USING 1-PARAMETER SUB-FAMILY OF DISTRIBUTIONS IN EMPIRICAL LIKELIHOOD RATIO WITH CENSORED DATA

Xiao-rong Pan and Mai Zhou

University of Kentucky

Abstract

Recently it has been shown that empirical likelihood ratios can be used to form confidence intervals and test hypothesis just like the parametric case. We illustrate here the use of a particular kind of 1-parameter sub-family of distributions in the analysis of empirical likelihood with censored data. This approach not only simplifies the theoretical analysis of the limiting behavior of the empirical likelihood ratio, it also gave us clues for the numerical search of constrained maxima of an empirical likelihood.

Key Words and Phrases: Empirical likelihood ratio, One parameter sub-families of distributions, Censored data.

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 its inverse. 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 its inverse) is difficult to estimate. Even when all 3 are easy to obtain, the LR method still holds some unique advantages: it produces a confidence interval that is invariant under transformations, and is always inside the parameter space. The drawback of the Wilks method is that we need to find the maximum of the likelihood under a constraint and this could be non-trivial.

Recently, Owen (1988, 1990) and others showed that the likelihood ratio method can also be used to produce confidence intervals in some *nonparametric* settings after appropriate modifications. He term this empirical likelihood ratio method. A key feature of his approach is that the maximum of the (empirical) likelihood under the constraint can (almost) be explicitly found. This is no longer the case when either the likelihood or the constraint gets more complicated. For example, when the constraint is not a linear function of distribution function or when the data are censored making the likelihood more complicated. For some recent work on the censored data empirical likelihood ratio method see Li (1995) and Murphy (1995).

In this paper we use the method of 1-parameter sub-family of distributions to deal with the difficult of maximizing the likelihood under a constraint. In essence, this method delays (part of) the task of finding the maximum under a constraint until after the limit ($n \to \infty$) was taken. In the limiting form, the models are smoother and often easier to be maximized. We present below in detail how this works for right censored data with the constraint of a linear functional of the hazard function, but the same method works for many other cases too.

Now we formally describe the type of data and the likelihood we will be considering in this paper. Suppose that $X_1, \cdots, X_n$ are iid nonnegative random variables denoting the lifetimes with a nondegenerate distribution function $F(t) = P(X \leq t)$. Independent of the lifetimes are the censoring times $C_1, \cdots, C_n$ which are iid with a distribution G. Only the censored observations are available to us:

$$T_i = \min(X_i, C_i) \quad ; \qquad \delta_i = I[X_i \leq C_i] \qquad i = 1, 2, \cdots, n. \tag{1.1}$$

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

$$\prod \Delta F(T_i)^{\delta_i} [1 - F(T_i)]^{1-\delta_i} \Delta G(T_i)^{\delta_i} [1 - G(T_i)]^{1-\delta_i} \ .$$

Since we are only concerned in this paper with the inference of X distribution F, we drop the terms involving G in the above. Thus the empirical likelihood pertaining F is

$$EL(F) = \prod_{i=1}^{n} \Delta F(T_i)^{\delta_i} [1 - F(T_i)]^{1-\delta_i} \ . \tag{1.2}$$

When the distributions are continuous, the above likelihoods reduces to the familiar forms

$$\prod [\Delta \Lambda_F(T_i)]^{\delta_i} [\Delta \Lambda_G(T_i)]^{1-\delta_i} [1 - F(T_i)][1 - G(T_i)]$$

and

$$\prod [\Delta \Lambda_F(T_i)]^{\delta_i} [1 - F(T_i)] \ .$$

where $\Lambda_F(\cdot)$ is the cumulative hazard function of X, etc.

But since the distributions here may not be continuous, in fact the nonparametric maximum likelihood estimators: the Kaplan-Meier and the Nelson-Aalen estimators are all purely discrete, we need to be more careful. We begin with a formula linking F and Λ_F that is valid for continuous as well as discrete distributions. The cumulative hazard function $\Lambda(t)$ is defined in this case (see. eg. Andersen, Borgan, Gill and Keiding (1993) p. 92)

$$\Lambda(t) = \Lambda_F(t) = \int_{[0,\ t]} \frac{dF(s)}{1 - F(s-)} \ , \tag{1.3}$$

and it is often more convenient to work with than distribution function when censoring are involved. Thus we have $\Delta F(t) = \Delta\Lambda(t)[1 - F(t-)]$ and $1 - F(t) = \prod_{s \leq t}(1 - \Delta\Lambda(s))$. This leads to the rewriting of (1.2) as

$$EL = \prod (\Delta\Lambda(T_i))^{\delta_i} [1 - F(T_i-)]^{\delta_i} [1 - F(T_i)]^{1-\delta_i} \ , \tag{1.4}$$

and

$$EL(\Lambda) = \prod_{i=1}^{n} (\Delta\Lambda(T_i))^{\delta_i} [\prod_{s<T_i} (1 - \Delta\Lambda(s))]^{\delta_i} [\prod_{s \le T_i} (1 - \Delta\Lambda(s))]^{1-\delta_i}. \tag{1.5}$$

If the distributions (and thus the cumulative hazards) have discrete as well as continuous part then the later two products above should be understood as product integrals. But for discrete distributions or hazards that have possible jumps only at the observed times, T_i, the likelihood is

$$EL(\Lambda) = \prod_{i=1}^{n} (\Delta\Lambda(T_i))^{\delta_i} [\prod_{T_j<T_i} (1 - \Delta\Lambda(T_j))]^{\delta_i} [\prod_{T_j \le T_i} (1 - \Delta\Lambda(T_j))]^{1-\delta_i}. \tag{1.6}$$

See Owen (1988) for a discussion on why we can restrict the support of F on the observed sample. See also Gill (1989) for more detailed discussion on the extension of the likelihood function that covers discrete case. One point made clear by Gill is that even though the true distribution of X may be continuous, we should write the likelihood to include the discrete case.

