Notes on Censored EL, and Harzard

Mai Zhou

In survival analysis, the statistics involving the hazard functions are often more relevant and also usually easier to handle mathematically then those involving the means or distributions. For example, it is easier to show the Nelson-Aalen estimator is an NPMLE of the cumulative hazard function compared to the Kaplan-Meier estimator (which is an NPMLE of distribution function). The log rank test is often more appropriate than the test based on means.

However, there is a catch: the (cumulative) hazard function $\Lambda(t)$ has two distinct pairs of formula connecting with the distribution function $F(t)$, one for the continuous hazard, one for the discrete hazard.

Discrete hazard

$$\Delta\Lambda(t) = \frac{\Delta F(t)}{1 - F(t-)}, \quad \text{and} \quad 1 - F(t) = \prod_{s_i \le t} [1 - \Delta\Lambda(s_i)]$$

Continuous hazard

$$\Lambda(t) = -\log[1 - F(t)] \quad \text{and} \quad 1 - F(t) = \exp[-\Lambda(t)]$$

Therefore there are two versions of everything related to hazard: two versions of empirical likelihood, two versions of the null hypothesis. And later we will proof two versions of the Empirical likelihood ratio Wilks theorem.

The continuous version of empirical likelihood is also called the Poisson empirical likelihood; the discrete version is also called the binomial empirical likelihood. The discrete version is a true likelihood, the continuous version is an approximation

1 Hazard Empirical Likelihood: continuous version

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 function $\Lambda_0(t)$. 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. Only the censored observations, (T_i, δ_i), are available to us:

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

For the empirical likelihood in terms of hazard, we use the Poisson extension of the likelihood (Murphy 1995), and it is defined as

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

where $\Delta\Lambda(t) = \Lambda(t+) - \Lambda(t-)$ is the jump of Λ at t.

Remark: The term $\exp(-\Lambda(T_i))$ in the first line above has its origin in the continuous formula, yet in the second line we assume a discrete $\Lambda(\cdot)$.

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

$$EL = \prod_{i=1}^{n} [w_i]^{\delta_i} \exp\{-\sum_{j=1}^{n} w_j I[T_j \le T_i]\} \ ,$$

and the log likelihood is

$$\log EL = \sum_{i=1}^{n} \left\{ \delta_i \log w_i - \sum_{j=1}^{n} w_j I[T_j \le T_i] \right\} \ .$$

If we maximize the log EL above (without constraint) we see that $w_i = \frac{\delta_i}{R_i}$, where $R_i = \sum_j I[T_j \ge T_i]$. This is the well known Nelson-Aalen estimator: $\Delta\hat{\Lambda}_{NA}(T_i) = \frac{\delta_i}{R_i}$. If we define $R(t) = \sum_k I[T_k \ge t]$ then $R_i = R(T_i)$.

The next step in our analysis is to find a (discrete) cumulative hazard function that maximizes the log $EL(\Lambda)$ under the constraints (1):

$$\begin{aligned} \int_0^\infty g_1(t) d\Lambda(t) &= \theta_1 \\ \int_0^\infty g_2(t) d\Lambda(t) &= \theta_2 \\ \cdots & \quad \cdots \\ \int_0^\infty g_p(t) d\Lambda(t) &= \theta_p \end{aligned} \tag{1}$$

where $g_i(t)(i = 1, 2, ..., p)$ are given functions satisfy some moment conditions (specified later), and θ_i $(i = 1, 2, ..., p)$ are given constants. The constraints (1) can be written as (for discrete hazard)

$$\begin{aligned} \sum_{i=1}^{n} g_1(T_i) w_i &= \theta_1 \\ \sum_{i=1}^{n} g_2(T_i) w_i &= \theta_2 \\ \cdots & \quad \cdots \\ \sum_{i=1}^{n} g_p(T_i) w_i &= \theta_p \ . \end{aligned} \tag{2}$$

A similar argument as in Owen (1988) will show that we may restrict our attention in the EL analysis to those discrete hazard functions that are dominated by the Nelson-Aalen estimator: $\Lambda(t) \ll \hat{\Lambda}_{NA}(t)$. Owen (1988) restricted his attention to those distribution functions that $F(t) \ll$ the empirical distribution.

Since for discrete hazard functions, the last jump must be one, this imply that $w_n = \delta_n = \Delta\hat{\Lambda}_{NA}(T_n)$ always. The next theorem gives the other jumps.

Theorem 1 *If the constraints above are feasible (which means the maximum problem has a hazard solution), then the maximum of log $EL(\Lambda)$ under the constraint is obtained when*

$$\begin{aligned} w_i &= \frac{\delta_i}{R_i + n\lambda^T G(T_i)\delta_i} \\ &= \frac{\delta_i}{R_i} \times \frac{1}{1 + \lambda^T(\delta_i G(T_i)/(R_i/n))} \\ &= \Delta\hat{\Lambda}_{NA}(T_i)\frac{1}{1+\lambda^T Z_i} \end{aligned}$$

where

$$G(T_i) = \{g_1(T_i), ..., g_p(T_i)\}^T, \quad Z_i = \frac{\delta_i G(T_i)}{R_i/n} = \{Z_{1i}, ..., Z_{pi}\}^T \quad \text{for} \;\; i = 1, 2, \ldots, n.$$

and $\lambda = \{\lambda_1, ..., \lambda_p\}^T$ *is the solution of the following equations*

$$\sum_{i=1}^{n-1} \frac{1}{n} \, \frac{Z_{ki}}{1+\lambda^T Z_i} + g_k(T_n)\delta_n = \theta_k \quad \text{for} \;\; k = 1, \ldots, p \ . \tag{3}$$

Proof. Use Lagrange Multiplier to find the constrained maximum of log EL. See Pan and Zhou (2002) for details.

