Hazard Estimating Equations and Empirical Likelihood

Yanling Hu and Mai Zhou

Department of Statistics, University of Kentucky, Lexington, KY 40536

Abstract

Qin and Lawless (1994) studied the empirical likelihood method with general estimating equations. They obtained very nice asymptotic properties especially when the number of estimating equations exceeds the number of parameters (over determined case). We study here a parallel setup to Qin and Lawless (1994) except uses a hazard-type estimating equations. The empirical likelihood we used here is also formulated in terms of the hazard. The advantage of using hazard is that right censored data can be handled easily through martingale techniques. We obtained similar asymptotic results for the maximum empirical likelihood estimators and the empirical likelihood ratio tests, including the over determined case. Three examples are provided to demonstrate the potential applications of the methodology.

Key Words: Hazard Empirical Likelihood; Estimating Equations; Censored data; Martingale theory; Asymptotic chi-square distribution.

1 Introduction

The first use of the empirical likelihood method was proposed by Thomas and Grunkemeier (1975) for right censored data and the Kaplan-Meier estimator. It is a nonparametric statistical inference method similar to the parametric likelihood ratio test. Empirical Likelihood has been widely studied since a series of papers by Owen starting at (1988) and ultimately summarized in his book of (2001).

Empirical likelihood method now found numerous applications the construction of confidence regions and hypothesis tests in nonparametric settings or distribution-free contexts.

Other asymptotic properties of empirical likelihood ratio statistics have been investigated by DiCiccio and Romano (1989), DiCiccio, Hall and Romano (1989) and others.

Qin and Lawless (1994) showed that the empirical likelihood method could be brought to bear on problems with over-determined estimating equations, where the number of estimating equations r exceeds the number of parameters p. They demonstrated how the maximum empirical likelihood estimators of parameters $\boldsymbol{\theta} \in \Re^p$ may be obtained and determined its asymptotic multivariate normal distribution. They also proved that the empirical likelihood ratio test statistic for parameters have asymptotic $\chi^2_{(p)}$ distributions so that confidence regions and hypothesis tests could be constructed. When $r = p$, their results are the same as those of Owen (1988, 1990).

However, Qin and Lawless's results are limited to uncensored data. For right censored data, no such results are available. Hence, we propose a parallel construct to that of Qin and Lawless, except uses a hazard-type empirical likelihood with over-determined *hazard-type* estimating equations/constraints. This approach naturally incorporates censoring and counting process martingales. The resulting empirical likelihood estimator and test statistic also have very nice asymptotic properties.

Some existing work on censored data empirical likelihood include empirical likelihood for a single constraint on the surviving probability by Li (1995); empirical likelihood ratio test for the equality of several medians, Naik-Nimbalkar and Rajarshi (1997); Empirical likelihood for the weighted hazards by Pan and Zhou (2002), etc. Li, Li and Zhou (2005) provides a review of empirical likelihood results in survival analysis.

This manuscript is organized as follows. Section 2 describes the background of the research and some basic theory. The asymptotic properties and results are given in Section 3. In Section 4, we show the efficiency of the maximum empirical likelihood estimator. The application of our results with three examples are provided in Section 5. Section 6 gives the concluding remarks. Detailed proofs are not shown in this manuscript due to space limit. But we refer reader to Hu (2011).

2 Empirical Likelihood, Over-determined Constraints in terms of Hazard

Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. nonnegative random variables denoting the lifetimes with a continuous distribution function F_0 and cumulative hazard Λ_0. Independent of the lifetimes there are censoring times $C_1, C_2, \ldots, C_n$ which are i.i.d. with a distribution G_0. F_0 and G_0 are unknown in practice. Only the censored observations, (t_i, δ_i), are available to us:

$$t_i = \min(x_i, c_i) \quad \text{and} \quad \delta_i = I[x_i \leq c_i] \quad \text{for } i = 1, 2, \ldots, n.$$

The empirical likelihood based on the censored observations (t_i, δ_i) in terms of distribution function F is:

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

We shall re-cast the empirical likelihood in terms of the hazard function. In general, the cumulative hazard function $\Lambda(t)$ related to a general CDF $F(t)$ is defined by

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

We will restrict our analysis (that is, search for the maximum) of the empirical likelihood to the purely discrete functions dominated by their NPMLE's. See Owen(1988) for the reason and discussion on this restriction.

Using the relation between the $\Lambda(t)$ and $F(t)$, the empirical likelihood above can be written in terms of hazard function:

$$EL(\Lambda) = \prod_{i=1}^{n} [\Delta\Lambda(t_i)]^{\delta_i} [\prod_{j:t_j<t_i} (1 - \Delta\Lambda(t_j))]^{\delta_i} [\prod_{j:t_j\le t_i} (1 - \Delta\Lambda(t_j))]^{1-\delta_i}$$

where $\Delta\Lambda(t) = \Lambda(t+) - \Lambda(t-)$ is the jump of Λ at t. The reason we use $EL(\Lambda)$ instead of $EL(F)$ will be discussed in Section 6.

In this manuscript, we will use a simpler version of the $EL(\Lambda)$, which is called a Poisson extension of the likelihood by Murphy (1995) and was also used by Pan and Zhou (2002):

$$\begin{aligned} AL(\Lambda) &= \prod_{i=1}^{n} [\Delta\Lambda(t_i))]^{\delta_i} \exp\{-\Lambda(t_i)\} && (1) \\ &= \prod_{i=1}^{n} [\Delta\Lambda(t_i))]^{\delta_i} \exp\{-\sum_{j:t_j\le t_i} \Delta\Lambda(t_j)\} \end{aligned}$$

Notice we have assumed a discrete $\Lambda(t)$ in the above.

The difference between $AL(\Lambda)$ and $EL(\Lambda)$ is small and negligible for large n. See Pan and Zhou (2002) for the comparison.

Let $w_i = \Delta\Lambda(t_i)$ for $i = 1, 2, \ldots, n$, where we notice $w_n = \delta_n$ because the last jump of a discrete cumulative hazard function must be one. The likelihood at this Λ can be written in term of the jumps

$$AL = \prod_{i=1}^{n} [w_i]^{\delta_i} \exp\{-\sum_{j=1}^{n} w_j I[t_j \le t_i]\} \ ,$$

and the log likelihood is

$$\begin{aligned} \log AL &= \sum_{i=1}^{n} \left\{ \delta_i \log w_i - \sum_{j=1}^{n} w_j I[t_j \le t_i] \right\} && (2) \\ &= \sum_{i=1}^{n} \delta_i \log w_i - \sum_{i=1}^{n} w_i R_i \ , && (3) \end{aligned}$$

where $R_i = \sum_j I[t_j \ge t_i]$.

