Empirical Likelihood Ratio in Terms of Cumulative Hazard Function for Censored Data

Xiao-Rong Pan and Mai Zhou

University of Kentucky and Searle
E-mail: mai@ms.uky.edu

Received April 17, 2000; published online July 19, 2001

It has been shown that (with complete data) empirical likelihood ratios can be used to form confidence intervals and test hypotheses 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.

Key words and phrases: weighted hazard; one sample log rank test; stochastic constraint; median.

1. INTRODUCTION

Based on the likelihood function there are three different methods to produce confidence intervals: namely Wald’s method, Rao’s method, and Wilks’ method. Among the three, the Wilks likelihood ratio (LR) method does not need the calculation of information or the inverse of that. It automatically adjust the statistics \(-2 \log \text{LR}\) 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 three are easy to obtain, the LR method still holds some unique advantages. For example, the confidence intervals produced by the LR method are always range respecting (confidence bounds inside the parameter space), while the other two are not. Therefore, transformation on the parameter is often used in connection with Wald’s and Rao’s methods 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 achieving the result comparable to 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 nonparametric settings after some modification. He termed this empirical likelihood ratio method. The empirical likelihood (EL) of \( n \) i.i.d. observations \( X_i \) is just

\[
EL(F) = \prod_{i=1}^{n} A(F(X_i)).
\]

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

\[
\int g(t) dF(t) = \mu,
\]  

(1.1)

Owen (1988, 1990) showed that the distribution function that maximizes 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 \( X_i \) equal to

\[
A(F(X_i)) = A\hat{F}_n(X_i) \times \frac{1}{1 + \lambda(g(X_i) - \mu)},
\]

where \( \lambda \) is defined by the equation

\[
\sum_{i=1}^{n} A\hat{F}_n(X_i) \times \frac{g(X_i)}{1 + \lambda(g(X_i) - \mu)} = \mu.
\]

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. However, a generalization of the above setting to the right censored data case is difficult. No explicit maximization under constraint (1.1) can be obtained.

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

\[
A(t) = \int_{(t, \infty)} dF(s) \frac{1}{1 - \hat{F}(s)},
\]  

(1.2)

It gives rise to a martingale formulation of the observations. For example, the 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 and Johnston, 1990) and the 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) \( \theta = \int g(t) \, dA(t) \) for arbitrary given \( g(t) \). (2) \( \theta_s = \int g_s(t) \, dA(t) \) where \( g_s(t) \) is a random but predictable function and depends on sample size \( n \); \( \theta_s \) can also change with sample size \( n \). (3) \( \theta \) is defined implicitly: \( \int g(t, \theta) \, dA(t) = C \) 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 \( A_\tau(t) = \int_0^t g(s) \, dA_s(t) \), where \( A_s \) is the baseline cumulative hazard.

The parameter of the second type is prompted by the one sample log-rank type tests. The weight function of the one sample log-rank test takes the form \( g(t) = Y(t)/n \) where \( Y(t) \) is the size of the risk set at time \( t \). See, for example, Andersen et al. (1993, Sect. V.1) for details and other similar types of tests. As a further example for the stochastic weight function \( g \), we take the mean, which can be obtained from the integration of the cumulative hazard with \( g(t) = \int t \, dF(t) \). Since \( F \) is unknown, we may use \( g_n(t) = \int t \, d\hat{F}_n(t) \).

The prime example for the implicit type parameters is the quantiles. For example, the parameter \( \theta \) of the median may be defined implicitly as \( \int I_{t \leq \theta} \, dA(t) = \log 2 \).

Another purpose of this paper is to serve as a starting point in the comparison of the two different types of empirical likelihoods with right censored data, (2.4) and (2.5). Section 4 shows that for continuous \( F \) and as \( n \to \infty \) the two are equivalent, but there are many differences when \( F \) is discrete and/or for small \( n \). We shall present the differences when using the three types of parameters discussed above in a forthcoming paper.

Murphy (1995) also studied the empirical likelihood ratio using counting process formulations. She obtained the explicit result when the constraint is the hazard function itself evaluated at a point, \( A(t_0) = -\log [1 - F(t_0)] \). Li (1995), building on the earlier work of Thomas and Grunkemeier (1975), studied the empirical likelihood method for censored data, but only for the parameters of the form \( F(t) \). Murphy and Van der Vaart (1997) proved a very general result but in each specific case one still needs to work out 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 (4.1)). 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 two 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 two 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.

