A NOTE ON NONPARAMETRIC LIKELIHOOD RATIO ESTIMATION OF SURVIVAL PROBABILITIES FOR CENSORED DATA

by

Gang Li
Purdue University

Technical Report #93–24

Department of Statistics
Purdue University

May 1993

A Note on Nonparametric Likelihood Ratio Estimation of Survival Probabilities for Censored Data

Gang Li
Department of Statistics
Purdue University

May 1993

Abstract

A common approach to the problem of estimating survival probabilities from censored data is to construct confidence limits using Kaplan-Meier's estimator and Greenwood's formula via a normal approximation. However, this method is not satisfactory for small samples and has a drawback that it often produces intervals that contain values less than zero or greater than one. Thomas and Grunkemeier (1975) proposed an alternative approach which uses a nonparametric likelihood ratio function to derive interval estimates. Their procedure always yields confidence limits between 0 and 1, and the resulting intervals were shown to have better coverage probabilities by some simulation studies. Heuristic arguments have been used in the literature to support this procedure. In this note we give a rigorous justification for the Thomas and Grunkemeier method.

Keywords and phrases: censoring; Kaplan-Meier's estimator; nonparametric likelihood ratio; survival probability.

1 Introduction and Summary

Let $X_1, \ldots, X_n$ be n independent positive random variables with a common unknown survival function $S(t) = P(X_i > t)$ and $Y_1, \ldots, Y_n$ be n independent positive random variables with a common unknown survival function $G(t) = P(Y_i > t)$. Assume that the X's and Y's are independent. A randomly censored data set consists of n i.i.d. pairs $(Z_1, \delta_1), \ldots, (Z_n, \delta_n)$ where $Z_i = \min\{X_i, Y_i\}$ and $\delta_i = I(X_i \leq Y_i)$ for $i = 1, \ldots, n$. In the context of survival analysis, X_i and Y_i usually refer to the life time and censoring time of the ith individual under study respectively. For example, X and Y may denote age at death and age at the end of study respectively. An important problem in survival analysis is to find confidence interval estimates for $S(a)$ ($a > 0$ is fixed), the proportion of subjects in the population whose life time would exceed a, on the basis of randomly censored observations.

A common approach to this problem is to construct confidence limits using Kaplan-Meier's (1958) estimator and "Greenwood's Formula" via a normal approximation. Kaplan-Meier's estimator of the survival function S may be derived by considering the likelihood function

$$L(S) = \prod_u \left[S(Z_i-) - S(Z_i)\right] \prod_c S(Z_i) \tag{1.1}$$

where the two products are taken over uncensored and censored individuals respectively. Let there be k distinct observed uncensored survival times $0 = T_0 < T_1 < \ldots < T_k < T_{k+1} = \infty$. Denote by r_j and d_j the number of subjects that are at risk and the number of deaths occurred at time T_j respectively ($r_0 \equiv n$, $r_{k+1} \equiv 0$, and $d_0 \equiv 0$). It was shown by Kaplan and Meier (1958) that a maximizer of the likelihood function in (1.1) must have support $\{T_1, \ldots, T_k\}$, and for such S

$$\begin{aligned} L(S) &= \prod_{j=1}^{k} \left[S(T_j-) - S(T_j)\right]^{d_j} \prod_{j=0}^{k} \left[S(T_j)\right]^{r_j - r_{j+1} - d_j} \\ &= \prod_{j=1}^{k} \left[S(T_{j-1}) - S(T_j)\right]^{d_j} \prod_{j=0}^{k} \left[S(T_j)\right]^{r_j - d_j} \prod_{j=1}^{k+1} \left[S(T_{j-1})\right]^{-r_j} \\ &= \prod_{j=1}^{k} \left[1 - \frac{S(T_j)}{S(T_{j-1})}\right]^{d_j} \left[\frac{S(T_j)}{S(T_{j-1})}\right]^{r_j - d_j} \end{aligned}$$

