EMPIRICAL ENVELOPE MLE AND LR TESTS

Mai Zhou

University of Kentucky

Summary

We study in this paper some nonparametric inference problems where the nonparametric maximum likelihood estimator (NPMLE) are not well defined. However, if we *enlarge* the parameter space, the NPMLE will be well defined. We propose to gradually shrink the enlarged parameter space by placing more and more restrictions on the parameter space, producing a sequence of (envelope) estimators. The approach is a counter part of the sieve MLE (Grenander, 1981).

Several different problems where this method can be applied effectively are discussed. The detailed treatment of 2-sample location problem is presented, including a Wilks type theorem for the empirical envelope likelihood ratio statistic and the asymptotic distribution of the empirical envelope MLE of location.

AMS 1991 Subject Classification: Primary 62G10; secondary 62G05.

Key Words and Phrases: K-sample Problem, Many Constraints, Symmetric Distributions, Weighted Hazards, Chi-square distribution.

1. Introduction

The maximum likelihood estimation method, when applied in the nonparametric settings, often have difficulties. The maximizer often do not exist or there may be many maximizers. Estimating a density function from a sample of i.i.d. observations is one such example. The method of sieve, (Grenander 1981, Geman and Huang 1982) is developed to get around this by adopting a sequence of smaller but growing parameter spaces that approximates the original parameter space. On the smaller parameter spaces, the MLEs are well defined.

In this paper, we consider empirical likelihood (Owen 1988, Thomas and Grunkemeier 1975). The Empirical Likelihood (EL) method parallels the parametric maximum likelihood method, except the assumption of a density is dropped and the likelihood function is replaced by the Empirical Likelihood. Many research papers have appeared since and all of them testify that the EL method is a very competitive and promising nonparametric statistical inference method, see the recent book by Owen (2001).

However, in some important cases, the maximum empirical likelihood estimator do not exist or there exist many maximizers of the empirical likelihood, and the empirical likelihood ratio test of Owen (1988) cannot be used. We propose an estimation method called empirical envelope likelihood. It *enlarges* the parameter space to make the maximum empirical likelihood estimator well defined and then gradually shrink the enlarged parameter space as sample size grows by putting more and more constraints on the parameter space.

To explain the idea clearly we start with the simplest 2-sample location problem and indicate later a variety of problems that can be solved similarly by the envelope MLE method.

2. The Location Problem, and the Proposed Solution

For n i.i.d. observations the log empirical likelihood is

$$\sum \log w_i \ ,$$

where $w_i > 0$ and $\sum w_i = 1$. See Owen (1988), (1990) among others. When there are two independent samples of size n and m, each with its own distribution, the log empirical likelihood is

$$\log Lik = \sum_{i=1}^{n} \log w_i + \sum_{j=1}^{m} \log p_j \ , \tag{1}$$

where $w_i > 0$, $\sum w_i = 1$ and $p_j > 0$, $\sum p_j = 1$.

The following 2-sample location shift problem presents some difficult for the empirical likelihood/NPMLE method.

Original Model: Suppose $X_1, \cdots, X_n$ are i.i.d. $F(t)$ and independently $Y_1 \cdots, Y_m$ are i.i.d. $F(t-\theta)$, where $\theta \in R^1$ is the parameter of interest. The distribution $F(t)$ is also unknown and arbitrary (if needed we may assume the mean is finite), which can be thought of as an infinite dimensional nuisance parameter.

If $F(t)$ is a continuous distribution there will be $n \times m$ maximizers of the log empirical likelihood (1), and those maximizers do not converge as sample size grows. Therefore neither of the NPMLE of F and θ exist.

