Notes on Censored EL, and Harzard

Mai Zhou

In survival analysis, the statistics involving the hazard functions are usually easier to handle mathematically then those involving the 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).

However, there is a catch: the hazard function Λ has two distinct pairs formula connecting with the distribution function F, one for the continuous hazard, one for the discrete hazard. 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 truly likelihood, the continuous version is an approximation.

Discrete

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

Continuous

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

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.

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 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 (specify 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 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*

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

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 Law of Large Numbers 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*

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

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.

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 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

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

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

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

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 $t \leq \tau$.

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 $\sup_{1\le i\le n}$) 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: 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 $0 = l(\eta)$.

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

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 sequence of local alternative hypothesis has a non-central chi squared distribution. (similar to Owen 1988).

The Poisson likelihood we defined in previous chapter has received some criticism. Since we assumed a discrete hazard/distribution function but at the same time we used a formula connecting the hazard and CDF that is only valid for the continuous case.

The discrepancy vanishes asymptotically but for finite samples, it is not an exact likelihood.

The 'binomial' likelihood we shall discuss here always strictly stick to a discrete CDF/hazard function, and the likelihood is a true probability. However, the class of statistic/parameter we shall be testing has a strange integrating format.

1 Censored Empirical Likelihood with $(k > 1)$ Constraints, Binomial likelihood

We will first study the one sample case. The results extend straightforwardly to the two sample situation in the next section.

1.1 One Sample Censored Empirical Likelihood

For n independent, identically distributed observations, $X_1, \cdots, X_n$, assume that the distribution of the X_i is $F_{x0}(t)$, and the cumulative hazard function of X_i is $\Lambda_{x0}(t)$. With right censoring, we only observe

$$T_i = \min(X_i, C_i) \qquad \text{and} \qquad \delta_i = I_{[X_i \le C_i]} \tag{1}$$

where the C_i's are the censoring times, assumed to be independent, identically distributed, and independent of the X_i's. Based on the censored observations, the log empirical likelihood pertaining to a distribution F_x is

$$\log EL(F_x) = \sum_i [\delta_i \log \Delta F_x(T_i) + (1-\delta_i)\log\{1 - F_x(T_i)\}] \ . \tag{2}$$

As shown in Pan and Zhou (2002), computations are much easier with the empirical likelihood reformulated in terms of the corresponding (cumulative) hazard function. However, there are different formula relating the CDF and the cumulative hazard function for discrete or continuous cases. Since the maximization of the EL will force the distribution to be discrete (for example: empirical distribution or the Kaplan-Meier) we shall use the discrete formula relating the F to Λ. The equivalent hazard formulation of (2) will be denoted by $\log EL(\Lambda_x)$. Using the relations

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

we can rewrite the empirical likelihood (Proof as homework) as follows. Denoting $\Delta\Lambda(T_i) = v_i$ the EL is given as follows:

$$\log EL(\Lambda_x) = \sum_i \{d_i \log v_i + (R_i - d_i)\log(1 - v_i)\} \tag{3}$$

where $d_i = \sum_{j=1}^n I_{[T_j = t_i]}\delta_j$, $R_i = \sum_{j=1}^n I_{[T_j \geq t_i]}$, and t_i are the ordered, distinct values of T_i. This EL is called the binomial version of the hazard empirical likelihood. See, for example, Thomas and Grunkemeier (1975) and Li (1995) for similar notation. Here, $0 < v_i \leq 1$ are the discrete hazards at t_i. The maximization of (3) with respect to v_i is known to be attained at the jumps of the Nelson-Aalen estimator: $v_i = d_i/R_i$. We further denote the maximum value achieved by EL as $EL(\hat{\Lambda}_{NA})$. Notice the similarity of this likelihood to the likelihood of a binomial sample, hence the name.