where the first equality is from the facts that $S(Z_i) = S(T_j)$ for $Z_i \in [T_j, T_{j+1})$ and that $r_j - r_{j+1} - d_j$ is the number of censored observations in the interval $[T_j, T_{j+1})$, and the third equality follows from $S(T_0) = S(0) = 1$ and $r_{k+1} = 0$. Therefore

$$L(S) = \prod_{j=1}^{k} h_j^{d_j} (1 - h_j)^{r_j - d_j} \tag{1.2}$$

where $h_j = 1 - S(T_j)/S(T_{j-1})$ for $j = 1, \ldots, k$. It is important to note that $S(T_j) = \prod_{i \leq j}(1 - h_i)$, by solving for $S(T_j)$, $j = 1, \ldots, k$, recursively. Kaplan-Meier's estimator

S_n of S is defined to be a maximizer of (1.2) and is given by

$$S_n(t) = \prod_{j=1}^{K(t)} (1 - \hat{h}_j), \quad 0 \leq t < \infty, \tag{1.3}$$

where $\hat{h}_j = d_j/r_j$ for $j = 1, \ldots, k$ and $K(t)$ denotes the number of distinct uncensored observations on the time interval $[0, t]$. Kaplan-Meier's estimator of G is defined in a symmetric way. It is well known that under mild conditions (see e.g. Theorem 4.2.2 of Gill (1980))

$$n^{\frac{1}{2}}\{S_n(a) - S(a)\} \xrightarrow{d} N(0, \{S(a)\}^2\sigma^2) \quad \text{as } n \to \infty,$$

where

$$\sigma^2 = -\int_0^a \frac{dS(u)}{S(u)S(u-)G(u-)} \tag{1.4}$$

can be consistently estimated by

$$\hat{\sigma}^2 = \sum_{j=1}^{K(a)} \frac{d_j}{r_j(r_j - d_j)}. \tag{1.5}$$

(We note that $\hat{\sigma}^2$ may be obtained by replacing S and G with their Kaplan-Meier's estimators in (1.4) respectively and the formula in (1.5) is usually referred to as "Greenwood's Formula".) One can therefore construct confidence intervals using

$$Z = \frac{n^{\frac{1}{2}}\{S_n(a) - S(a)\}}{\{S_n(a)\}^2\hat{\sigma}^2} \tag{1.6}$$

which has a standard normal limiting distribution.

Thomas and Grunkemeier (1975) pointed out that interval estimates obtained from (1.6) are not satisfactory for small samples and have a drawback that they may contain values less than zero or greater than one. They proposed an alternative approach that uses a nonparametric likelihood ratio function to set confidence limits. To describe the Thomas and Grunkemeier procedure, we restrict our attention to survival functions that have finite support on $T_1, \ldots, T_k$, i.e. $S \ll S_n$. For each p, $0 < p < 1$, define a nonparametric likelihood ratio function

$$R(p) = \frac{\sup\{L(S) | S(a) = p \text{ and } S \ll S_n\}}{L(S_n)} \tag{1.7}$$

where S_n is the Kaplan-Meier estimator defined by (1.3) and L is the likelihood function defined by (1.2). More explicitly,

$$R(p) = \frac{\sup\left\{\prod_{j=1}^{k} h_j^{d_j}(1 - h_j)^{r_j - d_j} \,\middle|\, \prod_{j=1}^{K(a)} (1 - h_j) = p\right\}}{\prod_{j=1}^{k} \hat{h}_j^{d_j}(1 - \hat{h}_j)^{r_j - d_j}} \tag{1.8}$$

where $\hat{h}_j = d_j/r_j$ for $j = 1, \ldots, k$. Thomas and Grunkemeier (1975) suggested that

$$\{p: \; -2 \ln R(p) \leq \chi_1^2(\alpha)\} \tag{1.9}$$

be a confidence set for $S(a)$ of approximate level α, where $\chi_1^2(\alpha)$ is the $(1-\alpha)^{th}$ quantile of the chi-square distribution with one degree of freedom. They further showed that the set produced by (1.9) is always a closed interval and never include values outside [0,1]. Their simulation results also indicate that, for small sample sizes, (1.9) has better coverage probability than that obtained from (1.6). These properties make the nonparametric likelihood ratio method very appealing.