Similar to the proof in the paper, it can also be shown the following Wilks theorem hold.

Theorem 2 *Let $(T_1, \delta_1), \ldots, (T_n, \delta_n)$ be n pairs of i.i.d. random variables as defined above. Suppose $g_i \;\; i = 1, \ldots, p$ are left continuous functions satisfy*

$$0 \;<\; \int \frac{|g_i(x) g_j(x)|}{(1-F_0(x))(1-G_0(x-))} d\Lambda(x) < \infty, \quad \text{all} \;\; 1 \le i, j \le p. \tag{4}$$

Furthermore, assume the matrix Σ, defined in the Lemma 2 below, is invertible.

Then, $\theta_0 = \{\int g_1(t)d\Lambda(t), ..., \int g_p(t)d\Lambda(t)\}^T$ will be a feasible vector with probability approaching one as $n \to \infty$ and

$$-2\log ELR(\theta_0) \xrightarrow{\mathcal{D}} \chi^2_{(p)} \quad as \quad n \to \infty$$

where $\log ELR(\theta_0) = \max \log EL(with\ constraints\ (2)) - \log EL(\hat{\Lambda}_{NA})$.

Proof. Here we briefly outline the proof. The complete proof is just a multivariate version of Pan and Zhou (2002). First, we need the following two lemmas. They are the Law of Large Numbers and CLT for the Nelson-Aalen estimator and can be proved via counting processes technique.

Lemma 1 *Under the assumption of Theorem 2, we have, for $1 \le k, r \le p$*

$$\frac{1}{n}\sum_{i=1}^{n} Z_{ki}Z_{ri} = \int \frac{g_k(t)g_r(t)}{R(t)/n} d\hat{\Lambda}_{NA}(t) \xrightarrow{P} \int \frac{g_k(x)g_r(x)}{(1-F_0(x))(1-G_0(x-))} d\Lambda_0(x)$$

as $n \to \infty$ where

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

Lemma 2 *Under the assumption of Theorem 2, we have*

$$\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} Z_i - \theta_0) = \sqrt{n}(\sum_{i=1}^{n} G(T_i)\Delta\hat{\Lambda}_{NA}(T_i) - \theta_0) \xrightarrow{\mathcal{D}} MVN(0, \Sigma),$$

as $n \to \infty$ where the limiting variance-covariance matrix is

$$\Sigma_{kr} = \int \frac{g_k(x)g_r(x)}{(1-F_0(x))(1-G_0(x-))} d\Lambda_0(x) \quad for \quad 1 \le k, r \le p; \tag{5}$$

and

$$\theta_0 = \{\int g_1(t)d\Lambda_0(t), \cdots, \int g_p(t)d\Lambda_0(t)\}^T .$$

We define the matrix A as below. Since $A \to \Sigma$ as $n \to \infty$ (Lemma 1) and we assumed Σ is invertible and thus positive definite, we conclude that for large enough n the symmetric matrix A is invertible. Next, we show the solution of λ to the constraint equations (3) is

$$\lambda = \lambda^* = A^{-1}b + o_p(n^{-1/2}) \tag{6}$$

where

$$A_{kr} = \frac{1}{n}\sum_{i=1}^{n} Z_{ki}Z_{ri} \quad \text{for} \quad 1 \le k, r \le p.$$

$$b = \{\frac{1}{n}\sum_{i=1}^{n} Z_{1i} - \theta_1, \quad \cdots, \quad \frac{1}{n}\sum_{i=1}^{n} Z_{pi} - \theta_p\}^T$$

This can be proved by an expansion of equation (3).

The stickier question is that why an expansion of (3) is valid and why the remainder term is $o_p(n^{-1/2})$. We deal with this in appendix by showing the solution λ^* is small in the first place.

Define

$$f(\lambda) = \log EL(w_i(\lambda)) = \sum_{i=1}^{n} \left(\delta_i \log w_i(\lambda) - \sum_j w_j(\lambda) I[T_j \leq T_i] \right)$$

and the test statistic $-2 \log ELR(\theta_0)$ can be expressed as

$$-2 \log ELR = 2[f(0) - f(\lambda^*)] = 2[f(0) - f(0) - \lambda^{*T} f'(0) - 1/2 \lambda^{*T} f''(0) \lambda^* + \ldots].$$

Straight forward calculation show $f'(0) = 0$ and $f''(0) = A$. Therefore

$$-2 \log ELR = -\lambda^{*T} f''(0) \lambda^* + \ldots \tag{7}$$

simplify it to the following

$$-2 \log ELR(\theta_0) = n b^T A^{-1} b + o_p(1)$$

Finally, by Lemma 1 and Lemma 2, we get

$$-2 \log ELR(\theta_0) \xrightarrow{\mathcal{D}} \chi^2_{(p)} \quad as \quad n \to \infty \ .$$

Appendix. We now give a proof of the Lemma 1, 'law of large number' for the Z_i or for the integral of the Nelson-Aalen estimator. The results is obviously true if we impose more moment conditions. We, however, try to give a proof that only assume the finiteness of the limiting integration and without the extra moment condition. Also, we could allow the $g(t)$ function to be a random sequence of functions. Notice here the random variables Z_i are not independent.