Enlarged model: If we enlarge the parameter space by

$$\text{replace} \qquad F(t-\theta) \qquad \text{with} \qquad G(t)$$

where G is also arbitrary, then the NPMLE of F and G both exist based on the sample of X's and Y's (they are the empirical distributions). Here the parameter space is enlarged from $\Theta_1 = (F(\cdot), \theta)$ to $\Theta_2 = (F(\cdot), G(\cdot))$. The estimator of location parameter $\hat\theta$ in the enlarged parameter space can be obtained, for example, as the difference of the 2 means or medians of $\hat{F}$ and $\hat{G}$, which is the same as the difference of the two sample means/medians.

This estimator of θ is consistent but not efficient since it did not take into account the information that X and Y sample have the same distribution, except a shift.

We propose a method of estimation (for θ and F) which we will call it the empirical envelope MLE. It uses the information that X and Y sample have same distribution, except a shift. It proceeds as follows.

Since the NPMLE do exist when we enlarge the parameter space, we shall start from there and try to shrink the parameter space gradually as the sample size increases. The idea is similar to the sieve MLE of Grenander (Grenander, 1981, see also Geman and Huang 1982) but with a sequence of shrinking envelopes (larger parameter spaces) instead of growing sieves (smaller parameter spaces).

Specifically, we shall impose constraints on Θ_2 that the two distributions F and G satisfy the following: for some given t_i's

$$F(t_i) = G(t_i + \theta), \qquad i = 1, 2, \cdots, k; \qquad \text{for some } \theta \in R^1. \tag{2}$$

We show that the maximizer of log likelihood (1) over Θ_2 with constraint (2) (i.e. the constrained NPMLE) is well defined. Other types of constraints are certainly possible and may have some advantages due to smoothness, see (5), (6) and Lemma 2 later.

If we increase the number of t_i points (thus the number of constraints) as sample size grows and let t_i's become dense in the support of F, then in the limit we have the constraint: $F(t) = G(t + \theta)$ for all t. The parameter space Θ_2 with these constraints falls back to the parameter space Θ_1.

The choice of the points t_i (or the choice of the $g_i(\cdot)$ functions in (5), (6)) in the finite sample is a problem similar to the choice of kernel/band-width in the density estimation.

The value $\hat{\theta}$, (and $\hat{F}$ and $\hat{G}$) that maximize the log likelihood among $\theta \in R^1$ and all possible F, G distributions that satisfy (2), is our **envelope MLE of** θ. The value $\hat{F}(t_i) \equiv \hat{G}(t_i + \hat{\theta})$ can also be thought of as the envelope MLE of $F(t_i)$.

Testing hypothesis for the parameter θ can be accomplished by the empirical likelihood ratio and Wilks type theorem.

To be specific let us treat a testing hypothesis problem of

$$H_0 : \theta = \theta_0 \qquad vs \qquad H_A : \theta \neq \theta_0$$

without loss of generality we take $\theta_0 = 0$.

The test statistic we propose is the likelihood ratio statistic

$$T_1 = -2\{\max \log Lik(with(2), \theta = 0) - \max \log Lik(with(2), \theta = \hat{\theta})\} \ . \tag{3}$$

This test statistic have an asymptotic chi-square distribution with 1 degree of freedom under null hypothesis. We reject H_0 for larger values of T_1.

The two log likelihood functions in (3) are defined as follows: the first term, $\max \log Lik$ $(with(2), \theta = 0)$, is maximized over Θ_2 under the constraint (2) but with $\theta \equiv 0$ in there. The second term, $\max \log Lik(with(2), \theta = \hat{\theta})$, is maximized over Θ_2 under the constraint (2) where θ is also maximized over R^1. As we defined above, the maximum over θ is assumed to attain at $\hat{\theta}$, and some $\hat{F}, \hat{G}$, where they satisfy $\hat{F}(t_i) = \hat{G}(t_i + \hat{\theta})$.

2.1 Reformulation in terms of hazard

Next we reformulate the problem in terms of (cumulative) hazard function instead of distribution function. It turns out that computations are much easier with the equivalent hazard formulation, especially for proportional hazard type constraints. (Fang and Zhou 2000). A second advantage of hazard formulation is that it can accommodate right censored data readily (Pan and Zhou 1999).