In the worst case, we may forget about whether EL is a true likelihood or not and only think of EL as just some statistic and using it we can construct a pivotal: $-2 \log$ ratio of maximized EL.

It is well known the unconstrained maximizer of $EL(\Lambda)$ is the Nelson-Aalen estimator:

$$\hat{\Lambda}_{NA}(t) = \sum_{i: T_i \le t} \frac{\delta_i}{R(T_i)}$$

where $R(t)$ is the number at risk at time t: $R(t) = \#\{i | \; T_i \ge t\}$. This estimator is a pure jump function. All of its jump sizes are less then one with the possible exception of the last jump which could be one. For detailed discussion of the cumulative hazard function and Nelson-Aalen estimator, see Andersen, Borgan, Gill and Keiding (1993) Chapter IV.

2. Main Theorems

Since the likelihood (1.6) is now in terms of the cumulative hazard function, we consider a constraint that is a linear functional of the (unknown) cumulative hazard function:

$$\int g(t) d\Lambda(t) = \theta \; . \tag{2.1}$$

Remark: Parameters of the form (2.1) can arise in a Cox model with a time-dependent covariate. In such a model the cumulative hazard up to time τ for a patient with a time-change multiplicative covariate $g(t)$ is $\int_0^\tau g(t) d\Lambda_b(t)$ where $\Lambda_b(t)$ is the baseline hazard. More specifically, if smoking doubles the hazard and a patient that smokes for 10 years and then quit will have a cumulative hazard up to time τ given by the above integral with $g(t) = 1 + I_{[t \le 10]}$.

With a constraint of this form the maximizer of $EL(\Lambda)$ is not easy to find explicitly thus we cannot mimic the approach Owen used. For technical reasons we require

$$g(t) \quad \text{is left continuous, and} \quad \sigma^2_\Lambda(g) = \int \frac{g^2}{(1-F)(1-G)} d\Lambda < \infty. \tag{2.2}$$

In order to make the problem of maximizing $EL(\Lambda)$ under constraint (2.1) easier, we define a 1-parameter sub-family of cumulative hazard functions, and only look for the maximum within this family first. The family we are going to define are all dominated by the Nelson-Aalen estimator, and indexed by λ as

$$\Lambda_\lambda(\cdot) \quad \text{has jump at } t = \begin{cases} \Delta\hat{\Lambda}_{NA}(t)\dfrac{1}{1+\lambda h(t)} & \text{if } \Delta\hat{\Lambda}_{NA}(t) < 1; \\ \Delta\hat{\Lambda}_{NA}(t) & \text{if } \Delta\hat{\Lambda}_{NA}(t) = 1. \end{cases} \tag{2.3}$$

where $h(t)$ is a given function. In the definition we need to single out the case of $\Delta\hat{\Lambda}_{NA}(T_i) = 1$, because any purely discrete cumulative hazard function must have the jump sizes ≤ 1. This is similar to the constrain of "total jump size adds up to one" in the distribution case.

The range of λ needs to be restricted to ensure the above defined Λ_λ is a true cumulative hazard function: i.e. we need to ensure $0 \leq \Delta\Lambda_\lambda < 1$ except when $\Delta\hat{\Lambda}_{NA}(t) = 1$. This is similar to the restriction that all the jumps of a distribution function must be non-negative. This imposes an interval of λ range $\mathcal{J}$ to guarantee that. For detailed definition of $\mathcal{J}$ see appendix. This interval will always contain the value 0 because when $\lambda = 0$, $\Lambda_{\lambda=0}$ falls back to $\hat{\Lambda}_{NA}$, a perfect legitimate hazard function. The function h has the interpretation as the direction Λ_λ passes through $\hat{\Lambda}_{NA}$.

For a given sample if the value θ in (2.1) is too far away from $\int g(t)d\hat{\Lambda}_{NA}(t) = \hat{\theta}$ then the constraint equation may not have a solution (at least not with a legitimate λ value that makes $0 \leq \Delta\Lambda_\lambda < 1$). This is exactly similar to Owen (1990) where the mean μ needs to be inside the convex hull of the sample. Following Owen (1991) we define the likelihood to be zero when θ value is too far away from $\hat{\theta}$.

Finding the maximizer of $EL(\Lambda)$ under constraint (2.1) for *this family of cumulative hazard functions only* is easy: there is only one such cumulative hazard function in this family that satisfy the constraint and thus it is the one that maximizes the EL under the constraint!