2. LIKELIHOOD IN TERMS OF HAZARD AND ITS MAXIMUM UNDER A CONSTRAINT OF TYPE 1

Suppose that \( X_1, \ldots, X_n \) 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 \( C_1, \ldots, C_n \) which are i.i.d. with a distribution \( G_0 \). Only the censored observations are available to us:

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

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

\[
EL(F) = \prod_{i=1}^{n} \left[ AF(T_i) \right]^{\delta_i} \left[ 1 - F(T_i) \right]^{1-\delta_i}. \quad (2.2)
\]

Since the NPMLE of the distribution \( F \) and hazard \( A \) 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 NPMLEs. 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

\[
1 - F(t) = \prod_{x \leq t} (1 - AA(x)) \quad \text{and} \quad AA(t) = \frac{AF(t)}{1 - F(t)} \quad (2.3)
\]
that is valid for purely discrete distributions we can rewrite (2.2) in terms of the cumulative hazard function. The empirical likelihood (2.2) becomes

$$EL(A) = \prod_{i=1}^{n} \left[ AA(T_i) \right]^{\delta_i} \left[ \prod_{j: T_j < T_i} (1 - AA(T_j)) \right]^{\delta_i}$$

$$\times \left[ \prod_{j: T_j < T_i} (1 - AA(T_j)) \right]^{1-\delta_i}. \quad (2.4)$$

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

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

$$AL(A) = \prod_{i=1}^{n} [AA(T_i)]^{\delta_i} \exp[-A(T_i)]. \quad (2.5)$$

See also Gill (1989) for a detailed discussion of different extensions of the 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 $AL(A)$ 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 maximizes $AL(A)$ under the constraint (of type 1)

$$\int g(t) dA(t) = \theta, \quad (2.6)$$

where $g(t)$ is a given function that satisfies some moment conditions, and $\theta$ 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 \( w_i = \Delta A(T_i) \) for \( i = 1, 2, \ldots, n \), where we notice \( w_n = \delta_n \). The constraint (2.6) for any \( \lambda \), that is dominated by the 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 \( \lambda \) can be written in terms of the jumps

\[
AL = \prod_{i=1}^{n} \left[ w_i \right]^6 \exp \left\{ -\sum_{j=1}^{i} w_j \right\}. \tag{2.8}
\]

Another important issue is that the constraint equation may not always have a solution for certain values of \( \theta \). 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 \( \theta \) 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 \( i = 1, 2, \ldots, n-1 \).

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

**Theorem 1.** The feasible values of \( \theta \) in the constraint (2.7) are given by the interval \( \theta^* \) 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

\[
w_i = W_i = \frac{\delta_i}{(n-i+1) + n \delta_i g(T_i) \delta_i} = \frac{\delta_i}{n-i+1} \times \frac{1}{1 + \lambda \delta_i g(T_i)/(n-i+1)}. \tag{2.9}
\]
where \( \lambda \) in turn is the solution of the equation

\[
\lambda(g(T)) = \delta
\]

where \( g(T) \equiv \sum_{i=1}^{n-1} \frac{\delta_i}{n-i+1} \left( \frac{1}{1 + \lambda(g(T))((n-i+1)/n)} \right) + g(T_n) \delta_n.
\]

(2.10)

**Proof.** To use the Lagrange multiplier, we form the target function

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

Taking partial derivative with respect to \( \omega_i \), for \( i = 1, ..., n-1 \), and letting them equal zero, we obtain

\[
\frac{\partial G}{\partial \omega_i} = \delta_i - (n-i+1) - n \lambda g(T_i) \delta_i = 0, \quad i = 1, 2, ..., n-1.
\]

By solving this equation we get the explicit expression for \( \omega_i \)

\[
W_i = \frac{\delta_i}{(n-i+1) + n \lambda g(T_i) \delta_i}
\]

\[
= \frac{\delta_i}{n-i+1} \times \frac{1}{1 + \lambda(g(T_i))((n-i+1)/n)}
\]

\[
= A \hat{A}_{\lambda}(T_i) \left( \frac{1}{1 + \lambda(g(T_i))((n-i+1)/n)} \right)
\]

for \( i = 1, 2, ..., n-1 \),

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

\[
\lambda(g(T)) = \delta
\]

The function \( \lambda(g(T)) \) above is monotone decreasing and continuous in \( \lambda \), a fact that can be verified by taking a derivative of \( \lambda(g(T)) \) with respect to \( \lambda \). On the other hand, any choice of legitimate value \( \lambda \) 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 \( \lambda \) range \( \mathcal{J} \).
All max and min in the following definitions are taken in the domain \( \{i: 1 \leq i \leq n - 1, \, \delta_i = 1, \, \text{and} \, g(T_i) \neq 0\} \); if there is any additional restriction then we specify in each individual case.