Let us consider a hypothesis testing problem for a k dimensional parameter $\theta = (\theta_1, \cdots, \theta_k)^T$ with $\theta_r = \int g_r(t)\log(1 - d\Lambda_x(t))$, where the $g_r(t)$ are given nonnegative functions. See also remark 1 after the theorem for the strange looking integration.

$$H_0 : \theta = \mu \quad \text{vs.} \quad H_A : \theta \neq \mu$$

where $\mu = (\mu_1, \cdots, \mu_k)^T$ is a vector of k constants. The constraints we shall impose on the discrete hazards v_i are: for given functions $g_1(\cdot), \cdots, g_k(\cdot)$ and constants $\mu_1, \cdots, \mu_k$, we have

$$\sum_i^{N-1} g_1(t_i)\log(1 - v_i) = \mu_1 \ , \quad \cdots \quad , \ \sum_i^{N-1} g_k(t_i)\log(1 - v_i) = \mu_k \ , \tag{4}$$

where N is the total number of distinct observation values. We need to exclude the last value as we always have $v_N = 1$ for discrete hazards. Let us abbreviate the maximum likelihood estimators of $\Delta\Lambda_x(t_i)$ under constraints (4) as v_i^*. Application of the Lagrange multiplier method shows

$$v_i^* = v_i(\lambda) = \frac{d_i}{R_i + n\lambda^T G(t_i)} \ ,$$

where $G(t_i) = \{g_1(t_i), \cdots, g_k(t_i)\}^T$, and λ is the solution to (4) when replace v_i by $v_i(\lambda)$ (Lemma 1 in the appendix).

Then, the likelihood ratio test statistic in terms of hazards is given by

$$W_2 = -2\{\log\max EL(\Lambda_x)(\text{with constraint (4)}) - \log EL(\hat{\Lambda}_{NA})\} \ .$$

We have the following result that is a version of Wilks' theorem for W_2 under some regularity conditions which include the standard conditions on censoring that allow the Nelson-Aalen estimators to have an asymptotic normal distribution (see, e.g., Gill, 1983; Andersen *et al.*, 1993). The proof of the following theorem, along with a detailed set of conditions, is provided in the appendix.

Theorem 1. *Suppose that the null hypothesis H_0 holds, i.e. $\mu_r = \int g_r(t) \log\{1 - d\Lambda_x(t)\}$, $r = 1, \ldots, k$. Then, under conditions specified in the appendix, the test statistic W_2 has asymptotically a chi-squared distribution with k degrees of freedom.*

Remark 1 The integration constraints are originally given as $\theta_r = \int g_r(t) d\log\{1 - F_x(t)\}$, $r = 1, \cdots, k$. (but this is not in terms of the hazard). The above formulation is found by using the identity $d\log\{1 - F(t)\} = \log\{1 - d\Lambda(t)\}$ which holds for **both** continuous and discrete $F(t)$. Again, the point t where $F(t) = 1$ have to be excluded from the integration.

Remark 2: If the functions $g_r(t)$ are random but predictable with respect to the filtration $\mathcal{F}_t = \sigma\{T_i I_{[T_i \le t]};\ \delta_i I_{[T_i \le t]};\ i = 1, \ldots, n\}$, then Theorem 1 is still valid (see the appendix for details).

Remark 3: One of the conditions for Theorem 1 is that the matrix Σ defined in Lemma 2 in appendix is invertible. If Σ is not invertible, then the k constraints may have redundancy within, in which case we may handle it by using the theory of over-determined EL.

1.2 Two Sample Censored Empirical Likelihood

Suppose that in addition to the censored sample of X-observations, we have a second sample $Y_1, \cdots, Y_m$ coming from a distribution function $F_{y0}(t)$ with a cumulative hazard function $\Lambda_{y0}(t)$. Assume that the Y_j's are independent of the X_i's. With censoring, we can only observe

$$U_j = \min(Y_j, S_j) \qquad \text{and} \qquad \tau_j = I_{[Y_j \le S_j]} \tag{5}$$

where the S_j's are the censoring variables for the second sample. Denote the ordered, distinct values of the U_j by s_j.