We have to chose a λ to satisfy the constraint:

$$\int g d\Lambda_\lambda = \sum_{i=1}^{n} g(T_i)\Delta\Lambda_\lambda(T_i) = \theta \ . \tag{2.4}$$

We require $h(t)g(t) \geq 0 \ a.s. \ F$ to guard against the possibility of multiple solutions of (2.4). This makes $\int g d\Lambda_\lambda$ a monotone function of λ, which can easily be checked by computing its derivative. Let us denote the λ root of (2.4) by λ_n.

Thus the un-constrained maximizer of the likelihood is $\hat{\Lambda}_{NA}$ (or $\Lambda_{\lambda=0}$) and the constrained maximizer (among the family (2.3)) is $\Lambda_{\lambda=\lambda_n}$.

We can thus form the empirical likelihood ratio for this family of cumulative hazard functions by

$$ELR_h(\theta) = \frac{\sup\{EL(\Lambda) | \Lambda \in \text{family (2.3) and satisfy constraint (2.4)}\}}{\sup EL(\Lambda)} = \frac{EL(\Lambda_{\lambda_n})}{EL(\hat{\Lambda}_{NA})} \ . \tag{2.5}$$

We are now ready to state the following Theorem. To clearly present the use of 1-parameter sub-family of distributions and minimize the digress to the technical details we shall assume a continuous distribution for the observations.

Theorem 1 *For the censored data (1.1) with a continuous distribution F, if the constraint is given by (2.1) with $\theta = \theta_0 = \int g d\Lambda_0$ and the given function $g(t)$ satisfy (2.2) and $h(t)$ in (2.3) is another function that is (i) left continuous (ii) $\int h^2 d\Lambda < \infty$ and (iii) $h(t)g(t) \geq 0$ a.s. F, then*

$$-2 \log ELR_h(\theta_0) \xrightarrow{\mathcal{D}} \chi^2_{(1)} \times r_h \qquad \text{as } n \to \infty,$$

where

$$r_h = \frac{\int \frac{g^2 d\Lambda}{(1-F)(1-G)} \int h^2 (1-G)(1-F) d\Lambda}{(\int g h d\Lambda)^2} .$$

Proof: See appendix. □

The constant r_h in the Theorem 1 above can easily be shown by Cauchy-Schwartz inequality to be always ≥ 1 and for a particular choice of h it becomes one. This amounts to find an h that maximize the (limit of) EL under constraint.

Theorem 2 *The constant in Theorem 2, r_h, is always greater then or equal to one. For the choice $h(t) = g(t)/[(1-F)(1-G)]$ the constant $r_h = 1$. Thus for this choice of h we have, under the conditions of Theorem 1,*

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

Proof: We begin by rewrite the 3 integrals in r_h with respect to Λ as expectations:

$$\begin{aligned}\int_0^\infty \frac{g^2(x)}{(1-F_0(x))(1-G_0(x))} d\Lambda_0(x) &= \int_0^\infty \frac{g^2(x)}{(1-F_0(x))^2(1-G_0(x))} dF_0(x) \\ &= \mathrm{E}_{F_0} \frac{g^2(X)}{(1-F_0(X))^2(1-G_0(X))},\end{aligned}$$

$$\begin{aligned}\int_0^\infty h^2(x)(1-G_0(x))(1-F_0(x)) d\Lambda_0(x) &= \int_0^\infty h^2(x)(1-G_0(x)) dF_0(x) \\ &= \mathrm{E}_{F_0} h^2(X)(1-G_0(X)),\end{aligned}$$

and

$$\begin{aligned}\left(\int_0^\infty g(x)h(x) d\Lambda_0(x)\right)^2 &= \left(\int_0^\infty \frac{g(x)h(x)}{1-F_0(x)} dF_0(x)\right)^2 \\ = \left(\mathrm{E}_{F_0} \frac{g(X)h(X)}{1-F_0(X)}\right)^2 &= \left(\mathrm{E}_{F_0} h(X)\sqrt{1-G_0(X)} \frac{g(X)}{(1-F_0(X))\sqrt{1-G_0(X)}}\right)^2 .\end{aligned}$$

By Cauchy-Schwartz inequality we know

$$\left(\mathrm{E}_{F_0}\left[h(X)\sqrt{1-G_0(X)}\frac{g(X)}{(1-F_0(X))\sqrt{1-G_0(X)}}\right]\right)^2$$
$$\leq \mathrm{E}_{F_0}\left[h^2(X)(1-G_0(X))\right]\mathrm{E}_{F_0}\left[\frac{g^2(X)}{(1-F_0(X))^2(1-G_0(X))}\right],$$

this immediately imply $r_h \geq 1$ and the equality holds if

$$h(X)\sqrt{1-G_0(X)} = \frac{g(X)}{(1-F_0(X))\sqrt{1-G_0(X)}}.$$

This implies $h_0 = \frac{g}{(1-F_0)(1-G_0)}$. With the choice $h = h_0$, the constant $r_{h_0} = 1$.

□

Remark: The above theorems actually point out a way to find the maximum of EL under constraint for finite sample sizes: solve (2.4) with $h = h_0$ and compute $EL(\Lambda^{h_0}_{\lambda_n})$ as the constrained maximum, albeit we need to estimate the h_0 in practice. An obvious estimator of h_0 is

$$h_0(t) = \frac{g(t)}{R(t)/n} \ .$$

In fact this is what we shall do in the examples of next section.