In their paper, Thomas and Grunkemeier (1975) only used heuristic arguments to support their procedure. An informal derivation is also given in Section 4.3 of Cox and Oakes (1984). Since this nonparametric likelihood ratio method has nice properties and seems to be the one that should be recommended in practice, it is desirable to have a theoretical justification for it. To our knowledge, there has not been one in the literature. The purpose of this short note is to provide a rigorous justification for the Thomas and Grunkemeier (1975) method. We prove the following.

Theorem 1 *Assume that S is continuous, $0 < S(a) < 1$, and $G(a) > 0$. Then, for $p = S(a)$,*

$$-2 \ln R(p) \xrightarrow{d} \chi_1^2 \quad \text{as } n \to \infty,$$

where χ_1^2 is a random variable having the chi-square distribution with one degree of freedom.

It is clear from Theorem 1 that the likelihood ratio set given by (1.9) consists of all values p for which the null hypothesis $H_0 : S(a) = p$ is rejected by a likelihood ratio test at the significance level α. One may also interpret the set given by (1.9) from another perspective. Define

$$C(r) = \Big\{S(a) \Big| S \ll S_n \text{ and } \frac{L(S)}{L(S_n)} > r\Big\}. \tag{1.10}$$

Noticing the fact that $R(p) = \sup\Big\{L(S)/L(S_n)\Big|S(a) = p \text{ and } S \ll S_n\Big\}$, it is not hard to see that

$$p \in C(r) \;\text{ if and only if }\; R(p) > r.$$

Therefore

$$\{p: \; -2 \ln R(p) \leq \chi_1^2(\alpha)\} = C(r) \quad \text{for} \quad r = \exp\Big\{-\frac{1}{2}\chi_1^2(\alpha)\Big\}.$$

If we think of the likelihood ratio $L(S_n)/L(S)$ as a "distance" between S_n and S, then $C(r)$ has an intuitive interpretation that it consists of the survival probabilities $S(a)$ for all distributions that have finite support on the uncensored observations and fall inside a "neighborhood" of the Kaplan-Meier estimator S_n.

The first rigorous treatment of nonparametric likelihood ratio methods is due to Owen (1990) and the literature on this topic has grown rapidly in recent years. Owen

(1990) studied the problem of estimating the expected value of a random vector and some of its "smooth" functions using an empirical likelihood ratio function under the classical i.i.d. setting. His results have been further extended to deal with problems arising from more general models such as linear regression, generalized linear regression, and projection pursuit. Refer to Owen (1990, 1991, 1992), DiCiccio et al (1991), Hall (1990), and Kolaczyk (1992) for more discussions on these topics.

In Section 2 we give a proof of Theorem 1. Details on the computation of confidence limits are reviewed in the last section. One may wonder whether Theorem 1 can be easily proved by making cosmetic changes in existing works, for example, the proof of Theorem 1 of Owen (1990). This is not the case. It is seen that calculation of the likelihood ratio defined in (1.8) involves a constrained optimization problem which can be solved using the Lagrangian multiplier principle. Because the constraint is different, it distinguishes this problem from others. As one will see in Section 2, a crucial step in proving Theorem 1 is to establish the asymptotic rate of a Lagrangian multiplier λ_n determined by (2.2). We note that the asymptotic rate of the Lagrangian multiplier in Owen (1990) is $O_p(n^{-1/2})$ (see (2.14) on page 101 of Owen (1990)). In the current situation, the different nature of constraints leads to a rate of $O_p(n^{1/2})$ (see Lemma 1). A different derivation is required and it uses the nontrivial weak convergence results of the Nelson-Aalen estimator for the cumulative hazard function and that of the Kaplan-Meier estimator. Finally we note that the order of $O_p(n^{1/2})$ occurs in many situations when one is interested in estimating survival probabilities. See Li (1993) for a discussion of nonparametric likelihood ratio estimation for randomly truncated data.