Similar to (3), the log empirical likelihood function based on the two censored samples pertaining to cumulative hazard functions Λ_x and Λ_y is simply $\log EL(\Lambda_x, \Lambda_y) = L_1 + L_2$ where

$$L_1 = \sum_i d_{1i} \log v_i + \sum_i (R_{1i} - d_{1i}) \log(1 - v_i) \quad \text{and}$$
$$L_2 = \sum_j d_{2j} \log w_j + \sum_j (R_{2j} - d_{2j}) \log(1 - w_j), \tag{6}$$

with d_{1i}, R_{1i}, d_{2j} and R_{2j} defined analogous to the one sample situation (see p. 3). Accordingly, let us consider a hypothesis testing problem for a k dimensional parameter $\theta = (\theta_1, \cdots, \theta_k)^T$ with respect to the cumulative hazard functions Λ_x and Λ_y such that

$$H_0 : \theta = \mu \quad \text{vs.} \quad H_A : \theta \neq \mu,$$

where $\theta_r = \int g_{1r}(t) \log\{1 - d\Lambda_x(t)\} - \int g_{2r}(t) \log\{1 - d\Lambda_y(t)\}$, $r = 1, \cdots, k$, for some predictable functions $g_{1r}(t)$ and $g_{2r}(t)$. Then, the constraints imposed on v_i and w_j are

$$\mu_r = \sum_{i=1}^{N-1} g_{1r}(t_i) \log(1 - v_i) - \sum_{j=1}^{M-1} g_{2r}(s_j) \log(1 - w_j), \quad r = 1, \ldots, k, \tag{7}$$

where N and M are the total numbers of distinct observation values in the two samples. As in the one sample case, we need to exclude the last value in each sample.

Let us abbreviate the maximum likelihood estimators of $\Delta\Lambda_x(t_i)$ and $\Delta\Lambda_y(s_j)$ under the constraints (7) as v_i^* and w_j^*, respectively, where $i = 1, \cdots, N$ and $j = 1, \cdots, M$. Application of the Lagrange multiplier method shows

$$v_i^* = v_i(\lambda) = \frac{d_{1i}}{R_{1i} + \min(n,m)\lambda^T G_1(t_i)} \ , \qquad w_j^* = w_j(\lambda) = \frac{d_{2j}}{R_{2j} - \min(n,m)\lambda^T G_2(s_j)} \ ,$$

where $G_1(t_i) = \{g_{11}(t_i), \cdots, g_{1k}(t_i)\}^T$, $G_2(s_j) = \{g_{21}(s_j), \cdots, g_{2k}(s_j)\}^T$, and λ is the solution to (7) when we plug in the v_i^* and w_j^*. Then, the two-sample test statistic is given as follows:

$$W_2^* = -2\{\log \max EL(\Lambda_x, \Lambda_y)(\text{with constraint (7)}) - \log EL(\hat{\Lambda}_x^{NA}, \hat{\Lambda}_y^{NA})\} \ ,$$

analogous to the one-sample case. The following theorem provides the asymptotic distribution result for W_2^*. The proof can be found in the appendix.

Theorem 2. *Suppose that the null hypothesis $H_0 : \ \theta_r = \mu_r$ holds. i.e. $\mu_r = \int g_{1r}(t) \log\{1 - d\Lambda_x(t)\} - \int g_{2r}(t) \log\{1 - d\Lambda_y(t)\}$, $r = 1, \ldots, k$. Then, as $\min(n, m) \to \infty$ and $n/m \to c \in (0, \infty)$, W_2^* has asymptotically a chi-squared distribution with k degrees of freedom.*

Remark: The two-sample setup we studies in this section took a particularly simple form: the difference of two parameters. For more involved parameters, we may not be able to write it as a simple difference. For example a two sample U statistics: $\theta = \int \int g(s,t) d\Lambda_x(s) d\Lambda_y(t)$. For the analysis of those, please see Barton (2010) and the R package `emplik2`.