If we maximize the log AL above over all possible (discrete) hazard functions, it is well known that this yields $w_i = \frac{\delta_i}{R_i}$. This is the well known Nelson-Aalen estimator: $\Delta\hat{\Lambda}_{NA}(t_i) = \frac{\delta_i}{R_i}$.

Next, we want to maximize the log AL subject to some estimating equations. For this purpose we need first discuss the estimating equations in terms of hazard. Denote the parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$. We assume the true value of the parameter $\boldsymbol{\theta_0}$ satisfy some finite functionals of the hazard function: $\boldsymbol{\theta} = T(\Lambda)$. In particular we assume $\boldsymbol{\theta}$ is defined by the equations

$$\left\{ \int g_1(t, \boldsymbol{\theta}) d\Lambda(t), \cdots, \int g_r(t, \boldsymbol{\theta}) d\Lambda(t) \right\}^\top = (k_1, \cdots, k_r)^\top \qquad (4)$$

where $g_i(t,\cdot)$ are given functions satisfy some moment conditions. Notice here the number of equations, r, can be larger than p. Denote $(k_1,\cdots,k_r)^\top = \boldsymbol{k}$.

Example 1: The parameter θ of median can be defined as $\int I_{[t\le\theta]} d\Lambda(t) = \log 2$.

Example 2: The paper of Naik-Nimbalkar and Rajarshi (1997) discuss the testing of the equality of the median of k samples. If we *assume* there is a common median, then how to estimate the common, unknown median become a question of over-determined estimating equation problem, since the median can be estimated from any of the k samples. In fact Naik-Nimbalkar and Rajarshi actually gave the definition of the maximum empirical likelihood estimator $\hat\theta$ of the common median (see their p. 269) and have noticed it actually is an optimal linear combination of the k individual median estimators.

For more examples please see the section 5 later.

The discrete version of the above estimating equations (4) is then:

$$\sum_{i=1}^{n-1} \delta_i \boldsymbol{g}(t_i,\boldsymbol{\theta}) w_i + \boldsymbol{g}(t_n,\boldsymbol{\theta})\delta_n = \boldsymbol{k}\ . \tag{5}$$

These equations can be used in two different ways. First, if $r=p$ then we may let $w_i = \delta_i/R_i = \Delta\hat\Lambda_{NA}(t_i)$ and solve the equations in term of $\boldsymbol{\theta}$. The solution is the estimator (in fact NPMLE).

Second, when $r>p$ these equations do not in general have a solution in terms of $\boldsymbol{\theta}$. But we can fix the $\boldsymbol{\theta}$ (at the null value for example) and solve these equations in term of w_i. These w_i will give rise to an empirical likelihood value and provide us a way to test the hypothesis of the null value of the $\boldsymbol{\theta}$.

The following Theorem 1 is just the solution of (5), in terms of w_i for fixed $\boldsymbol{\theta}$.

The first step in our empirical likelihood analysis is to find a (discrete) cumulative hazard function w_i that maximizes the $\log AL$ under the constraints (5), notice here w_i in general do not equal to the Nelson-Aalen jump.

Theorem 1 *If the constraints (5) are feasible (which means there is at least a genuine hazard w_i that solve (5)), then the maximum of AL under the constraints is obtained when*

$$\begin{aligned} w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}),\boldsymbol{\theta}) &= \frac{\delta_i}{R_i + n\boldsymbol{\lambda}(\boldsymbol{\theta})^\top \boldsymbol{g}(t_i,\boldsymbol{\theta})\delta_i} \\ &= \frac{\delta_i}{R_i} \times \frac{1}{1+\boldsymbol{\lambda}(\boldsymbol{\theta})^\top(\delta_i \boldsymbol{g}(t_i,\boldsymbol{\theta})/(R_i/n))} \\ &= \Delta\hat\Lambda_{NA}(t_i)\frac{1}{1+\boldsymbol{\lambda}(\boldsymbol{\theta})^\top \boldsymbol{Z}(t_i,\boldsymbol{\theta})}\ , \end{aligned} \tag{6}$$

where

$$\boldsymbol{Z}(t_i,\boldsymbol{\theta}) = \frac{\delta_i \boldsymbol{g}(t_i,\boldsymbol{\theta})}{R_i/n} = (Z_1(t_i,\boldsymbol{\theta}),\ldots,Z_r(t_i,\boldsymbol{\theta}))^\top \quad \textit{for} \;\; i=1,2,\ldots,n-1. \tag{7}$$

and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_r)^\top$ *are the solutions of the following* r *equations*

$$\begin{aligned}
\boldsymbol{h}(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) &= \sum_{i=1}^{n-1} \delta_i \boldsymbol{g}(t_i, \boldsymbol{\theta}) w_i(\boldsymbol{\lambda}(\boldsymbol{\theta}), \boldsymbol{\theta}) + \delta_n \boldsymbol{g}(t_n, \boldsymbol{\theta}) - \boldsymbol{k} \\
&= \sum_{i=1}^{n-1} \frac{\delta_i \boldsymbol{g}(t_i, \boldsymbol{\theta})}{R_i} \times \frac{1}{1 + \boldsymbol{\lambda}(\boldsymbol{\theta})^\top \boldsymbol{Z}(t_i, \boldsymbol{\theta})} + \delta_n \boldsymbol{g}(t_n, \boldsymbol{\theta}) - \boldsymbol{k} \qquad (8) \\
&= \frac{1}{n} \sum_{i=1}^{n-1} \frac{\boldsymbol{Z}(t_i, \boldsymbol{\theta})}{1 + \boldsymbol{\lambda}(\boldsymbol{\theta})^\top \boldsymbol{Z}(t_i, \boldsymbol{\theta})} + \delta_n \boldsymbol{g}(t_n, \boldsymbol{\theta}) - \boldsymbol{k} \\
&= \mathbf{0}. \qquad (9)
\end{aligned}$$

Proof. Use standard Lagrange multiplier calculation. Similar to Pan and Zhou (2002).