2 Proof of Theorem 1

The proof involves several steps. We first show that $R(p)$ is well defined by (1.8) for $0 < p < 1$. As a consequence an explicit expression of $-2\ln R(p)$ is given. Then in Lemma 1 we obtain the asymptotic rate of the Lagrangian multiplier λ_n determined by (2.2). This result is needed in order to make Taylor expansions for $-2\ln R(p)$. Finally we derive the asymptotic distribution of $-2\ln R(p)$ using Taylor expansions and some weak convergence results related to Kaplan-Meier's estimator.

To show that $R(p)$ is well defined by (1.8), we begin with a known result on the following constrained maximization problem:

$$(\mathcal{P}): \textit{ maximize } f(\boldsymbol{x}) \textit{ subject to } g(\boldsymbol{x}) = b \textit{ and } \boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^k,$$

where f and g are scalar-valued functions defined on the k-dimensional Euclidean space $\mathbb{R}^k$, $\mathcal{X}$ is a subset of $\mathbb{R}^k$, and b is a constant.

Proposition 2.1 *Suppose that a real number λ can be found such that a value maximizing $f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$ in $\mathcal{X}$, say $\boldsymbol{x}^*$, satisfies $g(\boldsymbol{x}^*) = b$. Then $\boldsymbol{x}^*$ is a solution for the constrained maximization problem ($\mathcal{P}$).*

Proof This is a special case of the first part of Theorem (3.1) of Whittle (1971).

Note that, for $0 < p < 1$, $R(p)$ is well defined by (1.8) if the constrained maximization problem

$$(\mathcal{P}1): \quad \textit{maximize}\, L(S) = \prod_{j=1}^{k} h_j^{d_j}(1-h_j)^{r_j-d_j}$$
$$\textit{subject to constraints}$$
$$\prod_{j=1}^{K(a)}(1-h_j) = p \quad \textit{and} \quad 0 \le h_j \le 1 \quad \textit{for} \quad j = 1, \ldots, n$$

has an unique solution.

Recall that r_j is the number of subjects that are at risk at time T_j and d_j is the number of uncensored deaths occurred at time T_j, $j = 1, \ldots, k$. This implies that $r_{j+1} - d_{j+1} \le r_{j+1} \le r_j - d_j$, $j = 1, \ldots, k$, and $\max_{1 \le j \le K(a)}\{d_j - r_j\} = d_{K(a)} - r_{K(a)}$. Therefore for any $\lambda > d_{K(a)} - r_{K(a)}$, $r_j + \lambda > d_j$ for $j = 1, \ldots, K(a)$ and

$$\begin{aligned} & \ln\left(\prod_{j=1}^{k} h_j^{d_j}(1-h_j)^{r_j-d_j}\right) + \lambda \ln\left(\prod_{j=1}^{K(a)}(1-h_j)\right) \\ = \; & \ln\left(\prod_{j=1}^{K(a)} h_j^{d_j}(1-h_j)^{(r_j+\lambda)-d_j}\right) + \ln\left(\prod_{K(a)+1}^{k} h_j^{d_j}(1-h_j)^{r_j-d_j}\right) \end{aligned}$$

is maximized at $\boldsymbol{h}(\lambda) = (h_1(\lambda), \ldots, h_k(\lambda))$, where

$$h_j(\lambda) \equiv \begin{cases} \dfrac{d_j}{r_j+\lambda} & \text{for } j = 1, \ldots, K(a) \\ \dfrac{d_j}{r_j} & \text{for } j = K(a)+1, \ldots, k \end{cases} \tag{2.1}$$

are between 0 and 1.

Moreover, if we define $g(\lambda) = \prod_{j=1}^{K(a)}\{1 - d_j/(r_j+\lambda)\}$, then $g(\lambda)$ is strictly increasing in λ and has the properties that $\lim_{\lambda \to d_{K(a)} - r_{K(a)}} g(\lambda) = 0$ and that $\lim_{\lambda \to \infty} g(\lambda) = 1$. Thus the equation

$$\prod_{j=1}^{K(a)}\left(1 - \frac{d_j}{r_j+\lambda}\right) = p \tag{2.2}$$

has an unique solution, say λ_n, on the interval $(d_{K(a)} - r_{K(a)}, \infty)$.