A Appendix

Assumptions for Theorem 1

Let $X_1, \cdots, X_n$ be independent, identically distributed random variables with cumulative distribution function $F_{x0}(t)$ and cumulative hazard function $\Lambda_{x0}(t)$. We observe $T_i = \min(X_i, C_i)$ and $\delta_i = I_{[X_i \le C_i]}$, where the C_i are independent, identically distributed censoring times, independent of the X_i. The cumulative distribution function of the C_i is $F_c(t)$. The distribution functions $F_{x0}(t)$ and $F_c(t)$ do not have common discontinuities.

Let $g_1(t), \ldots, g_k(t)$ be non-negative left continuous functions with

$$0 < \int \frac{|g_r(t)|^2(1 - \Delta\Lambda_{x0}(t))}{(1 - F_{x0}(t))(1 - F_c(t))} d\Lambda_{x0}(t) < \infty, \quad r = 1, \ldots, k. \tag{9}$$

This condition guarantees asymptotic normality of the Nelson-Aalen estimator (*cf.* Theorem 2.1 in Gill, 1983). Note that the factor $(1 - \Delta\Lambda_{x0}(t))$ is only needed for discrete distributions. It equals 1 when F_{x0} is absolutely continuous. Also, under the above condition, $\mu_r = \int g_r(t) \log(1 - d\Lambda_{x0}(t))$ is feasible with probability approaching 1 as $n \to \infty$. Note that the functions $g_r(t)$ may be random, but they have to be predictable with respect to the filtration $\mathcal{F}_t = \sigma\{T_i I_{[T_i \le t]};\ \delta_i I_{[T_i \le t]};\ i = 1, \ldots, n\}$ which makes $\hat{\Lambda}_{NA}(t) - \Lambda_{x0}(t)$ a martingale, so that the martingale central limit theorem can be applied. Here, $\hat{\Lambda}_{NA}(t)$ denotes the Nelson-Aalen estimator of hazard function. Furthermore, if the functions $g_r(t)$ are random, we require that there are non-random left continuous functions $g_{r0}(t)$ such that $\sup\limits_{t \le T_n} |g_r(t) - g_{r0}(t)| = o_p(1)$ and $\sup\limits_{t \le T_n} |\frac{g_r(t)}{g_{r0}(t)}| = O_p(1)$ for $r = 1, \ldots, k$ as $n \to \infty$.

Remark: The strength of the above assumption (9) is quite weak. This assumption is apparently of the same strength as the condition Akritas (2000) put on the mean function. Akritas was considering the CLT for $\int \phi(t) dF(t)$ and requires $\int \phi^2(t) dF(t)/[1 - G(t-)] < \infty$. If we put $g(t) = [1 - F(t)]\phi(t)$, then our condition (9) above is the same as Akritas condition. Why $g(t) = [1 - F(t)]\phi(t)$ is the connection? See Akritas and some more discussions in tech report.

Mathematical Derivations and Proofs for Theorem 1

Recall the column vectors $G(t) = \{g_1(t), \cdots, g_k(t)\}^T$ and $\lambda = \{\lambda_1, \cdots, \lambda_k\}^T$.

Lemma 1. *The hazards that maximize the log likelihood function (3) under the constraints (4) are given by*

$$v_i(\lambda) = \frac{d_i}{R_i + n\lambda^T G(t_i)} \ , \tag{10}$$

where λ is obtained as the solution of the following k equations.

$$\sum_i^{N-1} g_1(t_i) \log\{1 - v_i(\lambda)\} = \mu_1 \ , \quad \cdots \quad , \ \sum_i^{N-1} g_k(t_i) \log\{1 - v_i(\lambda)\} = \mu_k \ . \tag{11}$$

Proof of Lemma 1. The result follows from a standard Lagrange multiplier argument applied to (3) and (4). See Fang and Zhou (2000) for some similar calculations. $\diamondsuit$.

We denote the solution of (11) by λ_x.