Lemma 1 Under assumptions below, for given $k = 1, 2, \cdots, p$ we have

$$\frac{1}{n}\sum_{i=1}^{n} Z_{ki}^2 = \int \frac{g_k^2(t)}{R(t)/n} d\hat{\Lambda}_{NA}(t) \xrightarrow{P} \int \frac{g_k^2(x)}{(1-F_0(x))(1-G_0(x-))} d\Lambda_0(x)$$

Assumptions: (We omit the subscript k. These conditions should hold for all $k = 1, 2, \cdots, p$.)

(1) The limit that you hope to converge to must be finite: i.e. $\int_0^\infty \frac{g^2(x)}{(1-F_0(x))(1-G_0(x-))} d\Lambda_0(x) < \infty$

(2) If we use $g_n(t)$ on the left side, then we need to assume: it converges uniformly in any finite intervals, i.e. for any finite τ, $\sup_{t<\tau} |g_n(t) - g(t)|$ go to zero in probability and the ratio $\sup_i |g_n(T_i)/g(T_i)|$ is bounded in probability. These two conditions are satisfied by the empirical distributions, the Kaplan-Meier estimator and the Nelson-Aalen estimator.

Notice in the CLT of the martingale (Lemma 2), we will further require that $g_n(t)$ be predictable functions.

Proof:

We first proof the LLN for $\int_0^\tau$ for any given finite τ.

$$\int_0^\tau \frac{g_n^2(t)}{R(t)/n} d\hat{\Lambda}(t) = \sum_i I[T_i < \tau] \frac{g_n^2(T_i)}{R(T_i)/n} \frac{\Delta N(T_i)}{R(T_i)} \tag{8}$$

Minus and plus the term (recall $\Delta N(T_i) = \delta_i$)

$$\frac{1}{n}\sum_i I[T_i < \tau] \frac{g^2(T_i)}{[1 - H(T_i-)]^2} \delta_i$$

in the above, and regroup, we get

$$= \frac{1}{n}\sum_i I[T_i < \tau]\delta_i \left(\frac{g_n^2(T_i)}{[R(T_i)/n]^2} - \frac{g^2(T_i)}{[1 - H(T_i-)]^2} \right) + \frac{1}{n}\sum_i I[T_i < \tau] \frac{g^2(T_i)\delta_i}{[1 - H(T_i-)]^2} \tag{9}$$

The first term above is bounded by

$$\frac{1}{n}\sum_i I[T_i < \tau] \left| \frac{g_n^2(T_i)}{[R(T_i)/n]^2} - \frac{g^2(T_i)}{[1 - H(T_i-)]^2} \right| \delta_i \le \sup_{t<\tau} \left| \frac{g_n^2(t)}{[R(t)/n]^2} - \frac{g^2(t)}{[1 - H(t-)]^2} \right|$$

The term inside the absolute sign is uniformly convergent to zero, by the assumption 2 on $g_n(t)$. And it is well known that $R(t)/n \to [1 - H(t-)]$ uniformly. Therefore the reciprocal of it is at least uniformly convergent on $t \le \tau$.

The last term in (9) above is an iid sum with respect to (T_i, δ_i). By the classic LLN, it converge to its expectation, which is

$$E\left(I[T_i < \tau] \frac{g^2(T_i)\delta_i}{[1-H(T_i-)]^2} \right) = \int_0^\tau \frac{g^2(t)}{1-H(t-)} d\Lambda_0(t)$$

which by assumption 1 is finite. This proves that the Lemma holds for any finite τ.

We need to take care of the tail: $\int_\tau^\infty$. By assumption 1,

$$\int_\tau^\infty \frac{g^2(t)}{1-H(t-)} d\Lambda_0(t)$$

can be made arbitrary small by selecting a large τ. (say smaller than ϵ/C)

Since the ratio $g_n(T_i)/g(T_i)$ and $[1-H(T_i-)]/[R(T_i)/n]$ are both uniformly (in $\sup_{1\le i \le n}$) bounded in probability (assumption 2, and property of empirical distribution function) we have, that the term

$$\sum_i I[T_i \ge \tau] \frac{g_n^2(T_i)}{R(T_i)/n} \frac{\Delta N(T_i)}{R(T_i)}$$

is bounded in probability by

$$\le C \frac{1}{n} \sum_i I[T_i \ge \tau] \frac{g^2(T_i)\delta_i}{[1-H(T_i-)]^2}$$

This summation/average above converges to its mean (since it is an iid average)

$$C \int_\tau^\infty \frac{g^2(t)}{1-H(t-)} d\Lambda_0(t)$$

the absolute value of which, in turn, is smaller than the pre-selected ϵ. This finishes the proof.

Remark: For future work on the Edgeworth expansion/Bartlett correction, We need a LLN like the above but with rates, for the Edgeworth analysis of the empirical likelihood. Under suitable assumption, the following should be true (LIL):

$$\int g^2(t) d\hat{\Lambda}_n(t) - \int g^2(t) d\Lambda(t) = O(\sqrt{\frac{\log\log n}{n}}) \quad a.s.$$

or similar to lemma 1

$$\int \frac{g^2(t)}{R(t)/n} d\hat{\Lambda}_n(t) - \int \frac{g^2(t)}{1-H(t-)} d\Lambda(t) = O(\sqrt{\frac{\log\log n}{n}}) \quad a.s.$$

The proof of Lemma 2 is direct consequence of the martingale central limit theorem, See, for example, Kalbfleisch and Prentice 2002 chapter 5, Theorem 5.1 in particular.