3 Theory and Asymptotic Results

Consider the hypothesis:

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta_0}; \quad vs. \quad H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta_0}.$$

The true value of the parameter θ_0 are assumed to satisfy

$$\left\{ \int g_1(t, \boldsymbol{\theta_0}) d\Lambda_0(t), \cdots, \int g_r(t, \boldsymbol{\theta_0}) d\Lambda_0(t) \right\}^\top = (k_1, \cdots, k_r)^\top .$$

We propose an empirical likelihood ratio statistic as follows:

$$T = -2 \left\{ \max_{\boldsymbol{\theta} = \boldsymbol{\theta_0}, w_i} \log AL - \max_{\boldsymbol{\theta} \in R^p, w_i} \log AL \right\} . \qquad (10)$$

The first maximum in the test statistic T above can be obtained through Theorem 1, with $\boldsymbol{\theta} = \boldsymbol{\theta_0}$ and the w_i given there. The second maximization of above statistic T is taken over all possible $\boldsymbol{\theta}$ and w_i.

In the case $r = p$, assume the below equations have a unique solution $\hat{\boldsymbol{\theta}}$ (r equations and r unknowns)

$$\left\{ \int g_1(t, \boldsymbol{\theta}) d\hat{\Lambda}_{NA}(t), \cdots, \int g_r(t, \boldsymbol{\theta}) d\hat{\Lambda}_{NA}(t) \right\}^\top = (k_1, \cdots, k_r)^\top ,$$

then the maximum is achieved when we just use $w_i = \Delta \hat{\Lambda}_{NA}(t_i)$ for the second $\log AL$.

In the case $r > p$, we have to search over θ for the max. But once the $\boldsymbol{\theta}$ is fixed, the w_i can once again be obtained by Theorem 1. So, it is a search over the $\boldsymbol{\theta}$. This empirical likelihood ratio test statistic has an approximate chi-square distribution with p degrees of freedom under the null hypothesis. We reject H_0 for larger values of T. Confidence regions for $\boldsymbol{\theta}$ can be obtained by inverting the chi-square test.

Definition: The $\boldsymbol{\theta}$ value that achieve the maximum in the second term of the test statistic T in (10) will be our maximum empirical likelihood estimator, $\widehat{\boldsymbol{\theta}}$.

In this section, we will first use Weak Law of Large Number (Lemma 1) and Martingale Central Limit Theorem (Lemma 2) to get some asymptotic properties for $\boldsymbol{Z}$ defined in (7). Lemma 3 gives the asymptotic properties of $\boldsymbol{\lambda}$. Finally the main results are formulated in Theorem 2 and 3. When we take limit of a matrix, we are taking limit of each element of the matrix.

Lemma 1 *Let $(T_1, \delta_1), \ldots, (T_n, \delta_n)$ be n pairs of random variables as defined above. Suppose $g_i(x, \boldsymbol{\theta})$ $i = 1, \ldots, r$ are left continuous functions and*

$$0 < \int \frac{|g_i(x, \boldsymbol{\theta})||g_j(x, \boldsymbol{\theta})|}{(1 - F_0(x))(1 - G_0(x-))} d\Lambda_0(x) < \infty, \quad \forall i, j \quad 1 \leq i, j \leq r.$$

Then,we have, for $\boldsymbol{Z}$ defined in (7),

$$\frac{1}{n} \sum_{i=1}^{n} Z_u(t_i, \boldsymbol{\theta}) Z_v(t_i, \boldsymbol{\theta}) = \int \frac{g_u(t, \boldsymbol{\theta}) g_v(t, \boldsymbol{\theta})}{R(t)/n} d\hat{\Lambda}_{NA}(t) \xrightarrow{P} \int \frac{g_u(x, \boldsymbol{\theta}) g_v(x, \boldsymbol{\theta})}{(1 - F_0(x))(1 - G_0(x-))} d\Lambda_0(x)$$

where

$$R(t) = \sum I_{[T_i \geq t]}.$$

Lemma 2 *In addition to the assumptions of Lemma 1, we assume the matrix Σ_Z is positive definite. For $\boldsymbol{Z}$ defined in (7), we have, under null hypoyhesis*

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{Z}(t_i, \boldsymbol{\theta}_0) - \boldsymbol{k} \right) = \sqrt{n} \left(\sum_{i=1}^{n} \boldsymbol{g}(t_i, \boldsymbol{\theta}_0) \Delta \hat{\Lambda}_{NA}(t_i) - \boldsymbol{k} \right) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \Sigma_Z)$$

as $n \to \infty$ where

$$\boldsymbol{k} = \left\{ \int g_1(t, \boldsymbol{\theta_0}) d\Lambda_0(t), \ldots, \int g_r(t, \boldsymbol{\theta_0}) d\Lambda_0(t) \right\}^{\top},$$

$$\Sigma_{Zuv} = \int \frac{g_u(x, \boldsymbol{\theta}_0) g_v(x, \boldsymbol{\theta}_0)}{(1 - F_0(x))(1 - G_0(x-))} d\Lambda_0(x) \quad \forall u, v \quad 1 \leq u, v \leq r.$$

Lemma 3 *Under mild regularity conditions (smoothness conditions that allow Taylor expansion), the solution $\boldsymbol{\lambda}$ of the constraint equations in (9) under the null hypothesis has the following asymptotic representations:*

(i) Let $\boldsymbol{\theta}_0$ to be the true value of the parameters, and assume

$$\boldsymbol{h}'(\mathbf{0}, \boldsymbol{\theta}_0) = \frac{\partial \boldsymbol{h}(\boldsymbol{\lambda}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\lambda}} |_{\boldsymbol{\lambda} = \mathbf{0}}$$

is an invertible $r \times r$ matrix, then we have

$$\sqrt{n} \boldsymbol{\lambda}(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \Sigma_\lambda); \quad as \quad n \to \infty$$

where $\Sigma_\lambda = \Sigma_Z^{-1} = \lim_{n\to\infty}[\boldsymbol{h}'(\mathbf{0}, \boldsymbol{\theta}_0)]^{-1}$.