The log likelihood function in terms of hazard for the two independent samples is

$$\log_H Lik = \sum_i [d_{1i} \log v_i + (R_{1i} - d_{1i}) \log(1 - v_i)] + \sum_j [d_{2j} \log u_j + (R_{2j} - d_{2j}) \log(1 - u_j)] \tag{4}$$

where d_{1i} is the number of uncensored observations at x_i — the ordered X_i's, and R_{1i} is the number of X observations that are $\geq x_i$, etc. and $1 \geq v_i > 0$, $1 \geq u_j > 0$ are the discrete hazards at x_i and y_j respectively. See, for example, Thomas and Grunkemeier (1975), Li (1995) for similar notation.

The constraints we impose on the hazards v_i and u_j are: for given $g_1(\cdot), \cdots, g_k(\cdot)$ functions

$$\sum g_1(x_i) \log(1 - v_i) = \sum g_1(y_j - \theta) \log(1 - u_j) \ , \tag{5}$$

$$\cdots \quad \cdots$$

$$\sum g_k(x_i)\log(1-v_i) = \sum g_k(y_j-\theta)\log(1-u_j)\ . \tag{6}$$

If we take the function $g_1(t) = I_{[t\le t_1]}$ in the above then the first constraint becomes

$$\prod_{x_i\le t_1}(1-v_i) = \prod_{y_j-\theta\le t_1}(1-u_j)\ .$$

Since for discrete distributions $1-F(t) = \prod_{x_i\le t}(1-v_i)$, this is equivalent to the constraint we imposed on the distributions in (2). We in fact shall allow $g_i(\cdot)$ to be any predictable (or left continuous) functions for generality.

Remark: We could even use two different $g(\cdot)$ in one constraint equation. For example, in equation (5) if we take $g_1(x_i) = I_{[x_i\le t_1]}$ and replace $g_1(y_j-\theta)$ by $\alpha\cdot I_{[y_j-\theta\le t_1]}$ then it is easy to verify that we have the constraint

$$1-F(t_1) = [1-G(t_1+\theta)]^\alpha\ .$$

This is a *shift plus proportional hazard* constraint. We will discuss this model in a bit more detail in section 4. If we let $\alpha = 1$ then this is the same shift constraint as in (2). Notice that this constraint would be awkward to work with by using the distribution formulation (2) because of the nonlinearity of the constraint in terms of probabilities.

The test statistic in terms of hazards is

$$T_2 = -2\{\log_H \max Lik(with\ constraint\ (5,6),\theta=0) - \log_H \max_\theta Lik(with\ constraint\ (5,6),\theta)\}\ .$$

The term $\log_H \max Lik(with(5,6),\theta=0)$ can be easily computed numerically. The computation of the second term is a little more involved. But we can easily get an asymptotic expression for $\hat\theta$ (see Theorem 2).

The 2-sample location problem discussed here is a basic problem and there are many other possible nonparametric tests available for the location parameter in the original model. They include the Wilcoxon type tests based on ranks. We chose to present the idea of the envelope empirical likelihood method in this simple setting so that we can see clearly how it works. The envelope empirical likelihood method clearly applies to other more complicated cases. For example, it is non-trivial to generalize the ranks to the multivariate data case and/or censored data case. In contrast, the proposed empirical envelope MLE is easy to generalize to multivariate/censored data case, and have high efficiency. The Wilcoxon test is not always efficient. Besides, the confidence interval for θ based on the empirical likelihood ratio test has all the inherited nice properties of the likelihood ratio method (see Owen 1988, 1990, 2001).

3. Generality of the Problem/Method

In this section we discuss some other nonparametric problems where the envelope MLE method can apply.