So we have proved that $\boldsymbol{h}(\lambda_n) = (h_1(\lambda_n), \ldots, h_k(\lambda_n))$ defined by (2.1) and (2.2) maximizes

$$\ln\left(\prod_{j=1}^{k} h_j^{d_j}(1-h_j)^{r_j-d_j}\right) + \lambda_n \ln\left(\prod_{j=1}^{K(a)}(1-h_j)\right)$$

and satisfies

$$\prod_{j=1}^{K(a)}\{1 - h_j(\lambda_n)\} = p \quad \text{and} \quad 0 < h_j(\lambda_n) < 1 \quad \text{for} \quad j = 1, \ldots, K(a). \tag{2.3}$$

This, together with Proposition 2.1 and the uniqueness of λ_n, implies that $(\mathcal{P}1)$ has an unique solution $\boldsymbol{h}(\lambda_n)$ given by (2.1) and (2.2). We also have

$$-2\ln R(p) = -2\sum_{j=1}^{K(a)}\Big\{(r_j - d_j)\ln\Big(1+\frac{\lambda_n}{r_j-d_j}\Big) - r_j\ln\Big(1+\frac{\lambda_n}{r_j}\Big)\Big\} \tag{2.4}$$

where λ_n is uniquely determined by (2.2).

To obtain the limiting distribution of $-2\ln R(p)$, we shall need to make Taylor expansions for the expressions in (2.4). This requires the knowledge of the asymptotic behavior of λ_n. The following lemma establishes the asymptotic rate of λ_n as $n \to \infty$.

Lemma 1 *If the assumptions of Theorem 1 hold, then $\lambda_n = O_p(n^{1/2})$ as $n \to \infty$.*

Proof We first state some asymptotic results related to Kaplan-Meier's estimator, which will be needed later in the proof. Recall that the cumulative hazard function of X is defined to be $A(t) = -\int_0^t \{S(s-)\}^{-1} dS(s)$, $t \geq 0$, and that $A = -\ln S$ when S is continuous. A can be estimated by the Nelson-Aalen estimator (see e.g. page 92 of Fleming and Harrington (1991)) that can be written as

$$A_n(t) = \sum_{j=1}^{K(t)} \frac{d_j}{r_j}.$$

Under the assumptions of Theorem 1 (see e.g. Theorem 4.2.2 of Gill (1980))

$$n^{\frac{1}{2}}\{A_n(a) - (-\ln p)\} \xrightarrow{d} N(0, \sigma^2) \tag{2.5}$$

and

$$n^{\frac{1}{2}}\{S_n(a) - p\} \xrightarrow{d} N(0, \{S(a)\}^2\sigma^2),$$

where σ^2 is defined by (1.4). Furthermore, the $\hat{\sigma}^2$ defined in (1.5) converges to σ^2 in probability. An application of δ-method also gives

$$n^{\frac{1}{2}}\{\ln S_n(a) - \ln p\} \xrightarrow{d} N(0, \sigma^2). \tag{2.6}$$

Now we derive the asymptotic rate of λ_n. Recall that for each sample realization λ_n is the unique solution of (2.2) on the interval $(d_{K(a)} - r_{K(a)}, \infty)$. We shall obtain an upper bound for λ_n under each of the two possible cases $d_{K(a)} - r_{K(a)} < \lambda_n < 0$ and $\lambda_n \geq 0$ separately. (Note that we shall only have the latter case if $d_{K(a)} - r_{K(a)} = 0$.)