(ii) In addition, assume that $\boldsymbol{g}(\cdot)$ *are smooth and* $|\boldsymbol{\theta} - \boldsymbol{\theta}_0| = O(1/\sqrt{n})$, *we have*

$$\boldsymbol{\lambda}(\boldsymbol{\theta}) = \boldsymbol{\lambda}(\boldsymbol{\theta}_0) - \{\boldsymbol{h}'(\mathbf{0}, \boldsymbol{\theta}_0)\}^{-1} A(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_p(1/\sqrt{n})$$

where A *is an* $r \times p$ *matrix defined as*

$$A = \frac{\partial \boldsymbol{h}(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \,|_{\boldsymbol{\lambda}=0, \boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \boldsymbol{Z}(t_i, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$$

Proof. Use Taylor expansion on $\boldsymbol{h}$ with respect to $\boldsymbol{\lambda}$.
Remark. The $r \times r$ matrix

$$\boldsymbol{h}'(\mathbf{0}, \boldsymbol{\theta}_0) = \frac{\partial \boldsymbol{h}(\boldsymbol{\lambda}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\lambda}} \,|_{\boldsymbol{\lambda}=\mathbf{0}} = -\frac{1}{n} \sum_{i=1}^{n-1} \boldsymbol{Z}(t_i, \boldsymbol{\theta}_0) \boldsymbol{Z}^\top(t_i, \boldsymbol{\theta}_0)$$

is easy to verify to be symmetric and at least non-positive definite. Proper choice of the g function will guarantee it to be negative definite.

Theorem 2 *Under the null hypothesis, plus same regularity assumptions as in Lemma 3, the test statistic* T *has asymptotically a chi-square distribution with* p *degrees of freedom:*

$$T \xrightarrow{\mathcal{D}} \chi^2_{(p)} \quad , \quad as \quad n \to \infty.$$

Theorem 3 *Let* $\hat{\boldsymbol{\theta}}$ *be the over determined case maximum empirical likelihood estimator as defined earlier. Assume that the true parameter* $\boldsymbol{\theta_0}$ *and the true hazard function* $\Lambda_0(t)$ *satisfy the equations (4), then the asymptotic distribution of* $\hat{\boldsymbol{\theta}}$ *is given by*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \Sigma_\theta) \ ,$$

where

$$\Sigma_\theta = -\lim_{n\to\infty} \{A^\top [\boldsymbol{h}'(\mathbf{0}, \boldsymbol{\theta}_0)]^{-1} A\}^{-1}.$$

4 Efficiency

We now take a closer look at the asymptotic variance of the maximum empirical likelihood estimator obtained in Theorem 3 above. By noticing that the structure of Σ_θ is the same as the variance-covariance matrix in Qin and Lawless's paper, we can get the same corollary as follows.

Corollary 1 *When* $r > p$, *the asymptotic variance* Σ_θ *of* $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ *cannot decrease if a constraint equation is dropped.*

This shows that the maximum empirical likelihood estimator obtained in Theorem 3 is asymptotically efficient. The variance-covariance matrix will not decrease when we include fewer constraint equations. In other words, it is recommended that we should always use as much information as possible.

5 Examples

5.1 Example 1.

For $i = 1, \ldots, n$, let T_i be the failure time for the ith subject and let $\boldsymbol{X}_i$ be the associated p-vector of covariates. The accelerated failure time model specifies that

$$\log T_i = \boldsymbol{\beta}_0^\top \boldsymbol{X}_i + \epsilon_i, \quad i = 1, \ldots, n.$$

where $\boldsymbol{\beta}_0$ is a p-vector of unknown regression parameters and ϵ_i $(i = 1, \ldots, n)$ are independent error terms with a common, but completely unspecified, distribution.

Let C_i be the censoring time for T_i. $\tilde{T}_i = \min(T_i, C_i)$, $\delta_i = I_{\{T_i \le C_i\}}$. Define $e_i(\boldsymbol{\beta}) = \log \tilde{T}_i - \boldsymbol{\beta}^\top \boldsymbol{X}_i$. The weighted log-rank estimating function for parameter $\boldsymbol{\beta}_0$ takes the form

$$U_\phi(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \phi\{\boldsymbol{\beta}; e_i(\boldsymbol{\beta})\}[\boldsymbol{X}_i - \bar{\boldsymbol{X}}\{\boldsymbol{\beta}; e_i(\boldsymbol{\beta})\}],$$

where ϕ is a weight function, possibly data-dependent, and

$$\bar{\boldsymbol{X}}\{\boldsymbol{\beta}; e_i(\boldsymbol{\beta})\} = \frac{\sum_j X_j I[e_j \ge e_i]}{\sum_j I[e_j \ge e_i]} \ .$$

Two popular choices for the weight function ϕ are 'Gehan' weight and the 'Logrank' weight ($\phi = const.$).

Let $\hat{\boldsymbol{\beta}}_\phi$ be a root of the estimating function $U_\phi(\boldsymbol{\beta})$. We get different roots when we use different ϕ function.

Jin *et al* (2003) proposed new estimators for the regression parameters $\boldsymbol{\beta}$ by taking different forms of ϕ in the semiparametric accelerated failure time model with censored observations. They compared Gehan estimator and Log-rank estimator in the simulation studies and demonstrate their respective advantages in different scenarios. However, using our over determined estimating equation approach, we do not have to choose between them, we can use *both* of them and theoretically obtain a better estimator.

Using the estimating function above and re-writing it as estimating equations with $e_i(\beta) = s_i$ and with hazard summation:

$$0 = \sum_{i=1}^{n} \delta_i R(s_i) \phi\{\boldsymbol{\beta}; s_i\}[\boldsymbol{X}_i - \bar{\boldsymbol{X}}\{\boldsymbol{\beta}; s_i\}] w_i \ . \tag{11}$$

These equations needs to be coupled with the log empirical likelihood function (3).

One feature of the above equation is that if (1) $r = p$, and (2) $w_i = \Delta\hat{\Lambda}_{NA}(s_i)$ then the solution of the equations coincide with the usual rank estimator described in Jin et. al. (2003).

Some R code of computation:

This example is for the AFT model rank estimator. We illustrate the use of empirical likelihood (Poisson version) with the hazard estimating equations.