Remark: A better normal approximation for the martingales, which has Edgeworth expansion, is given by Lai and Wang.

$$P(\sqrt{n}(\hat{\Lambda}_n(t) - \Lambda(t)) \le \sigma z) = \Phi(z) - n^{-1/2}\phi(z)P_1(z) - n^{-1}\phi(z)P_2(z) + o(n^{-1})$$

Appendix. PROOF OF λ_0 SMALL.

We give a proof that validates the expansion of (3). In other words, we show the solution of (3) is small. We want to show $\lambda^T Z_i$ is small uniformly over i. We shall denote the solution as λ_0.

Lemma 3: Suppose $M_n = o_p(n^{1/2})$, then we have

$$\lambda_n = O_p(n^{-1/2}) \qquad \text{if and only if} \qquad \frac{|\lambda_n|}{1+|\lambda_n M_n|} = O_p(n^{-1/2}) \ .$$

PROOF: Homework.

Lemma 4: If $X_1, \cdots, X_n$ are identically distributed, and $E(X_1)^2 < \infty$, then we have $M_n = \max_{1\le i \le n} |X_i| = o_p(n^{1/2})$.

PROOF Since $\{M_n > a\} = \cup\{|X_i| > a\}$, we compute

$$P(M_n > n^{1/2}) = P(\bigcup_{i=1}^{n}(|X_i| > n^{1/2})) \le \sum_{i=1}^{n} P(|X_i| > n^{1/2}) \ .$$

By the identical distribution assumption,

$$= nP(|X_1| > n^{1/2}) = nP(X_1^2 > n) \ .$$

Since $EX_1^2 < \infty$, the right hand side above $\to 0$ as $n \to \infty$. Similar proof will show that, if $E|X_i|^p < \infty$ then $M_n = o_p(n^{1/p})$.

Lemma 5: We compute

$$E \ \frac{\delta_i g^2(T_i)}{[1-F(T_i-)]^2[1-G(T_i)]^2} = \int \frac{g^2(t)}{[1-F(t-)][1-G(t)]} d\Lambda(t) \ .$$

Therefore if we assume $\int \frac{g^2(t)}{[1-F(t-)][1-G(t)]} d\Lambda_0(t) < \infty$ then

$$M_n^* = \max_{1\le i\le n} \frac{\delta_i |g(T_i)|}{[1-F(T_i-)][1-G(T_i)]} = o_p(n^{1/2})$$

by Lemma 4 and 5.

Now, using a theorem of Zhou (1992) we can replace the denominator of M_n^* by $R(T_i)/n$:

$$M_n = \max_{1\le i\le n} |Z_i| = \max \frac{\delta_i |g(T_i)|}{R_i/n} \le M_n^* \max_i \frac{[1-F(T_i)][1-G(T_i)]}{R_i/n} = o_p(n^{1/2}) \ .$$

Now we proceed as follows: denote the solution by λ_0. We notice that for all i, $1+\lambda_0^T Z_i \ge 0$ since the solution w_i given in Theorem 1 must give rise to a legitimate jump of the hazard function, which must be ≥ 0. Clearly $w_i \ge 0$ imply $1+\lambda_0^T Z_i \ge 0$.

First we rewrite the equation (3) and notice that λ_0 is the solution of the following equation $0 = l(\eta)$.

$$0 = l(\lambda_0) = (\theta_0 - \frac{1}{n}\sum Z_i) + \frac{\lambda_0}{n}\sum_{i=1}^{n-1}\frac{Z_i^2}{1+\lambda_0 Z_i} \qquad (10)$$

Therefore,

$$\theta_0 - \frac{1}{n}\sum Z_i = -\frac{\lambda_0}{n}\sum_{i=1}^{n-1}\frac{Z_i^2}{1+\lambda_0 Z_i} \tag{11}$$

$$\left|\theta_0 - \frac{1}{n}\sum Z_i\right| = \frac{|\lambda_0|}{n}\left|\sum_{i=1}^{n-1}\frac{Z_i^2}{1+\lambda_0 Z_i}\right| \tag{12}$$

Since for every term (at least when $\delta_i = 1$, or $Z_i^2 > 0$), $Z_i^2/(1+\lambda_0 Z_i) \geq 0$, therefore we have

$$\left|\theta_0 - \frac{1}{n}\sum Z_i\right| = \frac{|\lambda_0|}{n}\sum\frac{Z_i^2}{|1+\lambda_0 Z_i|}$$

Replace the denominators $1+\lambda_0 Z_i$ by its upper bound: for any i we have

$$|1+\lambda_0 Z_i| \leq 1 + |\lambda_0| M_n$$

we got a lower bound in the fraction

$$\left|\theta_0 - \frac{1}{n}\sum Z_i\right| \geq \frac{|\lambda_0|}{1+|\lambda_0|M_n}\ \frac{1}{n}\sum_{i=1}^{n-1} Z_i^2 \geq 0$$

Since $\theta_0 - 1/n\sum Z_i = O_p(n^{-1/2})$ (CLT, Lemma 2), We see that

$$\frac{|\lambda_0|}{1+|\lambda_0|M_n}\ \frac{1}{n}\sum Z_i^2 = O_p(n^{-1/2})$$

and obviously $\frac{1}{n}\sum Z_i^2 = O_p(1)$ (Lemma 1) thus we must have

$$\frac{|\lambda_0|}{1+|\lambda_0|M_n} = O_p(n^{-1/2})\ .$$

By Lemma 3 above we must finally have $\lambda_0 = O_p(n^{-1/2})$.