3.1 Symmetric Distributions

Suppose a sample of n i.i.d. observations have a symmetric distribution, but otherwise both the point of symmetry (θ) and the (symmetric) distribution (F) are unknown and arbitrary. Maximizing the log empirical likelihood, $\sum \log w_i$, among all symmetric distributions based on the sample always yields many ($\approx n$) candidates for F and θ, when the true distribution is continuous. They all have the same likelihood value but are far apart. Therefore the NPMLE do not exist.

Empirical envelope MLE method can also be used here to get valid estimation/testing procedures. Professor Owen told me that this problem was considered by Qu for one dimensional data with no censoring and with moment constraints.

To use the envelope MLE method, we proceed first by enlarge the parameter space to all distributions, symmetric or not. The NPMLE exists now and is the empirical distribution function.

The envelope MLE method then calls for shrinking the parameter space by putting constraints like these on the parameter space $\{F: \text{ all distributions}\}$: for given t_i,

$$F(\theta - t_i) = 1 - F(\theta + t_i) \quad i = 1, 2, \cdots k; \qquad \text{for some } \theta \in R^1 .$$

Or we may use smooth constraints like: for given $g_i(\cdot)$,

$$\int g_i(t) dF(\theta - t) = \int g_i(t) d[1 - F(\theta + t)] \quad i = 1, 2, \cdots k .$$

Maximize the log empirical likelihood

$$\sum \log w_i$$

among distributions in the constrained parameter space yields both the empirical envelope MLE $\hat{\theta}$ and $\hat{F}(\hat{\theta} - t_i)$.

Empirical envelope likelihood ratio coupled with Wilks theorem can be used to obtain asymptotically correct α level tests and confidence intervals.

Remark: The efficiency of the estimator $\hat{\theta}$ can be shown to be close to the semiparametric information bound. (Kim and Zhou 2002).

Remark: The problem of estimating the center of a symmetric but unknown density was considered by many others. The approach of envelope MLE, however, do not assume a density. The same approach works for higher dimensional data. With a little effort, the symmetry of a distribution in R^2 can be defined. (van der Vaart 1988).

Other related problems in the symmetric distribution case include: the symmetry equality may only valid for a finite interval:

$$F(\theta - t) = 1 - F(\theta + t) \qquad \text{for } t \in [0, c]$$

or that the equality is valid only after a transformation on the F.

3.2. Location Related Problems

1). Two sample location-scale problems: where

$$x_1, \cdots, x_n \sim F(t) \ ; \qquad\qquad y_1, \cdots, y_m \sim G(t)$$

and

$$G\left(\frac{t+\theta}{\sigma}\right) = F(t) \ .$$

The parameters θ and σ are two parameters of interest and $F(\cdot)$ is treated as infinite dimensional nuisance parameter.

2). Location-transformation problems: similar to above model but

$$G(t + \theta) = h(F(t)) \ .$$

For example the $h()$ could be the proportional hazard transformation.

For example, the treatment of AZT for HIV positive patients may delay the onset of AIDS, as well as prolong the time to death after the onset of AIDS. So the effect of treatment on survival are twofold, a shift of onset time and a proportional hazards change in the distribution of death time after onset of AIDS. Such a parametric model assume exponential distribution is

$$F(t) = 1 - e^{-\lambda_1 t} \ ; \qquad G(t) = 1 - e^{-\lambda_2 (t-\theta)} \ .$$

A nonparametric model of the similar type should be more flexible/desirable and this leads to the *shift proportional hazard model*:

$$1 - F(t) = [1 - G(t-\theta)]^{\lambda} \ . \tag{7}$$

We can estimate θ and λ simultaneously by using the empirical envelope MLE method. The initial envelope would be an arbitrary F and an arbitrary G. We then impose constraints exactly as in (7) above except the equalities hold only at k given times $t_1, t_2, \cdots, t_k$. The estimation of $F(t_i)$ are also available.

3). Censored data or data from biased sampling poses no problem as long as we can compute the constrained NPMLE. In fact Theorem 1 and 2 in next section work without change for right censored data.

4). M samples shift problem, where θ will then be a $(M-1)$ dimensional shift parameter.