If $d_{K(a)} - r_{K(a)} < \lambda_n < 0$, then $|\lambda_n| < r_{K(a)} \leq n$ and $0 \leq d_j/(r_j + \lambda_n) < 1$ for $j = 1, \ldots, K(a)$ since $d_{K(a)} - r_{K(a)} = \max_{1\leq j \leq K(a)}\{d_j - r_j\}$. Because

$$-\ln(1-x) \geq x \quad \text{for } 0 \leq x < 1,$$

we have

$$\begin{aligned}
-\ln p &= -\sum_{j=1}^{K(a)} \ln\Big(1-\frac{d_j}{r_j+\lambda_n}\Big)\\
&\geq \sum_{j=1}^{K(a)} \frac{d_j}{r_j+\lambda_n}\\
&= \sum_{j=1}^{K(a)} \Big(\frac{d_j}{r_j}\Big)\Big(\frac{r_j}{r_j-|\lambda_n|}\Big)\\
&\geq \sum_{j=1}^{K(a)} \Big(\frac{d_j}{r_j}\Big)\Big(\frac{n}{n-|\lambda_n|}\Big)\\
&= A_n(a)\Big(\frac{n}{n-|\lambda_n|}\Big).
\end{aligned}$$

This implies that

$$|\lambda_n| \leq \frac{n\{-\ln p - A_n(a)\}}{-\ln p}. \tag{2.7}$$

If $\lambda_n \geq 0$, then

$$\ln\Big(1-\frac{d_j}{r_j+\lambda_n}\Big)+\frac{d_j}{r_j+\lambda_n} \geq \ln\Big(1-\frac{d_j}{r_j}\Big)+\frac{d_j}{r_j},$$

since $\ln(1-x)+x$ is decreasing on the interval $(0,1)$ and $d_j/(r_j+\lambda_n) \leq d_j/r_j$ for $j=1,\ldots,K(a)$. Thus

$$\begin{aligned}
\ln p &= \sum_{j=1}^{K(a)} \ln\Big(1-\frac{d_j}{r_j+\lambda_n}\Big)\\
&\geq \sum_{j=1}^{K(a)} \Big\{-\frac{d_j}{r_j+\lambda_n}+\ln\Big(1-\frac{d_j}{r_j}\Big)+\frac{d_j}{r_j}\Big\}\\
&= -\sum_{j=1}^{K(a)} \Big(\frac{d_j}{r_j}\Big)\Big(\frac{r_j}{r_j+|\lambda_n|}\Big)+\ln S_n(a)+A_n(a)\\
&\geq -\sum_{j=1}^{K(a)} \Big(\frac{d_j}{r_j}\Big)\Big(\frac{n}{n+|\lambda_n|}\Big)+\ln S_n(a)+A_n(a)\\
&= -A_n(a)\Big(\frac{n}{n+|\lambda_n|}\Big)+\ln S_n(a)+A_n(a),
\end{aligned}$$

which implies that

$$|\lambda_n| \leq \frac{n\{\ln S_n(a)-\ln p\}}{A_n(a)+\ln S_n(a)-\ln p}. \tag{2.8}$$

Therefore, $\lambda_n = O_p(n^{1/2})$ follows from (2.5)–(2.8).

Now we proceed to prove Theorem 1.

Proof of Theorem 1

We first show that

$$\lambda_n = \Big(\sum_{j=1}^{K(a)} \frac{d_j}{r_j(r_j-d_j)}\Big)^{-1}\Big(\ln S_n(a) - \ln p + O_p(n^{-\frac{1}{2}})\Big). \tag{2.9}$$

By the Glivenko-Cantelli Theorem, $r_j = \sum_{i=1}^n I(Z_i \geq T_j) = O_p(n)$ uniformly in $j = 1, \ldots, k$.

Since S is continuous, the probability that there is no tie among $X_1, \ldots, X_n$ is 1. Thus $d_j = O_p(1)$ uniformly in $j = 1, \ldots, k$.

Observe that $K(a) = \sum_{i=1}^n I(Z_i \leq a, \delta_i = 1)$ with probability 1. Thus $K(a) = O_p(n)$ by the Law of Large Numbers.