**Case 1.** When \( \min g(T_i) > 0 \)

\[
\mathcal{J} = \left( \max_{\text{ng}(T_i)} \frac{i-n}{ng(T_i)}, \, \infty \right) := (\hat{\lambda}, \infty).
\]

**Case 2.** When \( \max g(T_i) < 0 \)

\[
\mathcal{J} = \left( -\infty, \min_{\text{ng}(T_i)} \frac{i-n}{ng(T_i)} \right) := (-\infty, \bar{\lambda}).
\]

**Case 3.** When \( \max g(T_i) > 0 > \min g(T_i) \)

\[
\mathcal{J} = \left( \max_{g(T_i) > 0} \frac{i-n}{ng(T_i)}, \min_{g(T_i) < 0} \frac{i-n}{ng(T_i)} \right) := (\hat{\lambda}, \bar{\lambda}).
\]

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

**Case 1.**

\[
\mathcal{V}^- = \left( g(T_n) \delta_n, \sum_{i=1}^{n-1} \frac{\delta_i g(T_i)}{n-i+1 + n\hat{\lambda} g(T_i)} + g(T_n) \delta_n \right).
\]

**Case 2.**

\[
\mathcal{V}^- = \left( \sum_{i=1}^{n-1} \frac{\delta_i g(T_i)}{n-i+1 + n\hat{\lambda} g(T_i)} + g(T_n) \delta_n, \, g(T_n) \delta_n \right).
\]

**Case 3.**

\[
\mathcal{V}^- = \left( \sum_{i=1}^{n-1} \frac{\delta_i g(T_i)}{n-i+1 + n\hat{\lambda} g(T_i)} + g(T_n) \delta_n, \right).
\]

\[
\sum_{i=1}^{n-1} \frac{\delta_i g(T_i)}{n-i+1 + n\bar{\lambda} g(T_i)} + g(T_n) \delta_n.
\]
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 \( \lambda \) of (2.10).

**Lemma 1.** Suppose \( g(t) \) is a left continuous function and
\[
0 < \int \frac{|g(x)|^m dA_{0}(x)}{(1-F_{0}(x))(1-G_{0}(x))} < \infty, \quad m = 1, 2.
\]
Then \( \theta_{0} = \int g(t) dA_{0}(t) \) is feasible with probability approaching 1 as \( n \to \infty \), and the solution \( \lambda \) of (2.10) with \( \theta = \theta_{0} \) satisfies
\[
n^{2} \frac{\theta_{0}}{\lambda} \rightarrow \chi^{2}(1) \left( \int \frac{g^2(x) dA_{0}(x)}{(1-F_{0}(x))(1-G_{0}(x))} \right)^{-1} \quad \text{as } n \to \infty.
\]

**Proof.** See the Appendix.

Next we define the empirical likelihood ratio in terms of the hazard for the constraint (2.7) as
\[
\mathcal{L}R(\theta) = \sup \{ AL(A) \mid A \ll \hat{A}_{N_{A}}, \ \text{and} \ \hat{A}_{N_{A}}; \ A \text{ satisfy (2.7)} \}
\]

By Theorem 1, \( \mathcal{L}R(\theta) \) can be computed, when the constraint is feasible, by using \( W_{i} \) defined there and the known property of \( \hat{A}_{N_{A}} \):
\[
\hat{A}_{N_{A}}(T_{i}) = \delta_{i}/(n-i+1).
\]

**Theorem 2.** Let \( (T_{1}, \delta_{1}), \ldots, (T_{n}, \delta_{n}) \) be \( n \) pairs of random variables as defined in (2.1). Suppose \( g \) is a left continuous function and
\[
0 < \int \frac{|g(x)|^m dA_{0}(x)}{(1-F_{0}(x))(1-G_{0}(x))} < \infty, \quad m = 1, 2.
\]
Then, \( \theta_{0} = \int g(t) dA_{0}(t) \) will be a feasible value with probability approaching one as \( n \to \infty \) and
\[
-2 \log \mathcal{L}R(\theta_{0}) \rightarrow \chi^{2}(1) \quad \text{as } n \to \infty.
\]

**Proof.** In view of Lemma 2, we need only to prove the last claim:
\[
-2 \log \mathcal{L}R(\theta_{0}) \rightarrow \chi^{2}(1) \quad \text{as } n \to \infty.
\]
To this end, define
\[
Z_{i} = \delta_{i} g(T_{i}) \frac{1}{(n-i+1)/n} \quad \text{for } i = 1, 2, \ldots, n
\]
and consider

\[ -2 \log \mathcal{L}(\theta_0) = 2 \left[ \sum_{i=1}^{n} \delta_i \log \mathcal{A}_{NA}(T_i) - \sum_{i=1}^{n} (n - i + 1) \mathcal{A}_{NA}(T_i) \right] \]

\[ = 2 \left[ \sum_{i=1}^{n} \delta_i \log \mathcal{A}_{NA}(T_i) \right] \]

\[ + 2 \left[ \sum_{i=1}^{n-1} \delta_i \log(1 + \lambda Z_i) \right] \]

\[ + \sum_{i=1}^{n-1} \frac{(n - i + 1) \mathcal{A}_{NA}(T_i) + \mathcal{A}_{NA}(T_n)}{1 + \lambda Z_i} \]