Lemma 2. *Assume the data are such that the Nelson-Aalen estimator is asymptotically normal and the variance-covariance matrix Σ defined below (p. 12) is invertible. Then, for the solution λ_x of the constrained problem (11), corresponding to the null hypothesis $H_0: \ \mu_r = \int g_r(t) \log\{1 - d\Lambda_{x0}(t)\}, \ r = 1, \ldots, k$, we have that $n^{1/2}\lambda_x$ converges in distribution to $N(0, \Sigma)$.*

Preparation for the proofs of Lemma 2 and Theorem 1.

Let

$$f(\lambda) = \sum [d_i \log v_i(\lambda) + (R_i - d_i) \log\{1 - v_i(\lambda)\}] \ . \tag{12}$$

In order to show that $f'(0) = 0$, we compute

$$\frac{\partial}{\partial \lambda_r} f(\lambda) = \sum_i \frac{d_i}{v_i(\lambda)} \frac{\partial v_i(\lambda)}{\partial \lambda_r} - \frac{(R_i - d_i)}{v_i(\lambda)} \frac{\partial(1 - v_i(\lambda))}{\partial \lambda_r}, \ r = 1, \ldots, k.$$

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

$$\frac{\partial}{\partial \lambda_r} f(\lambda)|_{\lambda=0} = -\sum_i (R_i - R_i) \frac{d_i n g_r(t_i)}{R_i^2} \equiv 0 \ .$$

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

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

After straightforward but tedious calculations, we obtain

$$D_{rl} = -\left\{ \sum_i \frac{n^2 g_r g_l}{R_i} \frac{d_i}{R_i - d_i} \right\} \ .$$

By a now standard counting process martingale argument, we see that $-D_{rl}/n$ converges almost surely to D^*_{rl}.

Proof of Lemma 2. We derive the asymptotic distribution of λ. The argument is similar to, for example, Owen (1990) and Pan and Zhou (2002). Define a vector function $h(s) = \{h_1(s), \cdots, h_k(s)\}^T$ by

$$h_1(s) = \sum_i g_1(t_i)\log\{1 - v_i(s)\} - \mu_1 \ , \cdots , \ h_k(s) = \sum_i g_k(t_i)\log\{1 - v_i(s)\} - \mu_k \ . \tag{13}$$

Then, λ is the solution of $h(s) = 0$. Thus, we have

$$0 = h(\lambda) = h(0) + h'(0)\lambda + o_p(n^{-1/2}) \ , \tag{14}$$

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

Indeed, if we write $\lambda = \rho \cdot \tilde{\lambda}$, where $\|\tilde{\lambda}\| = 1$, then

$$\begin{aligned}
0 = \tilde{\lambda}^T h(\lambda) &= \sum_i \tilde{\lambda}^T G(t_i)\log\{1 - v_i(s)\} - \tilde{\lambda}^T\mu = \sum_i \tilde{\lambda}^T G(t_i)\log\{1 - \frac{d_i}{R_i + n\lambda^T G(t_i)}\} - \tilde{\lambda}^T\mu \\
&= \Big(\sum_i \tilde{\lambda}^T G(t_i)\log(1 - \frac{d_i}{R_i}) - \tilde{\lambda}^T\mu\Big) + \sum_i \tilde{\lambda}^T G(t_i)\log\Big[\frac{1 - d_i/\{R_i + n\lambda^T G(t_i)\}}{1 - d_i/R_i}\Big] \\
&= A + B \ ,
\end{aligned}$$

where the first expression A is of order $O_p(n^{-1/2})$. Considering the second expression, and noting that for any pair of numbers $\varepsilon_1, \varepsilon_2 \in (0, 1]$, the inequality $|\varepsilon_1 - \varepsilon_2| \le |\log(\varepsilon_1) - \log(\varepsilon_2)|$ holds, we have

$$\begin{aligned}
|B| &= |\sum_i \tilde{\lambda}^T G(t_i)\log\Big[\frac{1 - d_i/\{R_i + n\lambda^T G(t_i)\}}{1 - d_i/R_i}\Big]| \\
&\ge |\sum_i \tilde{\lambda}^T G(t_i)\frac{n\rho G(t_i)^T\tilde{\lambda} d_i}{R_i(R_i + n\rho\tilde{\lambda}^T G(t_i))}| \\
&\ge \frac{|\rho|}{1 + n|\rho|\max_i |\tilde{\lambda}^T G(t_i)/R_i|}\sum_i \frac{(\tilde{\lambda}^T G(t_i))^2 n d_i}{R_i^2}
\end{aligned}$$

The sum in the last expression is of order $O_p(1)$, and under assumption (9), the maximum in the denominator is of order $o_p(n^{1/2})$. Therefore, $|\rho|$ is of order $O_p(n^{-1/2})$, and hence, the expansion (14) is valid.