As a consequence, we also have $\lambda_0 M_n = o_p(1)$ and thus $\lambda_0 Z_i = o_p(1)$ uniformly for all i.

Problem: Using the similar techniques to show that the empirical likelihood ratio under a sequence of local alternative hypothesis has a non-central chi squared distribution (asymptotically). (similar to Owen 1988 Corollary 1).

PROOF OF THE LOWER BOUND FOR r_h

The inequality we are going to proof is just a Cauchy Schwarz inequality: $\langle a, b\rangle^2 \le \|a\|\|b\|$. But it is hidden behind several 'Advanced-time' and 'self-consistency' transformations.

Assume $\int \phi(t)dF(t) = 0$ (mean zero).

Theorem:

$$r_h = \frac{\sigma^2_{KM}(\phi) \times V_h}{\int \phi(t)h(t)dF(t)} \ge 1$$

and the minimum of 1 is achieved when $h(t)$ satisfy the equation (?) below. If we replace F everywhere in the above by the Kaplan-Meier estimator $\hat{F}_{KM}$ based on n observations, a similar inequality (finite sample version) also hold.

Lemma: (Advanced-time) [Efron and Johnston 1990] Define the 'advanced time' transformation for a function $g(t)$ with respect to a CDF $F(\cdot)$

$$\bar{g}(s) = \frac{\int_{(s,\infty)} g(x)dF(x)}{1-F(s)} = E_F[g(X)|X > s] \ .$$

Then we have

$$\mathrm{Var}_F(g) = \int [g(x) - E_F g]^2 dF(x) = \int [g(x) - \bar{g}(x)]^2 dF(x)$$

and

$$Cov_F(\phi, h) = \int [\phi(t) - E_F\phi]h(t)dF(t) = \int [\phi(t) - \bar{\phi}(t)][h(t) - \bar{h}(t)]dF(t)$$

where $E_F g = \int g(x)dF(x)$.

Proof: The result for the variance is directly from Efron and Johnston (1990). The result for the covariance can proved similarly. QED.

We are going to use this Lemma with the CDF being either the Kaplan-Meier estimator $\hat{F}_{KM}$ or F.

The asymptotic variance of the Kaplan-Meier integral $\int \phi(t)d\hat{F}_{KM}(t)$ can be written as (see Akritas 2000)

$$\sigma^2_{KM}(\phi) = \int [\phi(x) - \bar{\phi}(x)]^2 \frac{dF(x)}{1-G(x-)}$$

and a finite sample version is: the variance of the Kaplan-Meier integral can be consistently estimated by

$$\sum_i [\phi(t_i) - \bar{\phi}(t_i)]^2 \frac{d\hat{F}_{KM}(t_i)}{1-\hat{G}_{KM}(t_i)} \ .$$

Lemma (self-consistency) For the Kaplan-Meier estimator $\hat{F}_{KM}$ we have that for any function g

$$\sum_i g(t_i)\Delta\hat{F}_{KM}(t_i) = \sum_i \frac{\delta_i}{n} g(t_i) + \sum_i \frac{(1-\delta_i)}{n} \frac{\sum_{t_j > t_i} g(t_j)\Delta\hat{F}_{KM}(t_j)}{1-\hat{F}_{KM}(t_i)} \ .$$

Proof: The probability corresponding to $g(t_k)$ on left hand side is $\Delta\hat{F}(t_k)$. The probability for $g(t_k)$ on the left hand side is precisely those given by Turnbull (1976) self-consistent equation. QED.

Finally we need to verify the V_h is the norm of the said function. This is actually a finite sample quantity

Lemma (Simplification of second derivative)

Recall we have computed the second derivative and it is

$$\begin{aligned} f^{''}(0)/n &= (\sum_{i=1}^n h(x_i, T_i)\Delta \hat{F}_{KM}(T_i))^2 - \sum_{i=1}^n h^2(x_i, T_i)\Delta \hat{F}_{KM}(T_i) \\ &+ \sum_{i=1}^n \frac{1-\delta_i}{n} \frac{\sum_{j:T_j>T_i} h^2(x_j, T_j)\Delta \hat{F}_{KM}(T_j)}{1-\hat{F}_{KM}(T_i)} - \sum_{i=1}^n \frac{1-\delta_i}{n} \frac{(\sum_{j:T_j>T_i} h(x_j, T_j)\Delta \hat{F}_{KM}(T_j))^2}{(1-\hat{F}_{KM}(T_i))^2} . \end{aligned}$$

Use the identity $Eg^2 - (Eg)^2 = E[g - Eg]^2$ once for the first two terms, and again n times for the last two terms in the above, we get:

$$\begin{aligned} -f^{''}(0)/n &= \sum_{i=1}^n [h(x_i, T_i) - E_{F_{KM}} h]^2 \Delta \hat{F}_{KM}(T_i) \\ &- \sum_{i=1}^n \frac{1-\delta_i}{n} \frac{\sum_{j:T_j>T_i} [h(x_j, T_j) - E(h|t > T_i)]^2 \Delta \hat{F}_{KM}(T_j)}{1-\hat{F}_{KM}(T_i)} . \end{aligned}$$

Now use the Advanced-time Lemma for variance $n+1$ times we get:

$$\begin{aligned} -f^{''}(0)/n &= \sum_{i=1}^n [h(x_i, T_i) - \bar{h}(T_i)]^2 \Delta \hat{F}_{KM}(T_i) \\ &- \sum_{i=1}^n \frac{1-\delta_i}{n} \frac{\sum_{j:T_j>T_i} [h(x_j, T_j) - \bar{h}(T_j)]^2 \Delta \hat{F}_{KM}(T_j)}{1-\hat{F}_{KM}(T_i)} . \end{aligned}$$

We then apply the self-consistent Lemma to the above and finally have

$$V_h = -f^{''}(0)/n = \sum_{i=1}^n \frac{\delta_i}{n} [h(x_i, T_i) - \bar{h}(T_i)]^2 \ .$$

QED.