If every R functions are defined, you may test the hypothesis for β as in

```
library(emplik)
RankRegTestH(y=log(myeloma[,1]), d=myeloma[,2], x=myeloma[,3],
                                  beta= -2.46646 , type="Gehan")
```

Another option of the `type` is 'Logrank'. The relevant functions are defined as

```
RankRegTestH <- function(y, d, x, beta, type="Gehan") {
     n <- length(y)          ## dimension of x must be n x q.
     x <- as.matrix(x)       ## x must NOT including an intercept.
     xdim <- dim(x)
     if( xdim[1] != n ) stop("check dim of x")
     if( length(beta) != xdim[2] ) stop("check dim of beta and x")

e <- y - as.vector( x %*% beta )
ordere <- order(e, -d)
esort <- e[ordere]
dsort <- d[ordere]
xsort <- as.matrix(x[ordere,])
dsort[length(dsort)] <- 1       #last one as uncensored always?

##xbar <- rev(cumsum(rev(xsort)))/(n:1)

xbar <- xsort
####for(j in 1:(n-1)) xbar[j,] <- colMeans(xsort[j:n,])
for(j in 1:xdim[2]) xbar[,j] <- rev(cumsum(rev(xsort[,j])))/(n:1)

if(type == "Gehan") {A <- (n:1)^2 * (xsort - xbar)}
 else {if(type == "Logrank") A <- (n:1) * (xsort - xbar)
        else stop("type must be either Gehan or Logrank") }

## A1 <- (n:1)^2 * (xsort -xbar)
## A2 <- (n:1) * (xsort - xbar)
## A <- cbind(A1, A2)
AA <- as.matrix(A[dsort == 1,])

myfun <- function(t, A){ return(A) }
myfun2 <- function(t){ matrix(0, ncol=ncol(AA), nrow=length(t)) }
x20 <- runif(30)
d20 <- rep(1, 30)
```

```
temp2 <- emplikHs.test22(x1=esort, d1=dsort, x2=x20, d2=d20,
             theta=rep(0,ncol(AA)), fun1=myfun, fun2=myfun2, A=AA)

Samp1EL <- temp2$"-2LLR(sample1)"

list(loglikH0=temp2$"Llik(sample1)",  "-2LLR"=Samp1EL)
}
```

The following function is basically the same as the one in the package **emplik**, only the functions `fun1` can take extra inputs as in `...`. We need that for the way we define the rank estimator estimating equations.

```
emplikHs.test22 <- function (x1, d1, y1 = -Inf, x2, d2, y2 = -Inf,
     theta, fun1, fun2, maxit = 25, tola = 1e-07, itertrace = FALSE, ...)
{
    theta <- as.vector(theta)
    q <- length(theta)
    x1 <- as.vector(x1)
    n1 <- length(x1)
    if (n1 <= 2 * q + 1)
        stop("Need more observations in x1")
    if (length(d1) != n1)
        stop("length of x1 and d1 must agree")
    if (any((d1 != 0) & (d1 != 1)))
        stop("d1 must be 0/1's for censor/not-censor")
    if (!is.numeric(x1))
        stop("x1 must be numeric -- observed times")
    x2 <- as.vector(x2)
    n2 <- length(x2)
    if (n2 <= 2 * q + 1)
        stop("Need more observations for sample 2")
    if (length(d2) != n2)
        stop("length of x2 and d2 must agree")
    if (any((d2 != 0) & (d2 != 1)))
        stop("d2 must be 0/1's for censor/not-censor")
    if (!is.numeric(x2))
        stop("x2 must be numeric -- observed times")
    newdata1 <- Wdataclean2(z = x1, d = d1)
    temp1 <- DnR(newdata1$value, newdata1$dd, newdata1$weight,
        y = y1)
    newdata2 <- Wdataclean2(z = x2, d = d2)
    temp2 <- DnR(newdata2$value, newdata2$dd, newdata2$weight,
        y = y2)
    jump1 <- (temp1$n.event)/temp1$n.risk
    jump2 <- (temp2$n.event)/temp2$n.risk
```

```
index1 <- (jump1 < 1)
index2 <- (jump2 < 1)

funtime11 <- as.matrix(fun1(temp1$times, ...))

if (ncol(funtime11) != q)
    stop("check the output dim of fun1, and theta")
funtime21 <- as.matrix(fun2(temp2$times))
if (ncol(funtime21) != q)
    stop("check the output dim of fun2, and theta")

Kcent <- jump1 %*% funtime11 - jump2 %*% funtime21
if (itertrace)
    print(c("Kcenter=", Kcent))

K12 <- rep(0, q)
tm11 <- temp1$times[!index1]
if (length(tm11) > 1)
    stop("more than 1 place jump>=1 in x1?")
if (length(tm11) > 0) {
  ## This seems to calculate f(last point).
    K12 <- K12 ## + as.vector(fun1(tm11, ...))
  ## Since at the last point, the
  ##  score x - bar x is always 0, we just assign 0.
}
tm21 <- temp2$times[!index2]
if (length(tm21) > 1)
    stop("more than 1 place jump>=1 in x2?")
if (length(tm21) > 0) {
    K12 <- K12 - as.vector(fun2(tm21))
}
eve1 <- temp1$n.event[index1]
tm1 <- temp1$times[index1]
rsk1 <- temp1$n.risk[index1]
jmp1 <- jump1[index1]
funtime1 <- as.matrix(fun1(tm1, ...))

if(length(tm11) > 0) {funtime1 <- as.matrix(funtime1[-nrow(funtime1),])}

eve2 <- temp2$n.event[index2]
tm2 <- temp2$times[index2]
rsk2 <- temp2$n.risk[index2]
jmp2 <- jump2[index2]
funtime2 <- as.matrix(fun2(tm2))
```

```
TINY <- sqrt(.Machine$double.xmin)
if (tola < TINY)
    tola <- TINY
lam <- rep(0, q)

N <- n1 + n2
nwts <- c(3^-c(0:3), rep(0, 12))
gwts <- 2^(-c(0:(length(nwts) - 1)))
gwts <- (gwts^2 - nwts^2)^0.5
gwts[12:16] <- gwts[12:16] * 10^-c(1:5)
nits <- 0
gsize <- tola + 1
while (nits < maxit && gsize > tola) {
    grad <- gradf3(lam, funtime1, eve1, rsk1, funtime2, eve2,
        rsk2, K = theta - K12, n = N)
    gsize <- mean(abs(grad))
    arg1 <- as.vector(rsk1 + funtime1 %*% lam)

    arg2 <- as.vector(rsk2 - funtime2 %*% lam)
    ww1 <- as.vector(-llogpp(arg1, 1/N))^0.5
    ww2 <- as.vector(-llogpp(arg2, 1/N))^0.5
    tt1 <- sqrt(eve1) * ww1
    tt2 <- sqrt(eve2) * ww2
    HESS <- -(t(funtime1 * tt1) %*% (funtime1 * tt1) + t(funtime2 *
        tt2) %*% (funtime2 * tt2))
    nstep <- as.vector(-solve(HESS, grad))
    gstep <- grad
    if (sum(nstep^2) < sum(gstep^2))
        gstep <- gstep * (sum(nstep^2)^0.5/sum(gstep^2)^0.5)
    ninner <- 0
    for (i in 1:length(nwts)) {
        lamtemp <- lam + nwts[i] * nstep + gwts[i] * gstep
        ngrad <- gradf3(lamtemp, funtime1, eve1, rsk1, funtime2,
            eve2, rsk2, K = theta - K12, n = N)
        ngsize <- mean(abs(ngrad))
        if (ngsize < gsize) {
            lam <- lamtemp
            ninner <- i
            break
        }
    }
    nits <- nits + 1
    if (ninner == 0)
        nits <- maxit
```

```
        if (itertrace)
            print(c(lam, gsize, ninner))
    }
    lamfun1 <- as.vector(funtime1 %*% lam)
    lamfun2 <- as.vector(funtime2 %*% lam)
    onePlamh1 <- (rsk1 + lamfun1)/rsk1
    oneMlamh2 <- (rsk2 - lamfun2)/rsk2
    loglik1 <- (sum(eve1 * llog(onePlamh1, 1/N)) - sum(eve1 *
        (lamfun1)/(rsk1 + lamfun1)))
    loglik1fenzi <- -sum(eve1 * llog((rsk1 + lamfun1), 1/N)) -
        sum(eve1/onePlamh1)
    loglik2 <- (sum(eve2 * llog(oneMlamh2, 1/N)) - sum(eve2 *
        (-lamfun2)/(rsk2 - lamfun2)))
    loglik <- 2 * (loglik1 + loglik2)
    list(`-2LLR` = loglik, lambda = lam, `-2LLR(sample1)` = 2 *
        loglik1, `Llik(sample1)` = loglik1fenzi)
}
```

