Censored Empirical Likelihood, and Hazard, II

Mai Zhou

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. (the second line assumes 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 max 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 first 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, 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 Nelson-Aalen: $\Lambda(t) \ll \hat{\Lambda}_{NA}(t)$. [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

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

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

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

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*

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

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

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 LLN and CLT for 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

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

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

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.

Define

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

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 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

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

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 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 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 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 $\sup_{1\leq i\leq n}$) bounded in probability (assumption 2, and property of empirical distribution function) we have, that the term is bounded in probability by

$$\sum_i I[T_i \geq \tau] \frac{g_n^2(T_i)}{R(T_i)/n} \frac{\Delta N(T_i)}{R(T_i)} \leq C \frac{1}{n} \sum_i I[T_i \geq \tau] \frac{g^2(T_i)\delta_i}{[1 - H(T_i-)]^2}$$

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.

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.

Appendix. **Verification of λ_0 is 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: 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.

References

Pan, X. and Zhou, M. (1999). Using 1-parameter sub-family of distributions in empirical likelihood ratio with censored data. *Journal of Statistical Planning and Inference* **75**, 379-392.

Pan, X. and Zhou, M. (2002). Empirical likelihood ratio in terms of cumulative hazard function for censored data. *Journal of Multivariate Analysis.* **80**, 166-188.