Recall for the Kaplan-Meier estimator, we have

$$\frac{\delta_i}{n[1-\hat{G}_{KM}(t_i)]} = \Delta \hat{F}_{KM}(t_i)$$

and thus

$$V_h = -f^{''}(0)/n = \sum_i [h(t_i) - \bar{h}(t_i)]^2 [1-\hat{G}_{KM}(t_i)] \Delta \hat{F}_{KM}(t_i) \ .$$

For a finite n version of the inequality we stop here. For the asymptotic version, we let $n \to \infty$. This converge in the limit as $n \to \infty$

$$\lim_{n\to\infty} V_h = \int [h(t) - \bar{h}(t)]^2 [1 - G(t)] dF(t)$$

Therefore the inequality in the Theorem is just the Cauchy Scharwz inequality and the lower bound is achieved when

$$[h(t) - \bar{h}(t)]\sqrt{1 - G(t)} = \frac{[\phi(t) - \bar{\phi}(t)]}{\sqrt{1 - G(t)}} \quad a.s. \quad F(\cdot)$$

i.e.

$$[h(t) - \bar{h}(t)] = \frac{[\phi(t) - \bar{\phi}(t)]}{1 - G(t)} \quad a.s. \quad F(\cdot)\ . \tag{13}$$

Remark: There is a finite sample version of this inequality. Integrations becomes summations over n points.

Remark: This actually points out a least favorable direction of the distribution for the estimation of the Kaplan-Meier integral. And the quantity we dealt with in the theorem is in fact the information lower bound of the estimators.

Akritas, M. (2000). The central limit theorem under censoring. *Bernoulli* **6**, 1109-1120.

Efron, B. and Johnstone, I.M. (1990). Fisher's information in terms of the hazard rate. *Ann. Statist.* **18**, 38-62.

Residual Analysis, Model selection in the AFT model. (for the case p ¡ n]

(Buckley-James estimator, inverse censoring probability weighted estimator)

The investigations are mostly focused on the estimation of the model parameters, ASSUME the model is correct.

In practice, all models are wrong. So model selection is important. Residuals is an important instrument that model selection needs.

2 Estimating equations, Censored data

Empirical likelihood theory for iid, completely observed data are well developed. Owen's 2001 book covers.

However, for censored survival data, the empirical likelihood theory is sporadic. Some results are known but the picture is far from complete. Case in point, the censored data analog of Qin and Lawless (1994) theory for the estimating equations empirical likelihood is still nowhere to be found.

The theory of Qin and Lawless deals with *over-determined* estimating equations: we have more estimating equations than number of parameters. This is important in Economics, also important regression modeling as in CRQ, and AFT models.

Our general goal in this part: to develop Empirical Likelihood methods and theory for censored data estimating equations, in particular we shall establish a result similar to Qin and Lawless dealing with over-determined estimating equations with censored survival data.

We mostly work with right censored data in this investigation but techniques developed could be used for truncation or other censor type. The application of the theory includes the regression models (AFT, quantile, longitudenal).

One feature of the survival analysis is that many statistical models and inferences are in terms of the hazard function instead of the CDF. For example the regression model in terms of hazards is the well known Cox proportional hazards model. This feature carries over to the empirical likelihood: we shall investigate parallel results for empirical likelihood theory for the hazards as well as the distributions.

In the process of investigate both type of empirical likelihood, we propose to develop the equivalency (to be made clear in section 1.2) of the hazard based estimating equations to the distribution based estimating equations. So a unified theory ......

This is our game plan:

2.1 EL and Estimating Equations in terms of Hazard, Censored Data

Develop a complete EL theory with hazard based estimating equations and right censored data.

Many parameters, estimating equations, and This is to include the over-determined case, similar to Qin and Lawless.

Given the censored observations T_i, δ_i, we call those equations as estimating equations in term of hazard.

$$C_j = \sum_{i=1}^{n} h_j(t_i, \theta) \Delta \hat{\Lambda}_{NA}(t_i) \qquad j = 1, 2, \cdots, q.$$

Mimic the unbias requirement of the estimating equation in CDF, we require that $\int h_j(t, \theta) d\Lambda(t) = C_j$.

The estimating equations in terms of hazard are less common, but do exists. We use it here both as a tool to study the estimating equations in terms of CDF and in its own right.

For example the median can be defined as the solution of the estimating equation with censored data (T_i, δ_i)

$$\log 2 = \sum_{i=1}^{n} I[t_i \le \theta] \Delta \hat{\Lambda}_{NA}(t_i)$$

Similarly, the rank estimator of Tsiatis (1990), Lin, Ying and Wei (1990) can also be think of as the estimating equations in terms of hazard, where the estimator is based on the log-rank type test on the residuals.

The empirical likelihood for the censored observations should be defined as

$$EL_2 =$$

This can also be written in terms of hazard, using the relation between CDF and hazard function.

2.2 The Equivalency of Hazard and CDF Estimating Equations

We describe below what do we mean by and how we plan to establishing the equivalency of estimating equations in terms of hazard to the estimating equations in distribution. We work with censored data.