\[ = -2 \sum_{i=1}^{n} \delta_i + 2 \sum_{i=1}^{n-1} \delta_i \log(1 + \lambda Z_i) + 2 \sum_{i=1}^{n-1} \frac{\delta_i}{1 + \lambda Z_i} + 2\delta_n \]

\[ = -2 \sum_{i=1}^{n} \delta_i + 2 \sum_{i=1}^{n-1} \delta_i \log(1 + \lambda Z_i) + 2 \sum_{i=1}^{n-1} \delta_i \]

\[ - 2 \sum_{i=1}^{n-1} \frac{\delta_i \lambda Z_i}{1 + \lambda Z_i} + 2\delta_n \]

\[ = 2 \sum_{i=1}^{n-1} \delta_i \log(1 + \lambda Z_i) - 2 \sum_{i=1}^{n-1} \delta_i \lambda Z_i + 2 \sum_{i=1}^{n-1} \frac{\delta_i \lambda^2 Z_i^2}{1 + \lambda Z_i}. \]  

(3.2)

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

\[
\max_{1 \leq i \leq n} |Z_i| \leq \max_{1 \leq i \leq n} \left\{ \frac{\delta_i |g(T_i)|}{(1 - F_0(T_i)) (1 - G_0(T_i))} \right\} 
\times \max_{1 \leq i \leq n} \left\{ \frac{(1 - F_0(T_i)) (1 - G_0(T_i))}{(n - i + 1)/n} \right\} 
= o_p(n^{1/2}) O_p(1) = o_p(n^{1/2}).
\]

(3.3)

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

\[
\log(1 + \lambda Z_i) = \lambda Z_i - \frac{1}{2} \lambda^2 Z_i^2 + O_p(\lambda^3) Z_i^3.
\]

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

$$-2 \log \mathcal{L}(\theta_0) = 2 \sum_{i=1}^{n-1} \delta_i Z_i - \sum_{i=1}^{n-1} \delta_i^2 Z_i^2 + O_p(\lambda^3) \sum_{i=1}^{n-1} Z_i^3$$

$$-2 \sum_{i=1}^{n-1} \delta_i Z_i + 2 \sum_{i=1}^{n-1} \delta_i^2 Z_i^2 - 2 \sum_{i=1}^{n-1} \frac{\delta_i^3 Z_i^3}{1 + \lambda Z_i}$$

$$= 2 \lambda \sum_{i=1}^{n-1} \delta_i Z_i^2 + O_p(\lambda^3) \sum_{i=1}^{n-1} Z_i^3 - 2 \delta Z_i^3 \sum_{i=1}^{n-1} \frac{\delta_i^3 Z_i^3}{1 + \lambda Z_i},$$

(3.5)

where, as $n \to \infty$,

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

and, noting $\delta_i Z_i^3 = \frac{Z_i^3}{1 + \lambda Z_i}$,

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

By Lemma A3 and (3.3) we have

$$\text{Plim} \frac{1}{n} \sum_{i=1}^{n} Z_i^2 = \text{Plim} \frac{1}{n} \sum_{i=1}^{n-1} \delta_i Z_i^2 = \text{Plim} \frac{1}{n} \sum_{i=1}^{n-1} Z_i^2$$

$$= \int \frac{g^2(x) \, dA_o(x)}{(1 - F_0(x))(1 - G_0(x))} < \infty,$$

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 1, Lemma A3, and the Slutsky theorem. Thus we have as $n \to \infty$

$$-2 \log \mathcal{L}(\theta_0) \xrightarrow{d} \chi^2(1).$$

4. COMPARISON OF TWO VERSIONS OF LIKELIHOOD

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

\[ 176 \text{ PAN AND ZHOU} \]
Define
\[ \mathcal{D}(\theta) = \frac{EL(A^*)}{EL(A_{NA})}, \]
where \( A^* \) is given by the jumps \( W_i \) defined in Theorem 1.

**Theorem 3.** Suppose all the conditions of Theorem 2 hold. Then
\[ -2 \log \mathcal{D}(\theta_0) \overset{D}{\to} Z_{(1)}^2 \quad \text{as} \quad n \to \infty. \]

**Proof.** We shall prove that the two likelihood ratio statistics are asymptotically equivalent in the sense that their difference goes to zero in probability.

