1+|\lambda|\max|Z_i|} = O_p(n^{-1/2}),$$

which implies that

$$\lambda = O_p(n^{-1/2}). \tag{A.8}$$

Expanding (2.10), we obtain

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^{n} Z_i - \theta_0 - \frac{\lambda}{n}\sum_{i=1}^{n-1}\frac{Z_i^2}{1+\lambda Z_i} \\ &= \frac{1}{n}\sum_{i=1}^{n} Z_i - \theta_0 - \frac{\lambda}{n}\sum_{i=1}^{n-1} Z_i^2 + \frac{\lambda^2}{n}\sum_{i=1}^{n-1}\frac{Z_i^3}{1+\lambda Z_i}. \end{aligned} \tag{A.9}$$

The last term in (A.9) is bounded by (Lemma 2, (3.3) and Lemma A3)

$$\lambda^2\frac{1}{n}\sum_{i=1}^{n-1}|Z_i^3| \le \lambda^2 \max|Z_i| \frac{1}{n}\sum_{i=1}^{n-1} Z_i^2 = O_p(n^{-1})o_p(n^{1/2})O_p(1) = o_p(n^{-1/2}).$$

Therefore we get an expression of λ as follows

$$\lambda = \frac{\frac{1}{n}\sum_{i=1}^{n} Z_i - \theta_0}{\frac{1}{n}\sum_{i=1}^{n-1} Z_i^2} + o_p(n^{-1/2}). \tag{A.10}$$

By Lemma A4, as $n \to \infty$

$$\frac{1}{n}\sum_{i=1}^{n} Z_i - \theta_0 = \sqrt{n}\left(\sum_{i=1}^{n} g(T_i)\Delta\hat{\Lambda}_{NA}(T_i) - \theta_0\right) \xrightarrow{\mathcal{D}} N(0, \sigma_\Lambda^2(g)) \ .$$

Thus by Slutsky theorem and (A.7), as $n \to \infty$

$$n\lambda^2 \xrightarrow{\mathcal{D}} \chi^2_{(1)}\left(\int \frac{g^2(x)d\Lambda_0(x)}{(1-F_0(x))(1-G_0(x))}\right)^{-1}. \tag{A.11}$$

$\square$

References

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N., (1993). *Statistical Models Based on Counting Processes*, Springer Verlag, New York.

Chow, Y. S. and Teicher, H. (1980). *Probability Theory* Springer, New York.

Efron, B. and Johnstone, I. M. (1990). Fisher's information in terms of the hazard rate. *Ann. Statist.* **18**, 38-62.

Gill, R. D. (1989). Non- and Semi-Parametric Maximum Likelihood Estimators and the Von-Mises Method, Part I, *Scand. J. Statist.*, **16**, 97-128.

Lee, E. T. (1992). *Statistical Methods for Survival Data Analysis.* Wiley, New York.

Li, G. (1995). On Nonparametric Likelihood Ratio Estimation of Survival Probabilities for Censored Data. *Statistics and Probability Letters*, **25**, 95-104.

Murphy, S. A. (1995). Likelihood Ratio Based Confidence Intervals in Survival Analysis. *Journal of the American Statistical Association*, **90**, 1399-1405.

Murphy, S. A. and Van der Vaart, A. W. (1997). Semi-parametric Likelihood Ratio Inference. *Ann. Statist.* **25**, 1471-1509.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237-49.

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

Pan, X. R. and Zhou M. (1997a). Empirical likelihood ratio, one parameter sub-family of distribution functions and censored data. Tech. Report # 360, Department of Statistics, University of Kentucky.

Thomas, D. R. and Grunkemeier, G. L. (1975). Confidence Interval Estimation of Survival Probabilities for Censored Data. *Journal of the American Statistical Association*, **70**, 865-871.

Ying, Z. L. (1992). Minimum Hellinger-type distance estimation for censored data. *Ann. Statist.*, **20**, 1361-1390.

Venables, W.N. and Ripley, B.D. (1994). *Modern Applied Statistics with S-Plus.* Springer, New York.

Zhou, M. (1991). Some Properties of the Kaplan-Meier Estimator for Independent Non-identically Distributed Random Variables. *Ann. Statist.*, **19**, 2266-2274.

Department of Statistics
University of Kentucky
Lexington, KY 40506-0027