The method of 1 parameter sub-family of distributions is also useful in finding the maximum of the likelihood when the constraint is a linear functional of the distribution function $\int g dF = \theta$, and similar theorems can be proved by this approach. In this case the likelihood in terms of distribution, (1.2), is more convenient. The 1 parameter sub-family of distributions should also be defined via the distribution function:

$$F_\lambda \quad \text{is} << \hat{F}_n \text{ and has jump at } t = \Delta\hat{F}_n(t) \times \frac{1}{1+\lambda h(t)} \times \frac{1}{C(\lambda)} \ , \tag{2.6}$$

where $\hat{F}_n$ is the usual nonparametric maximum likelihood estimator of distribution function: the Kaplan-Meier estimator. The factor $1/C(\lambda)$ is the normalizing constant to ensure that the total jump size adds up to 1, and thus is

$$C(\lambda) = \sum_{i=1}^{n} \frac{\Delta\hat{F}_n(T_i)}{1+\lambda h(T_i)} \ .$$

For a given function $h(t)$ and $\hat{F}_n$, at least for small values of λ, the jumps of the so defined function will be positive. It thus define a bona fide distribution function at least for λ in a neighborhood of zero.

For a simple concrete example of the distribution defined by (2.6), suppose there is no censoring and there is no tie in the sample (1.1). In this case the Kaplan-Meier estimator is the usual empirical distribution which has jumps at each observation T_i with jump size 1/n. Further suppose the function

h is such that $h(T_i) = 1$ for $1 \le i \le k$ and $h(T_i) = 0$ for $k+1 \le i \le n$. Then it is not hard to see in this case

$$C(\lambda) = \frac{1}{n}\left[\frac{k}{1+\lambda} + n - k\right],$$

and

$$F_\lambda(\cdot) \quad \text{has jump at } T_i = \begin{cases} \dfrac{1}{k+(n-k)(1+\lambda)} & \text{for } 1 \le i \le k; \\ \dfrac{1}{k/(1+\lambda)+n-k} & \text{for } k+1 \le i \le n. \end{cases}$$

The range of the legitimate λ in this case is $-1 \le \lambda \le \infty$.

Similar theorems to the Theorem 1 and 2 can be shown to hold. The minimizing of the constant (theorem 2) can also be achieved by a simple application of Cauchy-Schwartz inequality when there is no censoring in the data. When there is censoring, however, the minimizing of the constant is not so apparent but with the help of van der Vaart (1991) we can still show that the minimum value is one. We summarize these in the following Corollary. For details see Pan (1997). Notice in the original JSPI version, there was a typo, the role of h and g was exchanged in (2.8). We also provide here some details.

Theorem 3 *For the censored data (1.1) with a continuous distribution F, if the constraint equation is*

$$\int g(t)dF(t) = \theta_0 \tag{2.7}$$

where θ_0 is the true value (i.e. $\theta_0 = \int gdF_0$) and $g(t)$ satisfies certain regularity conditions and $h(t)$ is another function that satisfy same regularity conditions as $g(t)$, then as $n \to \infty$,

$$-2\log ELR_h(\theta_0) = -2\log \frac{\sup\{(1.2)\text{among (2.6) and satisfy}(2.7)\}}{(1.2) \text{ with } F = \text{Kaplan-Meier}} \xrightarrow{\mathcal{D}} \chi^2_{(1)} \times r_h \ ,$$

where

$$r_h = \frac{\text{Asy Var}(\int gd\hat{F}_n) \times \left(\int h^2(1-G)dF + \int \frac{[\int_t^\infty h(s)dF(s)]^2}{1-F(t)}dG(t) - [\int hdF]^2\right)}{\left(\int ghdF\right)^2} . \tag{2.8}$$

Furthermore, the minimum value of the constant r_h over h is one.

Proof: We only sketch the proof for the minimum value of the constant r_h. For details of the convergence in distribution, please see Pan (1997).

First we notice that

$$\frac{\left(\int h^2(1-G)dF + \int \frac{[\int_t^\infty h(s)dF(s)]^2}{1-F(t)}dG(t) - [\int hdF]^2\right)}{\left(\int ghdF\right)^2}$$

is precisely the information defined by van der Vaart (1991), as i_α in his (4.1).

The infimum of i_α over all one-dimentional submodels is called "efficient Fisher information". And in this case (right censored observations), the reciprocal of it is given by the last equation on p. 193 of van der Vaart (1991), (as the lower bound for the asymptotic variance in estimate $\int g dF$):

$$\inf i_\alpha = \frac{1}{||\beta||_F^2} = \left(\int \frac{(R_{\tilde{\chi}_F})^2}{1-G} dF \right)^{-1} .$$

Lastly, we notice that $\int g d\hat{F}_n$ is an efficient estimate and therefore we can easily check

$$\text{Asy Var} \left(\int g d\hat{F}_n \right) = \int \frac{(R_{\tilde{\chi}_F})^2}{1-G} dF \; .$$

Therefore $\inf r_h = 1$. □

3. A Small Example

The computation of the empirical likelihood ratio statistic is quite easy for the one parameter subfamily of distribution functions. All we need to solve is the constrain equation for λ.