By (3.2) we have
\[ -2 \log \mathcal{D}(\theta_0) = 2 \sum_{i=1}^{n-1} \delta_i \log(1 + \lambda Z_i) - 2 \sum_{i=1}^{n-1} \delta_i \lambda Z_i, \]
where \( Z_i \) is defined as in (3.1). On the other hand, we also have
\[
2 \log \mathcal{D}(\theta_0) = 2 \sum_{i=1}^{n-1} \delta_i \log(1 + \lambda Z_i)
+ 2 \sum_{i=1}^{n-1} (n - i + 1 - \delta_i) \log(1 - A_{NA}(T_i))
- 2 \sum_{i=1}^{n-1} (n - i + 1 - \delta_i) \log \left( 1 - A_{NA}(T_i) \frac{1}{1 + \lambda Z_i} \right). \]

(4.1)

Observe
\[
\log \left( 1 - A_{NA}(T_i) \frac{1}{1 + \lambda Z_i} \right) = \log \left( 1 - A_{NA}(T_i) + A_{NA}(T_i) \frac{\lambda Z_i}{1 + \lambda Z_i} \right) \]

By the same reason as in (3.3), (3.4) we may expand
\[
\log \left( 1 - A_{NA}(T_i) \frac{1}{1 + \lambda Z_i} \right) = \log \left( 1 - A_{NA}(T_i) + A_{NA}(T_i) \frac{\lambda Z_i}{1 + \lambda Z_i} \right)
\]
\[= \log(1 - \bar{A}(T_i)) + \frac{A\bar{A}(T_i)}{1 - A\bar{A}(T_i)} \]
\[\times \frac{\bar{Z}_i}{1 + \bar{Z}_i} \left( \frac{\bar{A}(T_i)}{1 - A(T_i)} \right)^2 \eta_i^2 \]
\[= \log(1 - \bar{A}(T_i)) + \frac{\delta_i}{n - i + 1 - \delta_i} \]
\[\times \frac{\bar{Z}_i}{1 + \bar{Z}_i} \left( \frac{\delta_i}{n - i + 1 - \delta_i} \right)^2 \eta_i^2, \tag{4.2} \]

where \(|\eta_i| \leq |\bar{Z}_i/(1 + \bar{Z}_i)| \).

Substituting (4.2) in the expression of \(-2 \log \mathcal{L}(\theta_0)\), we obtain
\[-2 \log \mathcal{L}(\theta_0) = 2 \sum_{i=1}^{n-1} \delta_i \log(1 + \bar{Z}_i) - 2 \sum_{i=1}^{n-1} \frac{\delta_i \bar{Z}_i}{1 + \bar{Z}_i} \]
\[+ 2 \sum_{i=1}^{n-1} \eta_i^2 \frac{1}{n - i + 1 - \delta_i}. \]

Therefore
\[-2 \log \mathcal{L}(\theta_0) + 2 \log \mathcal{L}(\theta_0) = 2 \sum_{i=1}^{n-1} \eta_i^2 \frac{1}{n - i + 1 - \delta_i}, \]

where
\[0 \leq \sum_{i=1}^{n-1} \eta_i^2 \frac{1}{n - i + 1 - \delta_i} \leq \lambda^2 \sum_{i=1}^{n-1} Z_i^2 \frac{1}{n - i + 1 - \delta_i}. \]

By Lemma 1 and Lemma A3 we have
\[n\lambda^2 \frac{1}{n} \sum_{i=1}^{n-1} Z_i^2 \frac{1}{n - i + 1 - \delta_i} = O_p(1) \quad \Rightarrow \quad o_p(1). \]

Therefore
\[-2 \log \mathcal{L}(\theta_0) + 2 \log \mathcal{L}(\theta_0) \Rightarrow 0 \quad \text{as} \quad n \to \infty. \]

In view of Theorem 2, we have
\[-2 \log \mathcal{L}(\theta_0) \Rightarrow \chi^2_1 \quad \text{as} \quad n \to \infty. \blackslug \]
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 the mean from the integration of the cumulative hazard, we need to let \( g(t) = g_n(t) = [1 - \hat{F}_n(t)] \), again a random function. To accommodate this, we allow the function \( g \) to depend on the sample (of size \( n \)) but require that it be a predictable function with respect to the filtration that makes \( \hat{A}_{NA}(t) - A(t) \) a martingale. For example, the filtration

\[
\mathcal{F}_t = \sigma\{T_k I(T_k \leq t); \delta_k I(T_k \leq t); k = 1, 2, \ldots, n\}. \tag{5.1}
\]

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

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

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

\[
\int g_n(t) \, dA(t) = \theta_n. \tag{5.3}
\]

The \( \theta \) value may also depend on \( n \). For example, if we are testing the hypothesis \( H_0: A \equiv A_0 \) then we should take \( \theta_n = \int g_n(t) \, dA_0(t) \).

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

\[
-2 \log \mathcal{L}(A) = \sup \left\{ AL(A) \mid A \ll \hat{A}_{NA} \text{ and } A \text{ satisfy (5.3)} \right\},
\]

where the numerator of the ratio can be computed similarly as in Theorem 1 with \( g_n(t) \) and \( \theta_n \) replacing \( g(t) \) and \( \theta \) there.