Therefore,

$$n^{1/2}\lambda = \{h'(0)\}^{-1}\{-n^{1/2}h(0)\} + o_p(1) \ .$$

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

$$h'_{rl} = \sum_i \frac{n g_r g_l d_i}{R_i(R_i - d_i)} \ .$$

Notice that we have verified $nh'_{rl} = -D_{rl}$. By the counting process martingale central limit theorem (see, for example, Gill, 1980; Andersen *et al.*, 1993; or Fang and Zhou, 2000), we can show that $n^{1/2}h(0)$ converges in distribution to $N(0, \Sigma_h)$ with $\Sigma_h = \lim h'(0)$.

Finally, putting it together, we have that $n^{1/2}\lambda(0) = \{h'(0)\}^{-1}\{-n^{1/2}h(0)\} + o_p(1)$ converges in distribution to $N(0, \Sigma)$ with $\Sigma = \lim\{h'(0)\}^{-1}$. Recalling $nh'_{rl} = -D_{rl}$, we see that $\Sigma^{-1} = D^*$. $\diamondsuit$

Proof of Theorem 1. Let $f(\lambda)$ be defined as in (12). Then, we have $W_2 = -2\{f(\lambda_x) - f(0)\}$. By Taylor expansion, we obtain

$$W_2 = 2\{f(0) - f(0) - f'(0)\lambda_x - \frac{1}{2}\lambda_x^T D \lambda_x + o_p(1)\}, \tag{15}$$

where we use D to denote the matrix of second derivatives of $f(\cdot)$ with respect to λ. The expansion is valid in view of Lemma 2 (λ_x is close to zero).

Since we have $f'(0) = 0$ (see above), the expression above is reduced to

$$W_2 = -\lambda_x^T D \lambda_x + o_p(1) \ . \tag{16}$$

Notice that $-D$ is symmetric and positive definite for large enough n because $-D/n$ converges to a positive definite matrix, see below. Therefore, we may write

$$W_2 = \lambda_x^T (-D)^{1/2}(-D)^{1/2}\lambda_x + o_p(1) \ . \tag{17}$$

Recalling the distributional result for λ_x in Lemma 2 and noticing that $-D/n$ converges almost surely to D^*, and $D^* = \Sigma^{-1}$ (see above in the proof of Lemma 2), it is not hard to show that $n^{1/2}\lambda_x^T(D^{1/2}n^{-1/2})$ converges in distribution to $N(0, I)$. This together with (16) implies that W_2 converges in distribution to χ^2_k . $\diamondsuit$

About feasibility: Clearly when $\lambda = 0$ all the v_i's are between 0 and 1 or equivalently, μ_{NPMLE} is feasible. For the constraint imposed by a true H_0, as show above we have the order $\lambda_x = O_p(n^{-1/2})$. This imply $n\lambda^T G = O_p(n^{1/2})$. Notice $R(t) = O_p(n)$, so as $n \to \infty$ we always have $0 < d_i/(R + n\lambda G) < 1$, or that a true null hypothesis is feasible.