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

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

The 95% confidence interval using empirical likelihood ratio ELR for $\Lambda_0(9.8)$ is $(0.106, 0.945)$. On the other hand, the Wald confidence interval based on the Nelson-Aalen estimator and Aalen's formula of variance estimation is $(-0.063, 0.882)$. We use Aalen's formula of variance estimation because this is the recommended one after extended simulation by Klein (1991). This shows that the empirical likelihood ratio based confidence interval inherit some of the advantage from its parametric cousin: shorter and inside the natural parameter space.

5. Appendix

DEFINITION OF THE λ RANGE $\mathcal{J}$ IN (2.3) :

In order to ensure $0 < \Delta\Lambda_\lambda(t) < 1$, we find that the λ range depends on the function $h(\cdot)$ as well as the sample. The following max and min should all be taken in the domain $1 \le i \le n-1$, $h(T_i) \ne 0$ and $\delta_i = 1$. Additional restrictions, if any, are specified individually.

Case 1: all $h(T_i) \ge 0$.

$$\mathcal{J} = \left(\max \frac{\Delta\hat{\Lambda}_{NA}(T_i) - 1}{h(T_i)}, \; \infty \right) .$$

Case 2: all $h(T_i) \le 0$.

$$\mathcal{J} = \left(-\infty, \ \min \frac{\Delta\hat{\Lambda}_{NA}(T_i) - 1}{h(T_i)} \right) .$$

Case 3: $\max h(T_i) > 0 > \min h(T_i)$.

$$\mathcal{J} = \left(\max_{h(T_i)>0} \frac{\Delta\hat{\Lambda}_{NA}(T_i) - 1}{h(T_i)}, \ \min_{h(T_i)<0} \frac{\Delta\hat{\Lambda}_{NA}(T_i) - 1}{h(T_i)} \right) .$$

Lemma A *Let $g(\cdot)$ be a given function that satisfy (2.2) and $h(\cdot)$ be another given left continuous function such that $\int h^2 d\Lambda < \infty$ then as $n \to \infty$,*

$$\sum_{i=1}^{n} h^2(T_i)\Delta\hat{\Lambda}_{NA}(T_i) \xrightarrow{\text{P}} \int h^2(t) d\Lambda_0(t)$$

$$\sum_{i=1}^{n-1} g(T_i)h(T_i)\Delta\hat{\Lambda}_{NA}(T_i) \xrightarrow{\text{P}} \int g(t)h(t) d\Lambda_0(t)$$

and

$$\sqrt{n}\left[\sum g(T_i)\Delta\hat{\Lambda}_{NA}(T_i) - \theta_0\right] \xrightarrow{\mathcal{D}} N(0, \sigma^2_\Lambda(g))$$

where $\sigma^2_\Lambda(g) = \int \frac{g(t)}{[1-F_0(t)][1-G_0(t)]} d\Lambda_0(t)$.

Proof: For the first two limits, rewrite the sums as the integrals against $\hat{\Lambda}_{NA}(t)$ and then use Lenglart inequality to finish the proof.

As for the third limit, notice the left hand side can be written as

$$\sqrt{n} \int g(t) d[\hat{\Lambda}_{NA}(t) - \Lambda_0(t)] \ .$$

Now counting process and martingale argument similar to Andersen et. al. (1993) Chapter 4 can be used to analyze the integral. Since $g(t)$ is predictable (left continuous), martingale central limit theorem can be used here, which also give the expression of $\sigma^2_\Lambda(g)$.

□

Lemma B *Let g be a given function that satisfy (3.4) and h be any left continuous function such that $\int h^2 d\Lambda < \infty$ and $g(t)h(t) \ge 0$ a.s. F. If λ_n is the solution of (3.6) then*

(1) $\lambda_n = O_p(n^{-1/2})$,

(2) $n\lambda_n^2 \xrightarrow{\mathcal{D}} \chi^2_{(1)} \frac{\sigma^2_\Lambda(g)}{\left(\int_0^\infty g(x)h(x)d\Lambda_0(x)\right)^2}$ as $n \to \infty$, where

$$\sigma^2_\Lambda(g) = \int_0^\infty \frac{g^2(x)}{(1 - F_0(x))(1 - G_0(x))} d\Lambda_0(x) \ .$$

Proof:

The fact that $\int g^2 d\Lambda_0 < \infty$ and $\int h^2 d\Lambda_0 < \infty$ imply that

$$\mathrm{E}_{F_0,G_0} \frac{\delta_i g^2(T_i)}{(1-F_0(T_i))(1-G_0(T_i))} < \infty, \tag{1}$$

and

$$\mathrm{E}_{F_0,G_0} \frac{\delta_i h^2(T_i)}{(1-F_0(T_i))(1-G_0(T_i))} < \infty, \tag{2}$$

which also imply that $\mathrm{E}_{F_0,G_0} \delta_i g^2(T_i) < \infty$ and $\mathrm{E}_{F_0,G_0} \delta h^2(T_i) < \infty$ respectively. Therefore, by using a Lemma of Owen (1990, p98), we have

$$\max_{1 \le i \le n} \delta_i |g(T_i)| = o(n^{1/2}) \tag{3}$$

$$\max_{1 \le i \le n} \delta_i |h(T_i)| = o(n^{1/2}) \tag{4}$$

with probability 1 as $n \to \infty$. Define $M_n = \max_{1 \le i \le n} \delta_i |h(T_i)|$, by (4), $M_n = o(n^{1/2})$ with probability 1 as $n \to \infty$.