Given an estimating equation

$$0 = \frac{1}{n} \sum_{i=1}^{n} g(X_i, \theta),$$

we may rewrite it as

$$0 = \int g(x, \theta) d\hat{F}_n(x)$$

where $\hat{F}_n(\cdot)$ is the empirical distribution based on X_i's. The consistent consideration (or unbiasness) usually requires that $0 = \int g(x, \theta_0) dF(x)$, where $F(x)$ is the true CDF of X_i, θ_0 is the true parameter value. We have, by well known CLT that

$$\sqrt{n} \left[\int g(x, \theta) d\hat{F}_n(x) - \int g(x, \theta dF(x) \right]$$

converges to a normal distribution with zero mean.

This rewriting suggest the estimating equation with censored data: the censored data estimating equation is

$$0 = \int g(x, \theta) d\hat{F}_{KM}(x) \tag{14}$$

where $\hat{F}_{KM}(\cdot)$ is the Kaplan-Meier estimator based on the censored sample (T_i, δ_i). The unbias assumption $\int g(x, \theta_0) dF_0(x) = 0$ thus lead to

$$0 = \int g(x, \theta) d[\hat{F}_{KM}(x) - F(x)] \ .$$

If we can find a non-random function $g^*(t, \theta)$ such that

$$0 = \int g(x, \theta) d[\hat{F}_{KM}(x) - F(x)] \quad \text{if and only if} \quad 0 = \int g^*(t, \theta) d[\hat{\Lambda}_{NA}(t) - \Lambda(t)]$$

where $\Lambda_{NA}(\cdot)$ is the cumulative hazard function based on $\hat{F}_{KM}$, i.e.

$$\hat{\Lambda}_{NA}(t) = \int_{[0,t)} \frac{d\hat{F}_{KM}(s)}{1 - \hat{F}_{KM}(s-)}$$

(it is well known $\hat{\Lambda}_{NA}$ is just the Nelson-Aalen estimator) and Λ is based on F similarly then the estimating equation (**??**EE2)) can be replaced by the corresponding equations in terms of the hazard. The key point is that g^* cannot be random.

Unfortunately, the g^* that satisfy the "if and only if" requirement part above is hard to find. However, we do not have to find an exact "if and only if" substitute, the asymptotic equivalency can be established if the two are within a $o_p(1/\sqrt{n})$ of each other. Alternatively, the estimating equations do not have to be solved exactly, if a θ makes the value close to zero, it also constitutes a solution as pointed out by many authors REF()?

This is exactly what we propose to do. Based on the work of Akritas (2000). In Proposition 3 under some regularity conditions, he established that

$$\int g(s,\theta) d[\hat{F}_{KM}(s) - F(s)] = \int \tilde{g}(s,\theta) d[\hat{\Lambda}_{NA}(s) - \Lambda(s)] + o_p(n^{-1/2}) \tag{15}$$

where $\tilde{g}$ is non-random and given in his equation (9).

To use this type of equivalency in the EL analysis we not only need to establish this relationship between the Kaplan-Meier estimator and Nelson-Aalen estimator, but also for other estimators. The reason is that the likelihood ratio involves two likelihoods: one involves Kaplan-Meier/Nelson-Aalen in the denominator, the other has to do with maximization under constraints and are not achieved by the Kaplan-Meier/Nelson-Aalen estimator.

Fortunately, we believe the following result can be established alone the same line of Akritas (2000). For all discrete CDFs that are dominated by the Kaplan-Meier estimator and close to the Kaplan-Meier estimator:

$$F^* \ll \hat{F}_{KM} \quad \text{and} \quad \|F^* - \hat{F}_{KM}\| = O(1/\sqrt{n})$$

the above equivalency relationship ( ) still hold, when we replacing $\hat{F}_{KM}$ by F^* and the Nelson-Aalen estimator $\hat{\Lambda}_{NA}$ by

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

2.3 Unified EL Theory

With the equivalency results established in previous section as a powerful tool, we can derive the theory of the estimating equations for distribution functions along side the EL results for the hazards, which seems much tractable and we propose to develop further in section 1.1.

The idea of represent one in terms of the other has also been used for example in the iid representation of the Kaplan-Meier estimator by Lo and Chen (1995) and many others later. People have use the representation to establish the CLT of Kaplan-Meier integral, etc. as an iid sum. We carry this a step further in that the representation not only holds for the Kaplan-Meier/Nelson-Aalen but also for the (whole class of) constrained estimators.

Another benefit of the above proposed investigation is the nonparametric Maximum likelihood estimator in the over-determined estimating equations case with censored data. For example a central limit theorem could be obtained.

including the over-determined case. The theory itself will clarify some of the existing results as it clarify the regularity condition of the El theorem. Multivariate EL with censored data.

3 Comparison of Different EL approaches for Censored Data

Several EL approaches with censored data appeared in the literature in the last decade. Even with the same estimating equation, they differ in the definition of the likelihood function and the limiting null distribution is also different. Most cases, you either get a chi squared limiting distribution or a limiting distribution that can be described as linear combination of several independent chi square(df=1) random variables. In the latter case, some post adjust were also proposed to get a simpler limiting distributions.

We shall compare/clarify these different approaches to the censored EL, chi square or linear combination of chi squares in terms of Optimality of the resulting test: First, do they result the same estimator (MELE)? If not which estimator is more efficient? If yes, then are the resulting likelihood ratio test the same? Which test id more/most powerful (asymptotically)

To better illustrate the problem and our approach we consider the following

Example: (a tale of two empirical likelihoods)

Given $q > 1$ estimating equations

$$0 = \frac{1}{n} \sum_{i=1}^{n} g_j(X_i, \theta), \qquad j = 1, 2, \cdots, q$$

where $\theta \in R^q$ is the parameter.