5). The observations X and Y may be r dimensional vectors, and thus θ is also a r dimensional vector.

We shall study some of those problems in forthcoming papers.

In general, when the constraint we want to impose on the NPMLE(s) can not be obtained by adjusting the jump size of the un-constrained NPMLE(s), the usual empirical likelihood method will have difficulty. In those cases the empirical envelope MLE method can often be used.

4. Large Sample Results for the Location Envelope MLE

We shall prove in this section that under null hypothesis our proposed test statistic T_2 defined in section 2 has asymptotically a chi-square distribution with one degree of freedom, and obtain the asymptotic distribution for the envelope MLE $\hat{\theta}$.

Denote the column vectors

$$g(t) = \{g_1(t), \cdots, g_k(t)\}^T \ ; \qquad \text{and} \qquad \lambda = \{\lambda_1, \cdots, \lambda_k\}^T \ .$$

Lemma 1 *The hazards that maximize the log likelihood function (4) under the constraints (5), (6) with a fixed θ are given by*

$$v_i(\lambda) = \frac{d_{1i}}{R_{1i} + N\lambda^T \cdot g(x_i)}; \tag{8}$$

$$u_j(\lambda) = \frac{d_{2j}}{R_{2j} - N\lambda^T \cdot g(y_j - \theta)}; \tag{9}$$

where $N = n + m$ and $\lambda^T \cdot g(x_j)$ denote the inner product $\sum_i \lambda_i g_i(x_j)$. The λ value in equation (8), (9) above is obtained as the solution of the following k equations

$$\sum_i g_1(x_i)\log(1 - v_i(\lambda)) = \sum_i g_1(y_j + \theta)\log(1 - u_j(\lambda)) \tag{10}$$

$$\cdots \quad \cdots$$

$$\sum_i g_k(x_i)\log(1 - v_i(\lambda)) = \sum_i g_k(y_j + \theta)\log(1 - u_j(\lambda))\ . \tag{11}$$

Proof: The result follow from a standard Lagrange multiplier argument applied to (4), (5) and (6). $\diamondsuit$.

Since the solution of equations (10), (11) clearly depend on θ, we shall denote the solution λ as $\lambda(\theta)$ for the rest of this paper.

Lemma 2 *When X's and Y's have same distribution/hazard function, the solution $\lambda(\theta)$ of the constraint problem (10), (11) have the following asymptotic representations:*

(i)

$$\sqrt{N}\lambda(0) \xrightarrow{\mathcal{D}} N(0, \Sigma)\ ;$$

where Σ is defined by (21).

(ii) Assume $g(\cdot)$ is smooth and $h'(0)$ (defined in (20)) is invertible. For $|\theta| = O(1/\sqrt{N})$ we have

$$\lambda(\theta) = \lambda(0) + \theta a + o_p(1/\sqrt{N})$$

where

$$a = [h'(0)]^{-1}\left\{\int g_1'(t)\log(1 - d\Lambda(t)), \cdots, \int g_k'(t)\log(1 - d\Lambda(t))\right\}^T$$

and $\Lambda(t)$ is the common cumulative hazard function of X and Y.

Proof: See appendix. $\diamondsuit$.

The proof of the following two theorems actually work for right censored observations without change. Of course we will need the usual conditions on the censoring that ensure the Nelson-Aalen estimators to have asymptotic normal distributions. See Andersen et. al. (1993) for details.