5.2 Example 2.

This is the same example as the one in Kim (2003) dissertation.

An AML study by Embury *et al.* at Stanford University reports the results of a clinical trial to evaluate the efficacy of maintenance chemotherapy for acute myelogenous leukemia (AML). After reaching a status of remission through treatment by chemotherapy, the patients who enter the study are assigned randomly to two groups. The first, or treatment, group receives maintenance chemotherapy; the second, or control, group does not. Interest is on analyzing data if maintenance chemotherapy does either delay or prolong the time until relapse. The efficacy of maintenance chemotherapy for AML is evaluated on two aspects, a shift of the time to relapse and a proportional hazards change in the distribution of time to death.

Based on the Kaplan-Meier survival curves drawn from the data, we are convinced that a hybrid model with the shift parameter and the proportional hazard rate is appropriate to fit the data:

$$1 - G(t) = [1 - F(t - \theta)]^{\eta}, \quad \text{for any } t \in \Re^1.$$

where $G(t)$ and $F(t)$ are the two unknown distributions for survival times from two different groups.

This is a special case of our research. We only have two parameters θ and η, but $r > 2$ constraints since we assume the data fit the hybrid model.

$$\sum_j \delta_{yj} g_k(y_j) \log(1 - v_j) = \sum_i \delta_{xi} \eta g_k(x_i - \theta) \log(1 - w_i), \quad k = 1, \dots, r.$$

where $g_k, k = 1, \dots, r$ are given functions satisfying some conditions, (y_j, δ_{yj}) and $(x_i, \delta_{xi})$are censored observations from two samples, and v_j and w_i are hazard jumps from the corresponding two samples.

We hope to get the same or similar value as Kim did, but the programme is still in progress.

5.3 Example 3

Data generated from
$X_1, \ldots, X_n \sim$ Exponential ($\lambda_1 = 0.02$). $C_1, \ldots, C_n \sim$ Exponential ($\lambda_2 = 0.005$)
$T_i = \min(X_i, C_i)$ and $\delta_i = I[X_i \leq C_i]$.
We want to estimate a single parameter θ. with *two* estimating equations:

$$\int I[0 \leq x \leq 20] d\Lambda_0(x) = 20\lambda_1 = \theta \tag{12}$$

$$\int I[20 \leq x \leq 40] d\Lambda_0(x) = \theta \tag{13}$$

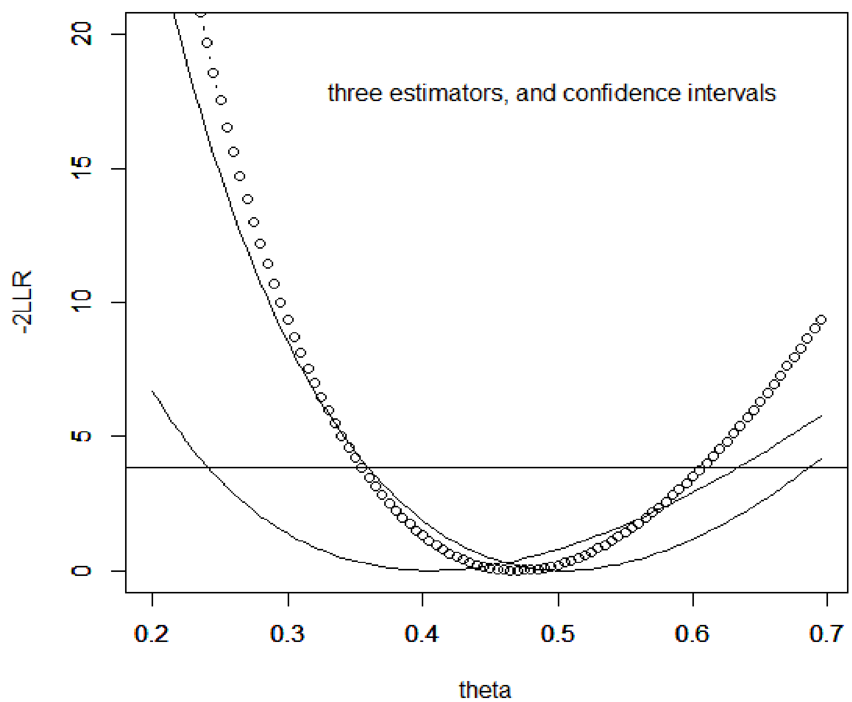
The true value of θ is 0.4.