If all $X_i, i = 1, 2, \cdots, n$ are iid and observable, the standard EL approach can be used here (see Owen 1991), with the empirical likelihood defined as

$$EL_1 = \prod_{i=1}^{n} p_i \ .$$

The null distribution is asymptotically chi squares with df=q as shown by Owen 1991.

However, when the X_i are subject to right censoring, and only $T_i = \min(X_i, C_i)$ and $\delta_i = I[X_i \le C_i]$ are available, the following two EL approaches were both used by many people.

(1) Since (see eg. Koul, Susarla and Van Ryzin 1981) for any $g_j(\cdot)$

$$\mathbb{E} \ \frac{\delta_i g_j(T_i, \theta)}{1 - G(T_i)} = \mathbb{E} \ g_j(X_i, \theta) \ ,$$

many people use the following estimating equations instead

$$0 = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i g_j(T_i, \theta)}{1 - G(T_i)} \ .$$

Notice this substituting estimating equation only involves the observable censored T_i, δ_i and the equations are consistent, in the sense that its mean is zero at the true parameter. This estimating equations can also be used with EL defined above: $EL_1 = \prod p_i$.

However, the censoring distribution $G(\cdot)$ is almost always unknown. This lead people to "plug-in" estimating equations, i.e. replace $G(\cdot)$ with a (root n) consistent estimator, in most case, the Kaplan-Meier estimator of G.

So, finally the estimating equations are

$$0 = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i g_j(T_i, \theta)}{1 - \hat{G}_{KM}(T_i)} \qquad j = 1, \cdots, q \tag{16}$$

and the EL remains $EL_1 = \prod_{i=1}^{n} p_i$.

The empirical likelihood ratio is then the ratio of two EL's, one with constraints derived from the estimating equation, one without. The maximum of EL_1 without constraints is $\prod 1/n$. The constraints used when maximize the EL_1 is

$$0 = \sum_{i=1}^{n} p_i \frac{\delta_i g_j(T_i, \theta)}{1 - \hat{G}_{KM}(T_i)} \ .$$

This leads to a log likelihood ratio statistic that under null hypothesis have a distribution as $l_1 \chi_1^2 + \cdots + l_p \chi_p^2$ asymptotically, where χ_j^2 are independent and each have a chi square with one degree of freedom. (see Hjort, McKeague and many others)

(2) The second approach. Start with the same 'plug-in' estimating equations (16) A key idea is to use the identity

$$\frac{\delta_i}{n[1 - \hat{G}_{KM}(T_i)]} = \Delta \hat{F}_{KM}(T_i)$$

so that the above estimating equation (16)can be written as

$$0 = \sum_{i=1}^{n} g_j(T_i, \theta) \Delta \hat{F}_{KM}(T_i) \qquad j = 1, \cdots, q. \tag{17}$$

This hinted that the EL to use here, should be the one that will make $\hat{F}_{KM}$ a NPMLE:

$$EL_2 = \prod_{i:\delta_i=1} p_i \prod_{i:\delta_i=0} (\sum_{T_j > T_i} p_j) \ .$$

Again, the empirical likelihood ratio is the ratio of two such EL_2's, one with the following constraints and one without. The constraint equation to use with this EL_2 when form the likelihood ratio is

$$0 = \sum_{i=1}^{n} g_j(T_i, \theta) \delta_i p_i \ .$$

We have, under the null hypothesis that

$$-2 \log ELR = \chi^2_{(q)} + o_p(1) \ ,$$

where $\chi^2_{(q)}$ is chi squares with df=q.

This result should be part of the results from part one of this proposal. Alternatively we may proof this directly, using approaches similar to Pan and Zhou.

Question: which approach is better?

Aside from the simplicity of the null distribution (chi squares with df=q), we argue that the second approach also have better power.

Develop and unify the theory,

Likely outcome is: identify more/most efficient EL test, and guideline for the future construction of EL for censored data.

1. The estimating equations EL, and MLE with over-determined case, more equations than the number of parameters i.e. Qin and Lawless (considered). But we do it for right censored data.

We first shall proof a similar result to Qin and Lawless for censored data, but with estimating equations in terms of hazard (not CDF).

Then we use the similar technique of 1 to show the estimating equations, over determined case, for CDF also holds for the censored data.

4 *Applications

(1) Economics. (2) quantile regression at several τ and suppose it has model with same β. (3) Rank regression, optimal weight, use both logrank and wilcoxon estimating equation for one sets of β. This achieves better efficiency.

2. Show the equivalency between the hypothesis about hazard integrals and the hypothesis of integrals wrt CDF. Equivalency is meant for the analysis, and asymptotic distribution results when empirical likelihood ratio test is used for test the hypothesis.

Since the proof for the EL test for hazard hypothesis are much easier and known (see Kim Bathke, and Zhou), this will lead to many results in the EL test on the CDF side, which are not well known, or the conditions are not explicitly worked out, or simply do not exist.

Multivariate/multiple hypothesis generalization of this is possible, leading to more new results in the EL for integrals of CDF.

For example we shall get that the EL for right censored data and hypothesis of integration wrt CDF is having a chi square limiting distribution under the minimal condition

$$\int \frac{[\phi(t)]^2}{1 - G(t-)} dF(t) < \infty$$