Assumptions for Theorem 2

Let $X_1, \cdots, X_n$ be independent, identically distributed random variables with cumulative distribution function $F_{x0}(t)$ and cumulative hazard function $\Lambda_{x0}(t)$. We observe $T_i = \min(X_i, C_i)$ and $\delta_i = I_{[X_i \le C_i]}$, where the C_i are independent, identically distributed censoring times, independent of the X_i. The cumulative distribution function of the C_i is $F_c(t)$. The distribution functions $F_{x0}(t)$ and $F_c(t)$ do not have common discontinuities. Further, let $Y_1, \cdots, Y_m$ be independent, identically distributed random variables with cumulative distribution function $F_{y0}(t)$ and cumulative hazard function $\Lambda_{y0}(t)$. We observe $U_j = \min(Y_j, S_j)$ and $\tau_j = I_{[Y_j \le S_j]}$, where the S_j are independent, identically distributed censoring times, independent of the Y_j. The cumulative distribution function of the S_j is $F_s(t)$. The distribution functions $F_{y0}(t)$ and $F_s(t)$ do not have common discontinuities. The (Y_j, S_j) are independent of the (X_i, C_i).

Let $g_{1r}(t)$ and $g_{2r}(t)$, $r = 1, \ldots, k$, be non-negative left continuous functions with

$$0 < \int \frac{|g_{1r}(t)|^2 (1 - \Delta\Lambda_{x0}(t))}{(1 - F_{x0}(t))(1 - F_c(t))} d\Lambda_{x0}(t) < \infty, \quad r = 1, \ldots, k, \qquad \text{and}$$
$$0 < \int \frac{|g_{2r}(t)|^2 (1 - \Delta\Lambda_{y0}(t))}{(1 - F_{y0}(t))(1 - F_s(t))} d\Lambda_{y0}(t) < \infty, \quad r = 1, \ldots, k.$$

The functions $g_{lr}(t)$, $l = 1, 2$, $r = 1, \ldots, k$, may be random, but they have to be predictable with respect to the filtration $\mathcal{F}_t = \sigma\{T_i I_{[T_i \le t]};\ \delta_i I_{[T_i \le t]};\ U_j I_{[U_j \le t]};\ \tau_j I_{[U_j \le t]};\ i = 1, \ldots, n;\ j = 1, \ldots, m\}$. Furthermore, if the functions $g_{lr}(t)$ are random, we require that there are non-random left continuous functions $g_{lr0}(t)$ such that $\sup_{t \le V_n} |g_{lr}(t) - g_{lr0}(t)| = o_p(1)$ and $\sup_{t \le V_n} |\frac{g_{lr}(t)}{g_{lr0}(t)}| = O_p(1)$ for $r = 1, \ldots, k$ as $\min(m, n) \to \infty$. Here $V_n = \min(\max T_i, \max U_j)$.

Mathematical Derivations and Proofs for Theorem 2

The proof of Theorem 2 is very similar to the one for the one-sample situation. In the two-sample case, the constraints are defined by

$$\mu_r = \sum_{i=1}^{N-1} g_{1r}(t_i) \log(1 - v_i) - \sum_{j=1}^{M-1} g_{2r}(s_j) \log(1 - w_j), \quad r = 1, \ldots, k.$$

Define $G_1(t_i) = \{g_{11}(t_i), \cdots, g_{1k}(t_i)\}^T$ and $G_2(s_j) = \{g_{21}(s_j), \cdots, g_{2k}(s_j)\}^T$. The vector λ_{xy} is the solution to maximizing $\log EL(\Lambda_x, \Lambda_y) = L_1 + L_2$ under the above constraints. Similar to Lemma 1, application of the Lagrange multiplier method yields the maximum likelihood estimators

$$v_i(\lambda) = \frac{d_{1i}}{R_{1i} + \min(n, m)\lambda_{xy}^T G_1(t_i)} \quad \text{and} \quad w_j(\lambda_{xy}) = \frac{d_{2j}}{R_{2j} - \min(n, m)\lambda_{xy}^T G_2(s_j)}\ .$$

In the two-sample situation, the function $f(\lambda)$ defined in (12) becomes

$$f(\lambda) = \sum [d_{1i} \log v_i(\lambda) + (R_{1i} - d_{1i}) \log\{1 - v_i(\lambda)\}] + \sum [d_{2j} \log w_j(\lambda) + (R_{2j} - d_{2j}) \log\{1 - w_j(\lambda)\}] \ .$$

