Notes on Censored EL, and Harzard

Mai Zhou

1 Empirical Likelihood Ratio Method in Terms of Cumulative Hazard function With Right-Censoring

The cumulative hazard function $\Lambda(t)$ is defined by

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

where $F(\cdot)$ is a CDF on the positive half line. Pan and Zhou (2002) studied the empirical likelihood ratio in terms of of hazard for right censored data subject to the constraint of the type $\int g(t) d\Lambda(t) = \theta$. Here g is a given function, (later we can see that g can be predictable). We outline a proof.

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. 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: $\Delta\hat{\Lambda}_{NA}(T_i) = \frac{\delta_i}{R_i}$.

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

$$\int_0^\infty g(t) d\Lambda(t) = \theta \tag{1}$$

where $g(t)$ is a given function satisfy some moment conditions, and θ is a given constant. The constraint (1) can be written as (for discrete hazard)

$$\sum_{i=1}^{n-1} \delta_i g(T_i) w_i + g(T_n)\delta_n = \theta \ . \tag{2}$$

Theorem 1 *If the constraint (2) is feasible (which means the maximum problem has a solution), then the maximum of $EL(\Lambda)$ under the constraint is obtained when*

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

where

$$Z_i = \frac{\delta_i g(T_i)}{R_i/n} \quad \textit{for} \ \ i = 1, 2, \ldots, n.$$

and λ is the solution of the following equation

$$\sum_{i=1}^{n-1} \frac{1}{n} \frac{Z_i}{1+\lambda Z_i} + g(T_n)\delta_n = \theta \ . \tag{3}$$

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

In the paper, it is also showed the following Wilks theorem.

Theorem 2 *Let* $(T_1, \delta_1), \ldots, (T_n, \delta_n)$ *be n pairs of random variables as defined above. Suppose g is a left continuous function and*

$$0 < \int \frac{|g(x)|^m}{(1-F_0(x))(1-G_0(x-))} d\Lambda(x) < \infty, \quad m = 1, 2.$$

Then, $\theta_0 = \int g(t) d\Lambda(t)$ *will be a feasible value with probability approaching one as* $n \to \infty$ *and*

$$-2 \log ELR(\theta_0) \xrightarrow{\mathcal{D}} \chi^2_{(1)} \quad as \quad n \to \infty$$

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

Proof. Here we briefly outline the proof. For the complete proof, see Pan and Zhou (2002). First, we proof the following two lemmas. They are the LLN and CLT for Nelson-Aalen estimator via counting processes technique.

Lemma 1 *Under the assumption of Theorem 2, we have*

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

where

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

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

$$\sqrt{n}\Big(\frac{1}{n} \sum_{i=1}^{n} Z_i - \theta_0\Big) = \sqrt{n}\Big(\sum_{i=1}^{n} g(T_i) \Delta \hat{\Lambda}_{NA}(T_i) - \theta_0\Big) \xrightarrow{\mathcal{D}} N(0, \sigma^2_\Lambda(g)).$$

where

$$\sigma^2_\Lambda(g) = \int \frac{g^2(x)}{(1-F_0(x))(1-G_0(x-))} d\Lambda_0(x) \quad and \quad \theta_0 = \int g(t) d\Lambda_0(t) \ .$$

Next, we show the solution of λ to the constraint equation (3) is

$$\lambda = \lambda^* = \frac{\frac{1}{n}\sum_{i=1}^{n} Z_i - \theta_0}{\frac{1}{n}\sum_{i=1}^{n-1} Z_i^2} + o_p(n^{-1/2}) \tag{4}$$

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

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

$$-2\log ELR = 2[f(0) - f(\lambda^*)] = 2[f(0) - f(0) - \lambda^* f'(0) - 1/2(\lambda^*)^2 f''(0) + \ldots].$$

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

$$-2\log ELR = -f''(0)(\lambda^*)^2 + \ldots \qquad (5)$$

simplify it to the following

$$-2\log ELR(\theta_0) = (\lambda^*)^2 \sum_{i=1}^{n-1} Z_i^2 + o_p(1) = \left(\frac{\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} Z_i - \theta_0)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n-1} Z_i^2}} \right)^2 + o_p(1)$$

Finally, by Lemma 1 and Lemma 2, we get

$$-2\log ELR(\theta_0) \xrightarrow{\mathcal{D}} \chi^2_{(1)} \quad as \quad n \to \infty \ .$$

We can easily get more accurate result by keeping more terms in the expansion of (4), and (5). It seems to me that we also need to get rate of the convergence in Lemma 1 and Lemma 2. Then the convergence rate of $-2\log ELR(\theta_0)$ to χ^2 can be obtained.

Remark: If g is a given function, it seems to me that g do not need to be even left continuous, it just need to be a measurable function and several integrals of it are well defined.

References

[1] Murphy, S (1995). Likelihood ratio based confidence intervals in survival analysis, *J. Amer. Statist. Assoc.* **90**, 1399-1405.

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

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

[4] Owen, A. (2001). Empirical likelihood. Chapman & Hall, London.

[5] 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.

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

[7] Zhou, M. (2005). Empirical likelihood ratio with arbitrarily censored/truncated data by EM algorithm. *J. of Computational and Graphical Statistics*,**14**, 643-656.