Code for Simulation 1, q-q plot. (which show the over determined empirical likelihood ratio has chi square distribution under null hypothesis)

```
estfun <- function(x){
  cbind(as.numeric(x<=20),as.numeric(20<=x&x<=40)) }
llikratio <- rep(NA,1000)
dloglik <- rep(NA,1000)
nuloglik <- rep(NA,1000)
for (j in 1:1000){
   t <- rexp(50,0.02)
   d <- rexp(50,0.005)
   x <- pmin(t,d)
   cen <- as.numeric(t <= d)
   newtemp <- newdataclean(x=x,d=cen, fun=estfun)
nuloglik[j] <- newloglik(newtemp$funt, newtemp$eve, newtemp$rsk,
               newtemp$n, maxit=25, newtemp$K12,0.4)$"nloglik"
testthetas<-matrix(NA, ncol=101, nrow=1000)
testloglik<-matrix(NA,ncol=101, nrow=1000)

for (i in 1:101){
testthetas[j,i] <- newtemp$ltheta+(i-1)/100*(newtemp$rtheta-newtemp$ltheta)
testloglik[j,i] <- newloglik(newtemp$funt, newtemp$eve, newtemp$rsk,
           newtemp$n, maxit=25, newtemp$K12, testthetas[j,i])$"nloglik" }
dloglik[j] <- max(-testloglik[j,])
llikratio[j] <- 2*(dloglik[j]+nuloglik[j]) }
plot(qchisq(1:1000/1001,1), sort(llikratio), xlab="chisq(1)
               quantiles", ylab= "-2 log likelihood ratio")
abline(a=0,b=1)
```

Next plot shows three log likelihood ratio value vs. θ curves. They are obtained using just estimating equation 1, just estimating equation 2, or both estimating equations. The location of the three minimums are the three MEL estimators. The curvature at the minimum reflect the variance of the estimators. We see that the over determined curve (the one with small bubbles) has minimum location very close to other two, and is sandwiched in between. The over determined curve also has the largest curvature. In other words, the over determined line has the largest second derivative at the minimum. We recall the second derivative is related to the information number. The horizontal line is drawn at 3.84, the 95 percentile of chi square df=1. Each curve meet the 3.84 line at two places, they are the lower and upper limit of the 95% confidence interval. We clearly see that the confidence interval based on the over determined curve are the shortest confidence interval.



R code for simulation 2

```
t <- rexp(100, 0.02)
d <- rexp(100, 0.005)
x <- pmin(t,d)
cen <- as.numeric(t <= d)
thetas1 <- 0.2+(0:99)*0.5/100
result1 <- rep(NA,100)
result2 <- rep(NA,100)
est1 <- function(x){as.numeric(x<=20)}
est2 <- function(x){as.numeric(20<=x & x<=40)}

for (i in 1:100){
```

```
  result1[i]<-emplikH1.test(x=x,d=cen,theta=thetas1[i],fun=est1)$"-2LLR"
  result2[i]<-emplikH1.test(x=x,d=cen,theta=thetas1[i],fun=est2)$"-2LLR"
  }
result <- rep(NA,100)
thetas2 <- thetas1

nuloglik <- rep(NA,100)
dloglik <- rep(NA,100)
testthetas <- matrix(NA,ncol=101,nrow=100)
testloglik <- matrix(NA,ncol=101,nrow=100)
newtemp <- newdataclean(x=x,d=cen, fun=estfun)
for (i in 1:100){
   nuloglik[i]<- newloglik(newtemp$funt,newtemp$eve, newtemp$rsk,
         newtemp$n, maxit = 25, newtemp$K12,thetas2[i])$"nloglik"
   for (j in 1:101){
    testthetas[i,j]<-newtemp$ltheta+(j-1)/100*(newtemp$rtheta-newtemp$ltheta)
    testloglik[i,j]<- newloglik(newtemp$funt, newtemp$eve, newtemp$rsk,
    newtemp$n, maxit = 25, newtemp$K12, testthetas[i,j])$"nloglik"
   }
dloglik[i]<- max(-testloglik[i,])
result[i]<- 2*(dloglik[i]+nuloglik[i])
}

plot(thetas1,result1,ylim=c(0,20),xlab="theta",
     ylab="-2loglikelihood ratio",main="Constraint (1)")
par(new=TRUE)
plot(thetas1,result2,ylim=c(0,20),xlab="theta",
     ylab="-2loglikelihood ratio",main="Constraint (2)")
par(new=TRUE)
plot(thetas2,result,ylim=c(0,20),xlab="theta",
     ylab="-2loglikelihood ratio", main="Constraints (1)&(2) ")
```

The above function needs the following to run. These are functions similar to those in the package `emplik`, but are modified to specifically work for the over determined hazard estimating equations.

```
newdataclean <- function(x, d, y = -Inf, fun, itertrace = FALSE) {
x <- as.vector(x)
n <- length(x)
if (n <= 2) stop("Need more observations in x")
if (length(d) != n) stop("length of x and d must agree")
if ( any( (d != 0) & (d != 1) ) ) stop("d must be 1 or 0, for death/censor")
if (!is.numeric(x)) stop("x must be numeric, the observed times")

newdata <- Wdataclean2(z=x, d=d)
temp <- DnR(newdata$value, newdata$dd, newdata$weight,y = y)

jump <- (temp$n.event)/temp$n.risk
funtime <- as.matrix(fun(temp$times))
if (ncol(funtime) !=2)
stop("check the output dim of fun")
esttheta <- t(jump) %*% funtime
if (itertrace) print(c("thetahat=", esttheta))
ltheta <-min(esttheta[1], esttheta[2])
rtheta <-max(esttheta[1], esttheta[2])

index <- (jump < 1)
K12 <-rep(0,2)
tm1 <- temp$times[!index]
if (length(tm1) > 1) stop("more than 1 places jump>=1 in x?")
if (length(tm1) > 0) {
K12 <- K12 + as.vector(fun (tm1))}
eve <-temp$n.event[index]
tm <- temp$times[index]
rsk<- temp$n.risk[index]

jmp <- jump [index]
funt <- as.matrix(fun (tm))
list(funt=funt, eve=eve, rsk=rsk, ltheta=ltheta, rtheta=rtheta,
K12=K12,n=n, jump=jump, tm=tm)
}
```

```
newgradf <- function(lam, funt, eve, rsk, K, n) {
   arg <- as.vector(rsk + funt%*% lam)
    VV <- (eve * llogp(arg, 1/n)) %*% funt - K
    return(as.vector(VV))
}
```

```
newloglik <- function(funt, eve, rsk, n, maxit = 25, K12, theta,
               tola=1e-07, itertrace=FALSE){
TINY <- sqrt(.Machine$double.xmin)
if (tola < TINY)
tola <- TINY
lam <- rep(0,2)
```

```
#Newton-Raphson process.
nwts <- c(3^-c(0:3), rep(0, 12))
gwts <- 2^(-c(0:(length(nwts) - 1)))
gwts <- (gwts^2 - nwts^2)^0.5
gwts[12:16] <- gwts[12:16] * 10^-c(1:5)
nits <- 0
gsize <- tola + 1

while (nits < maxit && gsize > tola) {
grad <- newgradf(lam, funt, eve, rsk, K = theta - K12, n = n)
gsize <- mean(abs(grad))
arg <- as.vector(rsk + funt %*% lam)
ww <- as.vector(-llogpp(arg, 1/n))^0.5
tt <- sqrt(eve) * ww
HESS <- - (t(funt * tt) %*% (funt * tt) )
nstep <- as.vector(-solve(HESS, grad))
gstep <- grad
if (sum(nstep^2) < sum(gstep^2))
gstep <- gstep * (sum(nstep^2)^0.5/sum(gstep^2)^0.5)
ninner <- 0

for (i in 1:length(nwts)) {
lamtemp <- lam + nwts[i] * nstep + gwts[i] * gstep
ngrad <- newgradf(lamtemp, funt, eve, rsk, K = theta - K12, n = n)
ngsize <- mean(abs(ngrad))
if (ngsize < gsize) {
lam <- lamtemp
ninner <- i
break
}
}
nits <- nits + 1
if (ninner == 0)
nits <- maxit
if (itertrace) print(c(lam, gsize, ninner))
}
#Calculate the log-likelihood.
lamfun <- as.vector(funt%*%lam)
onePlamf <- (rsk+lamfun)/rsk
nloglik <- sum(eve*llog(onePlamf*rsk,1/n))+sum(eve*llogp(onePlamf, 1/n))
list(nloglik=nloglik, lambda=lam) }
```