**Theorem 4.** Let \((T_1, \delta_1), \ldots, (T_n, \delta_n)\) 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 satisfying (5.2). Also assume

\[
0 < \int \frac{|g(x)|^m}{(1 - F_o(x))(1 - G_o(x))} \, dA_0(x) < \infty, \quad m = 1, 2.
\]
Then $\theta_n^0 = \int g_n(t) \, dA_0(t)$ will be a feasible value with probability approaching one as $n \to \infty$ and

$$-2 \log \mathcal{L}(\theta_n^0) \xrightarrow{d} \chi^2(1) \quad \text{as} \quad n \to \infty.$$  

5.2. Implicit Constraints

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

$$\int g(t, \theta) \, dA(t)$$

be monotone in $\theta$ for any given cumulative hazard function $A$, and (ii)

$$\int g(t, \theta) \, dA_0(t) = C$$

uniquely define the parameter $\theta_0$.

The likelihood ratio in this case is formed similarly. For given $\theta$ we first solve the following equation to get $\lambda$,

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

where $C$ is a given constant. Then $\mathcal{L}(\theta)$ is defined as the ratio of two ALS with the numerator computed as (2.8) with

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

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

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

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

Then,

$$-2 \log \mathcal{L}(\theta_0) \xrightarrow{d} \chi^2(1) \quad \text{as} \quad n \to \infty.$$
6. SIMULATIONS AND EXAMPLES

Notice our results in Section 2 reduce the computation of the maximization to a single parameter $\lambda$. All we need to solve is the constraint equation for $\lambda$ and it is monotone decreasing in $\lambda$. A 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: $A_0(t) = t$
- Sample size: $n = 20$
- $g(t) = e^{-t}$
- Parameter $\theta_0$: $\theta_0 = \int_0^\infty g(t) \, dA_0(t) = 1$

The 95% confidence interval for $\theta_0$ can be constructed as

$$\{ \theta \mid -2 \log \mathcal{L}(\theta) \leq 3.84 \}.$$  

Each time we compute $-2 \log \mathcal{L}(\theta = 1)$ and check to see if it is less than 3.84 (inside the interval). In 1000 independent such runs we recorded 947 inside for intervals that are supposed to have an asymptotical nominal coverage probability of 95%. For the same data the Wald confidence interval based on the Nelson–Aalen type estimator results in 920 inside 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 a slightly modified (break tie) version of Lee (1992, Example 4.2): 3, 6.5, 6.5, 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 $t = 9.8$, $A_0(9.8)$. Hence $\theta_0 = A_0(9.8)$. In this case the function $g$ is an indicator function: $g(t) = I_{t \leq 9.8}$.

The 95% confidence interval using the empirical likelihood ratio, $-2 \log \mathcal{L}(\theta)$, for $A_0(9.8)$ 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 inherits 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 take 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 distri-
butions, some smoothing or other modification may be needed, particu-
larly for small sample sizes. However, those modifications will become
negligible 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 obser-
vations. Since our formula developed in this paper assumes no ties in the
data, we shall break the ties by subtracting a small amount (0.00001) from
the successive observations. This is equivalent to assuming 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 dis-
tribution \( F_0 \) is continuous and the median is uniquely defined for
\( F_0 \). We shall take \( g(t, \theta) = I_{t < \theta} \) and constraint \( \int g(t, \theta) \, dA(t) = \log 2 \).

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

\[
\{ \theta \mid -2 \log \mathcal{D}(\theta) < 3.84 \}
\]

with the constraint \( \int g(t, \theta) \, dA(t) = \log 2 \). The 0.8 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.

**APPENDIX**

**Lemma A1.** For any random variable \( Y \), if \( E |Y|^k < \infty \) then for an i.i.d.
sample \( Y_1, Y_2, ..., Y_n \) that has the same distribution as \( Y \), we have

\[
\max_{1 \leq i \leq n} |Y_i| = o(n^{1/k}) \quad \text{a.s.}
\]

Lemma A.2. Let \((T_1, \delta_1), \ldots, (T_n, \delta_n)\) be \(n\) i.i.d. pairs of random variables, where each \((T_i, \delta_i)\) is defined by (2.1). Let also \(T^*_n = \max_{1 \leq i \leq n} T_i\). If \(\int h^2(x) \, dA_0(x) < \infty\), then

\[
\max_{1 \leq i \leq n} \frac{\delta_i |h(T_i)|}{\sqrt{(1 - F_0(T_i))(1 - G_0(T_i))}} = o(n^{1/2}) \text{ a.s. and } \delta^*_n h(T^*_n) = o_p(1),
\]

where \(\delta^*_n\) is the indicator function corresponding to \(T^*_n\).