Now we show that $\lambda_n = O_p(n^{-1/2})$. For the sake of simple notation we assume that $T_1 < \cdots < T_n$. Consider

$$\begin{aligned} 0 &\equiv \left| \sum_{i=1}^{n-1} g(T_i) \Delta \hat{\Lambda}_{NA}(T_i) \frac{1}{1+\lambda_n h(T_i)} + g(T_n) \Delta \hat{\Lambda}_{NA}(T_n) - \theta_0 \right| \\ &= \left| \sum_{i=1}^{n} g(T_i) \Delta \hat{\Lambda}_{NA}(T_i)) - \theta_0 + \lambda_n \sum_{i=1}^{n-1} \frac{g(T_i) h(T_i)}{1+\lambda_n h(T_i)} \Delta \hat{\Lambda}_{NA}(T_i) \right| \\ &\ge \left| \sum_{i=1}^{n} \delta_i g(T_i) \Delta \hat{\Lambda}_{NA}(T_i)) - \theta_0 \right| + \frac{|\lambda_n|}{1+|\lambda_n| M_n} \sum_{i=1}^{n-1} \delta_i g(T_i) h(T_i) \Delta \hat{\Lambda}_{NA}(T_i). \end{aligned} \tag{5}$$

The first term of (5) has a limiting normal distribution by Lemma A, so it is $O_p(n^{-1/2})$. Again by Lemma A we have that, as $n \to \infty$,

$$\sum_{i=1}^{n-1} \delta_i g(T_i) h(T_i) \Delta \hat{\Lambda}_{NA}(T_i) \xrightarrow{\mathrm{P}} \int_0^\infty g(x) h(x) d\Lambda_0(x). \tag{6}$$

It follows that

$$\frac{|\lambda_n|}{1+|\lambda_n| M_n} = O_p(n^{-1/2}),$$

which implies that $\lambda_n = O_p(n^{-1/2})$. This rate of λ_n is used in the following expansion.

Expanding the right hand side of the constraint equation,

$$\begin{aligned} 0 &= \sum_{i=1}^{n-1} g(T_i) \Delta \hat{\Lambda}_{NA}(T_i) \frac{1}{1+\lambda_n h(T_i)} + g(T_n) \Delta \hat{\Lambda}_{NA}(T_n) - \theta_0 \\ &= \sum_{i=1}^{n} g(T_i) \Delta \hat{\Lambda}_{NA}(T_i) - \theta_0 - \lambda_n \sum_{i=1}^{n-1} \frac{g(T_i) h(T_i)}{1+\lambda_n \delta_i h(T_i)} \Delta \hat{\Lambda}_{NA}(T_i) \\ &= \sum_{i=1}^{n} g(T_i) \Delta \hat{\Lambda}_{NA}(T_i) - \theta_0 - \lambda_n \sum_{i=1}^{n-1} g(T_i) h(T_i) \Delta \hat{\Lambda}_{NA}(T_i) + \lambda_n^2 \sum_{i=1}^{n-1} \frac{g(T_i) h^2(T_i)}{1+\lambda \delta_i h(T_i)} \Delta \hat{\Lambda}_{NA}(T_i), \end{aligned}$$

from which we get an expression of λ_n:

$$\lambda_n = \frac{\sum_{i=1}^{n} g(T_i)\Delta\hat{\Lambda}_{NA}(T_i) - \theta_0}{\sum_{i=1}^{n-1} g(T_i)h(T_i)\Delta\hat{\Lambda}_{NA}(T_i) - \lambda_n \sum_{i=1}^{n-1} \frac{g(T_i)h^2(T_i)}{1+\lambda_n\delta_i h(T_i)}\Delta\hat{\Lambda}_{NA}(T_i)},$$

where two of the three sums are familiar as in Lemma A. The third sum is (recall $\lambda_n = O_p(n^{-1/2})$)

$$\begin{aligned} & \left|\lambda_n \sum_{i=1}^{n-1} \frac{g(T_i)h^2(T_i)}{1+\lambda_n\delta_i h(T_i)}\Delta\hat{\Lambda}_{NA}(T_i)\right| \\ \leq \quad & \frac{|\lambda_n|M_n}{1-|\lambda_n|M_n} \sum_{i=1}^{n-1} g(T_i)h(T_i)\Delta\hat{\Lambda}_{NA}(T_i) = O_p(n^{-1/2})o(n^{1/2})O_p(1) = o_p(1), \end{aligned}$$

by Lemma A and (4). Again by Lemma A and Slutsky theorem it follows that, as $n \to \infty$

$$n\lambda_n^2 \xrightarrow{\mathcal{D}} \chi^2_{(1)} \frac{\sigma^2_\Lambda(g)}{\left(\int_0^\infty g(x)h(x)d\Lambda_0(x)\right)^2}.$$

□

Proof of Theorem 2:

Define

$$f(\lambda) = \log\left\{\prod_{i=1}^{n}[\Delta\Lambda_\lambda(T_i)]^{\delta_i}[\prod_{j:T_j<T_i}(1-\Delta\Lambda_\lambda(T_j))]^{\delta_i}[\prod_{j:T_j\leq T_i}(1-\Delta\Lambda_\lambda(T_j))]^{1-\delta_i})\right\}, \tag{7}$$

where Λ_λ is defined in (3.5). It is obvious that

$$f(0) = \log\left\{\prod_{i=1}^{n}[\Delta\hat{\Lambda}_{NA}(T_i)]^{\delta_i}[\prod_{j:T_j<T_i}(1-\Delta\hat{\Lambda}_{NA}(T_j))]^{\delta_i}[\prod_{j:T_j\leq T_i}(1-\Delta\hat{\Lambda}_{NA}(T_j))]^{1-\delta_i}\right\}. \tag{8}$$