6 Concluding Remarks

We have shown that the proposed empirical likelihood ratio test statistic with the constraints (5) in terms of hazards has an approximate chi-square distribution with p degrees of freedom under the null hypothesis. Confidence regions for $\boldsymbol{\theta}$ can be obtained by inverting the chi-square test.

This is a complement work to that of Qin and Lawless for handling censored data. However, computational issues arise as to the best ways to obtain T and $\hat{\boldsymbol{\theta}}$, especially in multi-dimensional parameter case. Further research on the calculation problem is still needed.

The reason we choose to use hazard-type empirical likelihood and constraints (or estimating functions) is that with right censored data we can easily get an estimate of the hazard function through Theorem 1.

Usually the estimating functions are given in terms of the CDF, like those studied by Qin and Lawless. Because most estimating equations are derived from the expectations, thus the integral with respect to CDF.

However, often the estimating equations in terms of the CDF can be approximated with estimating equations in terms of hazard, by choosing a proper $g^*(\cdot)$ function. It then is clear that the original (CDF estimating equations) problem can be transformed into an equivalent problem in hazard, and then treated using the results of this paper. It will be discussed in a forthcoming paper.

References

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

[2] DiCiccio, T.J., Hall, P. and Romana, J.P. (1989), "Comparisons of Parametric and Empirical Likelihood Functions." *Biometrika*.**76**, 465–476.

[3] DiCiccio, T.J. and Romana, J.P. (1989), "On Adjustments to the Signed Root of the Empirical Likelihood Ratio Statistic." *Biometrika*.**76**, 447–456.

[4] Hu, Yanling (2011), *Some Contributions to the Censored Empirical Likelihood with Hazard-Type Constraints.* Ph.D. Dissertation, University of Kentucky.

[5] Jin, Z., Lin, D.Y., Wei, L.J. and Ying, Z. (2003), "Rank-based Inference for the Accelerated Failure Time Model." *Biometrika.* **90**, 341–353.

[6] Kim, K. (2003), "Empirical Likelihood Ratio Method When Additional Information Is Known." *Ph.D Dissertation, Department of Statistics, University of Kentucky.*

[7] Kim, K. and Zhou, M. (2004), "Symmetric Location Estimation/Testing by Empirical Likelihood." *Communications in Statistics: Theory and Methods.* **33**, 2233–2243.

[8] Li, G. (1995), "On Nonparametric Likelihood Ratio Estimation of Survival Probabilities for Censored Data." *Statist. and Prob. Letters.* **25**, 95–104.

[9] Li, G., Li, R. and Zhou, M. (2005), "Empirical Likelihood in Survival Analysis." In *Contemporary Multivariate Analysis and Design of Experiments* Edited by J. Fan and G. Li. pp. 337–350. The World Scientific Publisher.

[10] Murphy, S.A. (1995), "Likelihood Ratio-based Confidence Intervals in Survival Analysis." *J. Amer. Statist. Assoc.*, **90**, 1399–1405.

[11] Naik-Nimbalkar, U.V. and Rajarshi, M.B. (1997), "Empirical likelihood ratio test for equality of k medians in censored data." *Statist. and Prob. Letters.* **34**, 267–273.

[12] Owen, A.B. (1988), "Empirical Likelihood Ratio Confidence Intervals For A Single Functional." *Biometrika*, **75**, 237–249.

[13] Owen, A.B. (1990), "Empirical Likelihood Ratio Confidence Regions." *Ann. Stat.*, **18**, 90–120.

[14] Owen, A.B. (1991), "Empirical Likelihood for Linear Models." *Ann. Stat.*, **19**, 1725–1747.

[15] Owen, A.B. (2001), *Empirical Likelihood.* Chapman & Hall/CRC.

[16] Pan, X.R. (1997), "Empirical Likelihood Ratio Method for Censored Data." *Ph.D Dissertation, Department of Statistics, University of Kentucky.*

[17] Pan, X.R. and Zhou, M. (2002), "Empirical Likelihood Ratio in Terms of Cumulative Hazard Function for Censored Data." *Journal of Multivariate Analysis*, **80**, 166-188.

[18] Qin, J. and Lawless, J. (1994), "Empirical Likelihood and General Estimating Equations." *Ann. Stat.*, **22**, 300-325.

[19] Thomas, D.R. and Grunkemeier, G.L. (1975), "Confidence Interval Estimation of Survival Probabilities for Censored Data." *J. Amer. Statist. Assoc.*, **70**, 865-871.

[20] Zhou, M. (2006). "The Cox proportional hazards model with partially known baseline" In *Random Walk, Sequential Analysis and Related Topics*, Chao Agnes Hsiung, Zhiliang Ying and Cun-Hui Zhang Eds. World Scientific, Singapore, ISBN 978-981-270-355-2. pp 215-232.