Proof. Since \(\int h^2(x) \, dA_0(x) < \infty\), we have

\[
E_{F_0, G_0} \frac{\delta_i h^2(T_i)}{(1 - F_0(T_i))(1 - G_0(T_i))} = \int h^2(x) \, dA_0(x) < \infty.
\]

Therefore, by Lemma A1, we have

\[
\max_{1 \leq i \leq n} \frac{\delta_i |h(T_i)|}{\sqrt{(1 - F_0(T_i))(1 - G_0(T_i))}} = o(n^{1/2}), \quad (A.1)
\]

with probability 1 as \(n \to \infty\).

The fact that

\[
\frac{\delta^*_n |h(T^*_n)|}{\sqrt{(1 - F_0(T^*_n))(1 - G_0(T^*_n))}} \leq \max_{1 \leq i \leq n} \frac{\delta_i |h(t_i)|}{\sqrt{(1 - F_0(T_i))(1 - G_0(T_i))}}
\]

implies

\[
\frac{\delta^*_n |h(T^*_n)|}{\sqrt{(1 - F_0(T^*_n))(1 - G_0(T^*_n))}} = o(n^{1/2}), \quad (A.2)
\]

with probability 1 as \(n \to \infty\).

Let \(H_0(t)\) be the distribution function of \(T_n\), 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

\[
1 - H_0(T^*_n) = O_p(n^{-1}), \quad (A.3)
\]

or

\[
\sqrt{(1 - F_0(T^*_n))(1 - G_0(T^*_n))} = O_p(n^{-1/2}),
\]

then it follows from (A.2) that \(\delta^*_n h(T^*_n) = o_p(1)\).
Now we show $1 - H_0(T^*_n) = O_p(n^{-1})$. For any $\varepsilon > 0$, there exists $M_0 > 0$ such that $\exp(-M_0) < \varepsilon$. For $M > M_0$, consider
\[
P \left( \frac{1 - \max_{1 \leq i \leq n} H_0(T_i)}{n^{-1}} > M \right) = P \left( \frac{1 - \max_{1 \leq i \leq n} H_0(T_i)}{n^{-1}} > M \right)
\]
\[
= P \left( \max_{1 \leq i \leq n} H_0(T_i) < (1 - n^{-1} \times M) \right)
\]
\[
= \left( 1 - \frac{M}{n} \right)^n \leq \exp(-M) < \varepsilon.
\]
Therefore $1 - H_0(T^*_n) = O_p(n^{-1})$.

**Lemma A3.** Under the assumptions of Theorem 2, we have, for $Z_i$ defined in (3.1),
\[
\frac{1}{n} \sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \frac{\delta_i g^2(T_i) n}{(n-i+1)^2} \rightarrow \int \frac{g^2(t) dA_i(t)}{(1-F)(1-G)}, \tag{A.4}
\]
and
\[
\frac{1}{n} \sum_{i=1}^{n-1} Z_i^2 = \int \frac{I_{\{Y_i > 1\}} g^2(t)}{(Y_i) \cdot Y(t)/n} dA_i(t) \rightarrow 0 \quad \text{as } n \to \infty, \tag{A.5}
\]
where $Y(t) = \sum I_{\{T_i > t\}}$.

**Proof.** For (A.5), use Lenglart's inequality on the integral to switch to a similar integral except with respect to $A_i(t)$, and then use uniform convergence of the 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),
\[
\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} Z_i - \theta_0 \right) = \sqrt{n} \left( \sum_{i=1}^{n} g(T_i) \cdot A_i N_x(T_i) - \theta_0 \right) \overset{d}{\rightarrow} N(0, \sigma^2_\alpha(g)),
\]
where $\sigma^2_\alpha(g) = \int \left( g^2(x) dA_0(x)/(1-F_0(x))(1-G_0(x)) \right)$ and $\theta_0 = \int g(t) dA_0(t)$. 

Proof. Notice the summation can be written as an integral
\[n \sum_{i=1}^n g(T_i) \Delta \hat{A}_{\text{Na}}(T_i) - \theta_0 = \int g(t) d[\hat{A}_{\text{Na}}(t) - A_0(t)].\]

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

Proof of Lemma 1. First we notice that if we set \(\lambda = 0\) in the constraint equation (2.10), the jumps \(W_i\) reduce to those of the Nelson–Aalen estimator, implying that \(\theta = \hat{\theta}_n = \int g(t) d\hat{A}_{\text{Na}}(t)\) is always a feasible value, i.e., \(\theta_0 \in \mathcal{Y}\).