Hence, $\lambda_n/r_j = O_p(n^{-1/2})$ and $d_j/r_j = O_p(1/n)$ uniformly in $j = 1, \ldots, k$, and

$$\begin{aligned}
\ln p &= \sum_{i=1}^{K(a)} \ln\Big(1 - \frac{d_j}{r_j+\lambda_n}\Big) \\
&= \sum_{i=1}^{K(a)} \ln\Big\{1 - \Big(\frac{d_j}{r_j}\Big)\Big(\frac{1}{1+\frac{\lambda_n}{r_j}}\Big)\Big\} \\
&= \sum_{i=1}^{K(a)} \ln\Big\{1 - \frac{d_j}{r_j}\Big(1 - \frac{\lambda_n}{r_j} + O_p(\frac{1}{n})\Big)\Big\} \\
&= \sum_{i=1}^{K(a)} \ln\Big\{1 - \frac{d_j}{r_j} + \frac{d_j\lambda_n}{r_j^2} + O_p(\frac{1}{n^2})\Big\} \\
&= \sum_{i=1}^{K(a)} \ln\Big(1 - \frac{d_j}{r_j}\Big) + \sum_{i=1}^{K(a)} \ln\Big(1 + \frac{\frac{d_j\lambda_n}{r_j^2} + O_p(\frac{1}{n^2})}{1-\frac{d_j}{r_j}}\Big) \\
&= \ln S_n(a) + \sum_{i=1}^{K(a)} \Big\{-\frac{\frac{d_j\lambda_n}{r_j^2} + O_p(\frac{1}{n^2})}{1-\frac{d_j}{r_j}} + O_p(n^{-\frac{3}{2}})\Big\} \\
&= \ln S_n(a) - \lambda_n \sum_{i=1}^{K(a)} \frac{d_j}{r_j(r_j-d_j)} + O_p(n^{-\frac{1}{2}})
\end{aligned}$$

where in the third step we have used the fact that $1/(1+x) = 1 - x + O(x^2)$ as $x \to 0$ and the sixth equality follows from the facts that $\{\frac{d_j\lambda_n}{r_j^2} + O_p(\frac{1}{n^2})\}/(1-\frac{d_j}{r_j}) = O_p(n^{-3/2})$ and that $\ln(1+x) = -x + O(x^2)$ as $x \to 0$. This implies (2.9) immediately.

Because $\lambda_n/(r_j - d_j) = O_p(n^{-1/2})$ and $\lambda_n/r_j = O_p(n^{-1/2})$ uniformly in $j = 1, \ldots, k$, and $\ln(1+x) = x - x^2/2 + O(x^3)$ as $x \to 0$, we have

$$\begin{aligned}
-2\ln R(p) &= -2\sum_{j=1}^{K(a)} \Big\{(r_j-d_j)\ln\Big(1+\frac{\lambda_n}{r_j-d_j}\Big) - r_j\ln\Big(1+\frac{\lambda_n}{r_j}\Big)\Big\} \\
&= -2\sum_{j=1}^{K(a)} \Big[(r_j-d_j)\Big\{\frac{\lambda_n}{r_j-d_j} - \frac{1}{2}\Big(\frac{\lambda_n}{r_j-d_j}\Big)^2 + O_p(n^{-\frac{3}{2}})\Big\}
\end{aligned}$$

$$-r_j\Big\{\frac{\lambda_n}{r_j}-\frac{1}{2}\Big(\frac{\lambda_n}{r_j}\Big)^2+O_p(n^{-\frac{3}{2}})\Big\}\Big]$$

$$= \lambda_n^2\Big(\sum_{j=1}^{K(a)}\frac{d_j}{r_j(r_j-d_j)}\Big)+O_p(n^{-\frac{1}{2}})$$

$$= \Big(\sum_{j=1}^{K(a)}\frac{d_j}{r_j(r_j-d_j)}\Big)^{-1}\Big\{\ln S_n(a)-\ln p+O_p(n^{-\frac{1}{2}})\Big\}^2+O_p(n^{-\frac{1}{2}})$$

where the last step is obtained by replacing λ_n with the expression in (2.9). This, together with (2.6) and the consistency of $\hat{\sigma}^2$, implies that