We still assume that $T_1 < T_2 < ... < T_n$ for the sake of simple notation, thus $f(\lambda)$ can be written as follows

$$\begin{aligned} f(\lambda) &= \sum_{i=1}^{n}\delta_i \log\Delta\Lambda_\lambda(T_i) + \sum_{i=1}^{n}\delta_i\left\{\sum_{j=1}^{i-1}\log(1-\Delta\Lambda_\lambda(T_j))\right\} \\ &\quad + \sum_{i=1}^{n}(1-\delta_i)\left\{\sum_{j=1}^{i}\log(1-\Delta\Lambda_\lambda(T_j))\right\} \\ &= \sum_{i=1}^{n}\delta_i \log\Delta\Lambda_\lambda(T_i) + \sum_{i=1}^{n}\sum_{j=1}^{i-1}\log(1-\Delta\Lambda_\lambda(T_i)) \\ &\quad + \sum_{i=1}^{n}(1-\delta_i)\log(1-\Delta\Lambda_\lambda(T_i)) \\ &= \sum_{i=1}^{n}\delta_i \log\Delta\Lambda_\lambda(T_i) + \sum_{i=1}^{n}(n-i)\log(1-\Delta\Lambda_\lambda(T_i)) \end{aligned}$$

$$
\begin{aligned}
& + \sum_{i=1}^{n} (1-\delta_i) \log (1 - \Delta\Lambda_\lambda(T_i)) \\
= \; & \sum_{i=1}^{n} \delta_i \log \Delta\Lambda_\lambda(T_i) + \sum_{i=1}^{n} (n-i+1-\delta_i) \log \left(1 - \Delta\Lambda_\lambda(T_i)\right)
\end{aligned}
$$

By Lemma B, $\lambda_n = O_p(n^{-1/2})$. Now we may apply Taylor's theorem to get:

$$
f(\lambda_n) = f(0) + \lambda_n f^{'}(0) + \frac{\lambda_n^2}{2} f^{''}(\xi), \qquad |\xi| \leq |\lambda_n|.
$$

Consider the first derivative of f with respect to λ

$$
f^{'}(\lambda) = -\sum_{i=1}^{n-1} \delta_i \frac{h(T_i)}{1+\lambda h(T_i)} + \sum_{i=1}^{n-1} (n-i+1-\delta_i) \frac{\Delta\hat{\Lambda}_{NA}(T_i) \dfrac{h(T_i)}{(1+\lambda h(T_i))^2}}{1 - \Delta\hat{\Lambda}_{NA}(T_i) \dfrac{1}{1+\lambda h(T_i)}} \ ,
$$

Thus

$$
\begin{aligned}
f^{'}(0) &= -\sum_{i=1}^{n-1} \delta_i h(T_i) + \sum_{i=1}^{n-1} (n-i+1-\delta_i) h(T_i) \frac{\Delta\hat{\Lambda}_{NA}(T_i)}{1-\Delta\hat{\Lambda}_{NA}(T_i)} \\
&= -\sum_{i=1}^{n-1} \delta_i h(T_i) + \sum_{i=1}^{n-1} (n-i+1-\delta_i) h(T_i) \frac{\delta_i}{n-i+1-\delta_i} \\
&= 0 \ . \qquad (9)
\end{aligned}
$$

The second derivative of f with respect of λ evaluated at ξ is given by

$$
\begin{aligned}
f^{''}(\xi) &= \sum_{i=1}^{n-1} \frac{\delta_i h^2(T_i)}{(1+\xi h(T_i))^2} + \sum_{i=1}^{n-1} (n-i+1-\delta_i) \Delta\hat{\Lambda}_{NA}(T_i) \frac{\dfrac{-2h^2(T_i)}{(1+\xi h(T_i))^3} + \dfrac{\Delta\hat{\Lambda}_{NA}(T_i) h^2(T_i)}{(1+\xi h(T_i))^4}}{\left(1 - \Delta\hat{\Lambda}_{NA}(T_i) \dfrac{1}{1+\xi h(T_i)}\right)^2} \\
&= \sum_{i=1}^{n-1} \frac{\delta_i h^2(T_i)}{(1+\xi h(T_i))^2} - 2 \sum_{i=1}^{n-1} \frac{(n-i+1-\delta_i) \Delta\hat{\Lambda}_{NA}(T_i) h^2(T_i)}{[1+\xi h(T_i)][1+\xi h(T_i) - \Delta\hat{\Lambda}_{NA}(T_i)]^2} \\
&\quad + \sum_{i=1}^{n-1} \frac{(n-i+1-\delta_i) \Delta\hat{\Lambda}_{NA}^2(T_i) h^2(T_i)}{[1+\xi h(T_i)]^2 [1+\xi h(T_i) - \Delta\hat{\Lambda}_{NA}(T_i)]^2}.
\end{aligned}
$$