Theorem 1 *Suppose that $F \equiv G$, the test statistics T_2 has asymptotically a chi-square distribution with one degree of freedom.*

PROOF: Let

$$f(\lambda(\theta)) = \sum \{d_{1i} \log v_i(\lambda(\theta)) + (R_{1i} - d_{1i}) \log[1 - v_i(\lambda(\theta))]\}$$
$$+ \sum \{d_{2j} \log u_j(\lambda(\theta)) + (R_{2j} - d_{2j}) \log[1 - u_j(\lambda(\theta))]\} \quad (12)$$

then we have

$$T_2 = -2\{f(\lambda(0)) - f(\lambda(\hat{\theta}))\} \ .$$

Or, recall the definition of $\hat{\theta}$ we have also

$$T_2 = -2 \min_\theta \{f(\lambda(0)) - f(\lambda(\theta))\} \ .$$

By Taylor expansion we have

$$T_2 = -2 \min_\theta \{f(0) + \lambda^T(0) f'(0) + 1/2 \lambda^T(0) D \lambda(0) + o_p(1)$$
$$- f(0) - \lambda(\theta)^T f'(0) - 1/2 \lambda^T(\theta) D \lambda(\theta) + o_p(1)\} \quad (13)$$

where we used D to denote the matrix of second derivatives of $f(\cdot)$ with respect to λ. The expansion are valid in view of Lemma 2.

Notice we have $f'(0) = 0$ (appendix), the above is reduced to

$$T_2 = -\min_\theta \{\lambda^T(0) D \lambda(0) - \lambda^T(\theta) D \lambda(\theta) + o_p(1)\} \ . \quad (14)$$

Use the representation $\lambda(\theta) = \lambda(0) + \theta a + o_p(1/\sqrt{N})$ from Lemma 2 and ignore the $o_p(1)$ term, we readily find the minimization over θ and obtain the minimized value (Lemma 5)

$$T_2 = \frac{[\lambda^T(0) D a]^2}{-a^T D a} + o_p(1) \ . \quad (15)$$

Recall the distributional result for $\lambda(0)$ in Lemma 2 and notice that (appendix)

$$-\frac{D}{N} \to D^* \ ,$$

it is not hard to show that

$$\sqrt{N}\lambda^T(0)(D/N)a \xrightarrow{\mathcal{D}} N(0, a^T D^* \Sigma D^* a) \ .$$

Finally we check $D^* = (\Sigma)^{-1}$ (see appendix) to get

$$\sqrt{N}\lambda^T(0)(D/N)a \xrightarrow{\mathcal{D}} N(0, a^T D^* a) \ . \tag{16}$$

This together with (15) imply that

$$T_2 \xrightarrow{\mathcal{D}} \chi^2(1) \ .$$

$\diamondsuit$.

Theorem 2 *The asymptotic distribution of the empirical envelope ML estimator $\hat{\theta}$ is given by*

$$\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N(0, \sigma^2)$$

where

$$\sigma^2 = \frac{1}{(a^T D^* a)}.$$

Proof: Recall we assumed WLOG that $\theta_0 = 0$. Let us focus on the expression (14). Aside from the $o_p(1/\sqrt{N})$ term, the θ that achieves the minimum is easily computed (Lemma ?):

$$\hat{\theta} = -\frac{\lambda^T(0)Da}{a^T Da} + o_p(1/\sqrt{N}) \ .$$

Therefore

$$\sqrt{N}\hat{\theta} = -\frac{\sqrt{N}\lambda^T(0)Da}{a^T Da} + o_p(1) \ .$$

In view of (16) we see that the theorem is proved. $\diamondsuit$.

Remark 1: The asymptotic variance of $\hat{\theta}$ is of interest. In fact we can show $a^T D^* a$ is equal to a summation that approximates the integral which in turn is the semiparametric Fisher information of the parameter θ in the original model. The approximation gets better when more constraints are used as in (2).

Remark 2: The asymptotic distribution of $\hat{F}(t_i)$ or more generally the asymptotic distribution of $\sum g_1(x_i) \log[1 - v_i(\lambda(\hat{\theta}))]$ can also be obtained easily using Theorem 2. We shall study the estimation of the nuisance parameter F and efficiency elsewhere.

Remark 3: The limiting distribution in Theorem 1 do not depend on k – the number of constraints we imposed on the distribution/hazard function. On the other hand, the limiting distribution in Theorem 2 and the confidence interval of θ based on the likelihood ratio statistic for a given sample do change with k.