$$-2\ln R(p)\xrightarrow{d}\chi_1^2.$$

3 Computation of Confidence Limits

We give a brief review on the calculation of confidence limits set by (1.9). (Also see Appendix B of Thomas and Grunkemeier (1975) whose notations are slightly different from ours.) In the following discussion we shall consider $-2\ln R$ as a function of λ:

$$-2\ln R(\lambda)=-2\sum_{j=1}^{K(a)}\Big\{(r_j-d_j)\ln\Big(1+\frac{\lambda}{r_j-d_j}\Big)-r_j\ln\Big(1+\frac{\lambda}{r_j}\Big)\Big\}$$

on the interval $[d_{K(a)}-r_{K(a)},\infty)$. It is also useful to note that $d_{K(a)}-r_{K(a)}<0$ is equivalent to $S_n(a)=\prod_{j=1}^{K(a)}(1-d_j/r_j)>0$ since $d_{K(a)}-r_{K(a)}=\max_{1\le j\le K(a)}\{d_j-r_j\}$.

If $d_{K(a)}-r_{K(a)}<0$, i.e. $S_n(a)>0$, then it is easy to check that $-2\ln R(\lambda)$ is strictly decreasing on the interval $(d_{K(a)}-r_{K(a)},0]$ and increasing on $[0,\infty)$. In addition, $-2\ln R(\lambda)\to+\infty$ as $\lambda\to+\infty$, $-2\ln R(\lambda)\to+\infty$ as $\lambda\searrow d_{K(a)}-r_{K(a)}$, and $\ln R(0)=0$. Therefore, the equation

$$-2\ln R(\lambda)=\chi_1^2(1-\alpha),\ (\alpha>0)$$

has exactly one solution on each of the intervals $(d_{K(a)}-r_{K(a)},0)$ and $(0,\infty)$, say λ_L and λ_U, and

$$\{\lambda:\ -2\ln R(\lambda)\le\chi_1^2(1-\alpha)\}\cap[d_{K(a)}-r_{K(a)},\infty)=[\lambda_L,\lambda_U].$$

If $d_{K(a)}-r_{K(a)}=0$, i.e. $S_n(a)=0$, then,

$$\{\lambda:\ -2\ln R(\lambda)\le\chi_1^2(1-\alpha)\}\cap[d_{K(a)}-r_{K(a)},\infty)=[\lambda_L,\lambda_U]$$

where $\lambda_L=0$ and λ_U is the unique solution of the equation

$$-2\ln R(\lambda)=\chi_1^2(1-\alpha),\ (\alpha>0)$$

on the interval $(0,\infty)$.

In both cases, confidence limits for $p = S(a)$ are set by

$$p_L = \prod_{j=1}^{K(a)} \left(1 - \frac{d_j}{r_j + \lambda_L}\right)$$

and

$$p_U = \prod_{j=1}^{K(a)} \left(1 - \frac{d_j}{r_j + \lambda_U}\right).$$

References

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data.* Chapman and Hall, London.

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

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis.* John Wiley and Sons, New York.

Gill, R. D. (1980). *Censoring and Stochastic Integrals.* Math. Centre Tracts 124, Mathematical Centre, Amsterdam.

Hall, P. (1990). Pseudo-likelihood theory for empirical likelihood. *Ann. Statist.* ***18*** 121-140.

Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* ***53***, 457-481.

Kolaczyk, E. D. (1992). Empirical likelihood for generalized linear models. Technical Report No. 389. Department of Statistics, Stanford University.

Li, G. (1993). Non-parametric and semi-parametric likelihood ratio estimation of survival probabilities from truncated data. (Preliminary draft).

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

Owen, A. (1991). Empirical likelihood for linear models. *Ann. Statist.* ***19***, 1725-1747.

Owen, A. (1992). Empirical likelihood and generalized projection pursuit. Technical Report No. 393. Department of Statistics, Stanford University.

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

Whittle, P. (1971). *Optimization under Constraints.* John Wiley and Sons, London.