Notice that $\max \xi \delta_i h(T_i) = o(1)$, it implies that $1/(1+\xi \delta_i h(T_i))^2 = 1 + o(1)$ uniformly for $1 \leq i \leq n-1$ etc. Using this and remember $\Delta\hat{\Lambda}_{NA}(T_i) = \delta_i/(n-i+1)$ we have

$$
f^{''}(\xi) = \sum_{i=1}^{n-1} \delta_i h^2(T_i)(1+o(1)) - 2 \sum_{i=1}^{n-1} \delta_i h^2(T_i) \frac{[1+o(1)][\frac{n-i+1}{n-i+1-\delta_i}]}{[1+o(1)\frac{n-i+1}{n-i}]^2} + \sum_{i=1}^{n-1} \delta_i h^2(T_i) \frac{[1+o(1)][\frac{1}{n-i+1-\delta_i}]}{[1+o(1)\frac{n-i+1}{n-i}]^2}
$$

and obviously $1 + o(1)\frac{n-i+1}{n-i} = 1 + o(1)$ uniformly for $1 \leq i \leq n-1$, thus

$$
f^{''}(\xi) = -\sum_{i=1}^{n-1} \delta_i h^2(T_i)(1+o(1)) - 2 \sum_{i=1}^{n-1} \delta_i h^2(T_i) \frac{1}{n-i}(1+o(1)) + \sum_{i=1}^{n-1} \delta_i h^2(T_i) \frac{1}{n-i}(1+o(1)) \ .
$$

Now we show that, as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n-1}\frac{\delta_i}{(n-i)}h^2(T_i) \xrightarrow{\text{P}} 0. \tag{10}$$

Observe

$$\begin{aligned} 0 \le \sum_{i=1}^{n-1}\frac{\delta_i}{(n-i)}h^2(T_i) &= \sum_{i=1}^{n-1}\frac{\delta_i}{n-i+1}h^2(T_i)\frac{n-i+1}{n-i} \\ \le \quad & \sum_{i=1}^{n-1}h^2(T_i)\Delta\hat{\Lambda}_{NA}(T_i)2 \le 2\sum_{i=1}^{n}h^2(T_i)\Delta\hat{\Lambda}_{NA}(T_i), \end{aligned}$$

we have, by lemma A, as $n \to \infty$,

$$\sum_{i=1}^{n}h^2(T_i)\Delta\hat{\Lambda}_{NA}(T_i) \xrightarrow{\text{P}} \int_0^\infty h^2 d\Lambda_0 < \infty,$$

so (10) holds.

Since $\text{E}_{F_0,G_0}\delta_i h^2(T_i) < \infty$, by the law of large numbers

$$\frac{1}{n}\sum_{i=1}^{n}\delta_i h^2(T_i) \xrightarrow{\text{P}} \text{E}_{F_0,G_0}\delta_1 h^2(T_1) \qquad \text{as} \qquad n \to \infty$$

where

$$\begin{aligned} \text{E}_{F_0,G_0}\delta_1 h^2(T_1) &= \iint_{x\le c} h^2(x)\, dG_0(c)\, dF_0(x) \\ &= \int_0^\infty h^2(x)(1-G_0(x))(1-F_0(x))d\Lambda_0(x)\ . \end{aligned}$$

Note that $\delta_n h^2(T_n) = o_p(n)$ by the assumption $\int h d\Lambda < \infty$, in view of the just mentioned law of large number and (10) we have

$$-\frac{1}{n}f''(\xi) \xrightarrow{\text{P}} \int_0^\infty h^2(x)(1-G_0(x))(1-F_0(x))d\Lambda_0(x), \qquad \text{as } n \to \infty. \tag{11}$$

By (7) and (8) we can write $-2\log ELR_h(\mu)$ as follows

$$\begin{aligned} -2\log ELR_h(\mu) &= 2(\log EL(\hat{\Lambda}_{NA}) - \log EL(\Lambda_{\lambda_n})) \\ &= 2(f(0) - f(\lambda_n)) \\ &= 2\left(f(0) - f(0) - \lambda_n f'(0) - \frac{\lambda_n^2}{2}f''(\xi)\right) \\ &= -\lambda_n^2 f''(\xi) = n\lambda_n^2 \times \frac{-1}{n}f''(\xi). \end{aligned}$$

Thus by (11) and Lemma B we have

$$-2\log ELR_h(\mu) \xrightarrow{\mathcal{D}} \chi^2_{(1)} \times r_h,$$

where

$$r_h = \frac{\sigma^2_\Lambda(g) \int_0^\infty h^2(x)(1-G_0(x))(1-F_0(x))d\Lambda_0(x)}{\left(\int_0^\infty g(x)h(x)d\Lambda_0(x)\right)^2} \ .$$

□

References

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

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

Klein, J.P. (1991). Small sample moments of some estimators of the variance of the Kaplan-Meier and Nelson-Aalen estimators. *Scand. J. Statistics* **18**, 333-40.

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.

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. (1997). *Empirical Likelihood Ratio Method for Censored Data.* Ph. D. dissertation, Department of Statistics, University of Kentucky.

van der Vaart, A. (1991). On differentiable functionals. *Ann. Statist.* **19**, 178-204.

Department of Statistics
University of Kentucky
Lexington, KY 40506-0027
mai@ms.uky.edu