5. An Example

We present one small example to illustrate the empirical envelope MLE procedure. We shall use the famous iris data set (of Fisher), focus on the petal length measurements for two species of iris: versicolor and virginica. There are 50 measurements for each species. The data is available as a built-in data set from the software R (Gentleman and Ihaka 1996).

The sample mean of petal length for versicolor is 4.26. The sample mean of petal length for virginica is 5.552, with a difference of 1.292.

We shall use three constraints in estimating θ, that for $t = 3.8, t = 4.2$ and $t = 4.5$

$$F_{vc}(t) = G_{vg}(t - \theta) \ .$$

The empirical envelope MLE of θ is found to be any value between 1.2 and 1.3. The non-uniqueness of the solution can be avoided if we use constraints (9), (10) with a continuous g.

Appendix

We first check $f'(0) = 0$. To this end we compute

$$\frac{\partial}{\partial \lambda_r} f(\lambda) = A + B$$

where

$$A = \sum_i d_{1i} \frac{(v_i)'_r}{v_i(\lambda)} - (R_{1i} - d_{1i}) \frac{(v_i)'_r}{1 - v_i(\lambda)} \ .$$

Letting $\lambda = 0$ and after some simplification we have

$$A = -\sum_i (R_{1i} - R_{1i}) \frac{d_{1i} N g_r(x_i)}{R_{1i}^2} \equiv 0 \ .$$

The calculation for term B is similar. $\diamondsuit$.

We now compute $f''(0) = D$. The rl^{th} element of the $k \times k$ matrix D is

$$D_{rl} = \frac{\partial^2}{\partial \lambda_r \partial \lambda_l} f(\lambda)|_{\lambda=0} \ .$$

After straight forward but tedious calculations we obtain

$$D_{rl} = -\left\{ \sum_i \frac{N^2 g_r g_l}{R_{1i}} \frac{d_{1i}}{R_{1i} - d_{1i}} + \sum_j \frac{N^2 g_r g_l}{R_{2j}} \frac{d_{2j}}{R_{2j} - d_{2j}} \right\} .$$

Now by a standard counting process argument, we see

$$-\frac{D_{rl}}{N} \to D^*_{rl} .$$

$\diamondsuit$

PROOF OF LEMMA 2. We show the asymptotic distribution of $\lambda(0)$. The argument is similar to, for example, Owen (1990) and Pan and Zhou (1999). Define a vector function $h(s) = (h_1(s), \cdots, h_k(s))$ by

$$h_1(s) = \sum_i g_1(x_i) \log(1 - v_i(s)) - \sum_j g_1(y_j) \log(1 - u_j(s)) , \tag{17}$$

$$\cdots \quad \cdots$$

$$h_k(s) = \sum_i g_k(x_i) \log(1 - v_i(s))) - \sum_j g_k(y_j) \log(1 - u_j(s)) , \tag{18}$$

then $\lambda(0)$ is the solution of $h(s) = 0$. Thus we have

$$0 = h(\lambda(0)) = h(0) + h'(0)\lambda(0) + o_p(1/\sqrt{N}) , \tag{19}$$

where $h'(0)$ is a $k \times k$ matrix. Therefore

$$\sqrt{N}\lambda(0) = [h'(0)]^{-1}(-\sqrt{N}h(0)) + o_p(1) .$$

The elements of $h'(0)$ are easily computed as

$$h'_{rl} = \sum_i \frac{N g_r(x_i) g_l(x_i) d_{1i}}{R_{1i}(R_{1i} - d_{1i})} + \sum_j \frac{N g_r(y_j) g_l(y_j) d_{2j}}{R_{2j}(R_{2j} - d_{2j})} . \tag{20}$$

Notice we have $Nh'_{rl} = -D_{rl}$. By the standard counting process martingale central limit theorem (see, for example, Gill (1981), Andersen et. al. (1993)) we can show that

$$\sqrt{N}h(0) \xrightarrow{\mathcal{D}} N(0, \Sigma_h)$$

with $\Sigma_h = \lim h'(0)$.