On the other hand, notice that the derivative
\[
\frac{\partial l(\lambda)}{\partial \lambda} = -\sum_{i=1}^{n-1} \frac{\delta g(T_i)}{n-i+1} \frac{Z_i}{[1 + z^2]^2},
\]
and when evaluated at \(\lambda = 0\) we have
\[
\frac{\partial l(\lambda)}{\partial \lambda}|_{\lambda = 0} = -\frac{1}{n} \sum_{i=1}^{n-1} Z_i^2.
\]

By Lemmas A2 and A3 it converges (in fact almost surely) to
\[
-\int \frac{g^2(x) dA_0(x)}{(1 - F_0(x))(1 - G_0(x))}.
\]
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 \(\lambda, \mathcal{Y}\), covers at least an open interval of length \(1/\sigma_n(n^{1/2})\) for all \(n\) centered at 0, then the feasible value of \(\theta, \mathcal{Y}\), will also contain an open interval of length \(1/\sigma_n(n^{1/2})\) centered at \(\hat{\theta}_n\). Since \(\hat{\theta}_n - \theta_0 = O_p(n^{-1/2})\), this will ensure that \(\theta_0\) will be in \(\mathcal{Y}\), i.e., a feasible value, for large \(n\).

The fact that the legitimate value of \(\lambda, \mathcal{Y}\), covers at least an open interval of length \(1/\sigma_n(n^{1/2})\) for all \(n\) centered at zero can easily be seen from the definition of \(\mathcal{Y}\) by noticing that
\[
\frac{1}{|\lambda|} = o_p(n^{1/2})
\]
which can be proved similarly to (3.3). The argument for \(\tilde{\lambda}\) is the same.
Now we turn to the asymptotic distribution of the solution $\lambda$ when $\theta = \theta_0$. The first step is to show that $\lambda = O_p(n^{-1/2})$ where $\lambda$ 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

$$0 = \left| \theta_0 - \frac{1}{n} \sum_{i=1}^{n} Z_i \frac{Z_i}{1 + \lambda Z_i} - \frac{1}{n} Z_n \right|$$

$$= \left| \theta_0 - \frac{1}{n} \sum_{i=1}^{n} Z_i + \frac{\lambda}{n} \sum_{i=1}^{n} \frac{Z_i^2}{1 + \lambda Z_i} - \frac{1}{n} Z_n \right|$$

$$= \left| \left( \theta_0 - \frac{1}{n} \sum_{i=1}^{n} Z_i \right) + \frac{\lambda}{n} \sum_{i=1}^{n} \frac{Z_i^2}{1 + \lambda Z_i} \right|$$

$$\geq \frac{|\lambda|}{1 + |\lambda|} \max |Z_i| \left| \frac{1}{n} \sum_{i=1}^{n} Z_i \right| - \left| \theta_0 - \frac{1}{n} \sum_{i=1}^{n} Z_i \right|. \quad (A.6)$$

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

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

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

$$\frac{1}{n} \sum_{i=1}^{n} Z_i^2 \overset{p}{\to} \int \frac{g^2(x)}{(1-F_0(x))(1-G_0(x))} dA_0(x), \quad (A.7)$$

and it follows that

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

which implies that

$$\lambda = O_p(n^{-1/2}). \quad (A.8)$$
Expanding (2.10), we obtain
\[
0 = \frac{1}{n} \sum_{i=1}^{n} Z_i - \theta_0 - \frac{1}{n} \sum_{i=1}^{n-1} \frac{Z_i^2}{1 + \lambda Z_i}
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} Z_i - \theta_0 - \frac{1}{n} \sum_{i=1}^{n-1} Z_i^2 + \frac{\lambda}{n} \sum_{i=1}^{n} \frac{1}{1 + \lambda Z_i}
\]
(A.9)

The last term in (A.9) is bounded by ((A.8), (3.3), and Lemma A3)
\[
\lambda^2 \frac{1}{n} \sum_{i=1}^{n-1} |Z_i^3| \leq \lambda^2 \max |Z_i| \frac{1}{n} \sum_{i=1}^{n-1} Z_i^2
\]
\[
= O_p(n^{-1}) \sigma_p(n^{1/2}) = O_p(1) = o_p(n^{-1/2}).
\]

Therefore we get an expression of \( \hat{\lambda} \) as
\[
\hat{\lambda} = \left( \frac{1}{n} \sum_{i=1}^{n} Z_i - \theta_0 \right) \left( \frac{1}{n} \sum_{i=1}^{n-1} Z_i^2 \right)^{-1} + o_p(n^{-1/2}).
\]
(A.10)

By Lemma A4, as \( n \to \infty \)
\[
\frac{1}{n} \sum_{i=1}^{n} Z_i - \theta_0 = \sqrt{n} \left( \sum_{i=1}^{n} g(T_i) A \hat{N}_h(T_i) - \theta_0 \right) \xrightarrow{d} N(0, \sigma_\lambda^2(g)).
\]

Thus by the Slutsky theorem and (A.7), as \( n \to \infty \)
\[
n\lambda^2 \xrightarrow{d} \hat{\lambda}^2 \left( \int \frac{g^2(x) \, dA_0(x)}{(1 - F_0(x))(1 - G_0(x))} \right)^{-1}.
\]
(A.11)

REFERENCES