Finally, putting it together we have

$$\sqrt{N}\lambda(0) = [h'(0)]^{-1}(-\sqrt{N}h(0)) + o_p(1) \xrightarrow{\mathcal{D}} N(0, \Sigma)$$

with

$$\Sigma = \lim[h'(0)]^{-1} \ . \tag{21}$$

Recall $Nh'_{rl} = -D_{rl}$, we see that $\Sigma^{-1} = \lim[h'(0)] = \Sigma_h = D^*$. This completes the proof of (i).

The result (ii) can be obtained by noticing

$$\sum_j g_k(y_j) \log(1-u_j) - \sum_j g_k(y_j+\theta) \log(1-u_j) = -\theta \sum_j g'_k(y_j) \log(1-u_j) + o(|\theta|) \ ,$$

and thus

$$-\theta \sum_j g'(y_j) \log(1-u_j) \approx h(\lambda(\theta)) - h(\lambda(0)) \approx h'(0)[\lambda(\theta) - \lambda(0)].$$

$\diamondsuit$

References

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

Fang, H. and Zhou, M. (2000). On empirical likelihood ratio method with k-sample problems. *Univ. of Kentucky, Dept. of Statist. Tech Report* .

Geman, S. and Huang, C.R. (1982). Nonparametric maximum likelihood estimation by the method of sieves. Ann. Statist. **10**, 401-414.

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

Grenander, U. (1981). *Abstract Inference.* Wiley, New York.

Chen, K., Zhou, M. and Huang, B. (2000). Using quadratic programing to compute empirical likelihood ratio. *Univ. of Kentucky, Dept. of Statist. Tech Report* .

Li, G. (1995). On nonparametric likelihood ratio estimation of survival probabilities for censored data. *Statist. & Prob. Letters*, **25** , 95-104.

Owen, A. (1988). Empirical Likelihood Ratio Confidence Intervals for a Single Functional. *Biometrika*, **75**, 237-249.

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

Owen, A. (2001). *Empirical Likelihood* Chapman & Hall/CRC, London.

Pan, X.R. and Zhou, M. (1999). Empirical likelihood in terms of cumulative hazard function for censored data. *Univ. of Kentucky, Dept. of Statist. Tech Report # 361.* To appear in *J. Multivariate Analysis.*

Gentleman, R. and Ihaka, R. (1996). R: A Language for data analysis and graphics. *J. of Computational and Graphical Statistics*, **5**, 299-314.

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

van der Vaart, (1988). Statistical Estimation in Large Parameter Spaces. CWI tract 44.

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

Lemma 5 Suppose D is a positive definite matrix of $k \times k$, and a is a vector of $k \times 1$, b is a matrix of $k \times p$ and u is a vector of $p \times 1$.

The minimization of the following quadratic form

$$\min_u \ (a - bu)^T D(a - bu)$$

occurs when $u = u^*$ where u^* is the solution of the following equations

$$b^T D(a - bu^*) = 0 \ .$$

The minimum value achieved is

$$a^T Da - a^T Dbu^* = a^T Da - (bu^*)^T Dbu^* \ .$$

PROOF: Define an inner product of any two $k \times 1$ vectors as

$$(v_1 \cdot v_2) = v_1^T Dv_2 \ ,$$

and then define the length of the vector as $||v|| = \sqrt{(v \cdot v)}$ in terms of this inner product. Then the minimization problem above can be viewed as the minimization of the length of the vector $(a - bu)$.

The u^* that achieve the minimization makes the vector $(a - bu^*)$ orthogonal to the linear space spanned by b. Therefore we have

$$b \cdot (a - bu^*) = b^T D(a - bu^*) = 0 \ .$$

The minimum value can then be easily calculated. $\diamondsuit$