The same calculation as above (see p. 10) yields $f'(0) = 0$ and $f''(0) = \sum$ where the rl^{th} element of the $k \times k$ matrix $\sum$ is

$$D_{rl} = -\left\{ \sum_i \frac{n^2 g_{1r} g_{1l}}{R_{1i}} \frac{d_{1i}}{R_{1i} - d_{1i}} + \sum_j \frac{m^2 g_{2r} g_{2l}}{R_{2j}} \frac{d_{2j}}{R_{2j} - d_{2j}} \right\} \ .$$

Since we assume that $n/m \to c \in (0, \infty)$ as $\min(m, n) \to \infty$, we have again that $-D_{rl}/n$ converges almost surely to D^{**}_{rl}.

In order to show the asymptotic normality of $n^{1/2}\lambda_{xy}$, we proceed analogous to the proof of Lemma 2. Define $h(u) = \{h_1(u), \cdots, h_k(u)\}^T$, where

$$h_r(u) = \sum_i g_{1r}(t_i) \log\{1 - v_i(u)\} - \sum_j g_{2r}(s_j) \log\{1 - w_j(u)\} - \mu_r \ , \ r = 1, \ldots, k,$$

let $\lambda_{xy} = \rho \cdot \tilde{\lambda}$, where $\|\tilde{\lambda}\| = 1$, and notice that

$$\begin{aligned} 0 &= \tilde{\lambda}^T h(\lambda) = A + B, \\ \text{where} \quad A &= \sum_i \tilde{\lambda}^T G_1(t_i) \log\{1 - \frac{d_{1i}}{R_{1i}}\} - \sum_j \tilde{\lambda}^T G_2(s_j) \log\{1 - \frac{d_{2j}}{R_{2j}}\} - \tilde{\lambda}^T \mu = O_p(n^{-1/2}) \\ \text{and} \quad B &= \sum_i \tilde{\lambda}^T G_1(t_i) \log \Big[\frac{1 - d_{1i}/\{R_{1i} + \min(m,n)\rho\tilde{\lambda}^T G_1(t_i)\}}{1 - d_{1i}/R_{1i}} \Big] \\ &\quad - \sum_j \tilde{\lambda}^T G_2(s_j) \log \Big[\frac{1 - d_{2j}/\{R_{2j} + \min(m,n)\rho\tilde{\lambda}^T G_2(s_j)\}}{1 - d_{2j}/R_{2j}} \Big] \ . \end{aligned}$$

A similar calculation as in the proof of Lemma 2 yields

$$|B| \geq \frac{|\rho|}{1 + n|\rho| \cdot |\max\big(\max_i(\tilde{\lambda}^T G(t_i)/R_{1i}), \max_j(\tilde{\lambda}^T G(s_j)/R_{2j})\big)|} \times \Big(\sum_i \frac{(\tilde{\lambda}^T G(t_i))^2 n d_{1i}}{R^2_{1i}} + \sum_j \frac{(\tilde{\lambda}^T G(s_j))^2 n d_{2j}}{R^2_{2j}} \Big) \ .$$

Again, the sum in the last expression is of order $O_p(1)$, and $|\rho|$ is therefore of order $O_p(n^{-1/2})$. Thus, the expansion $0 = h(\lambda_{xy}) = h(0) + h'(0)\lambda_{xy} + o_p(n^{-1/2})$ is valid, where $h'(0)$ is a $k \times k$

matrix. Application of the counting process martingale central limit theorem shows that $n^{1/2}\lambda_{xy}$ converges to $N(0, \Sigma)$ with $\Sigma = \lim\{h'(0)\}^{-1}$.

The final step in the proof of Theorem 2 is a Taylor expansion of $W_2^* = -2\big(f(\lambda_{xy}) - f(0)\big)$ as

$$W_2^* = -2\big(f'(0)\lambda_{xy} + \frac{1}{2}\lambda_{xy}^T D \lambda_{xy}\big) + o_p(1) = \lambda_{xy}^T(-D)^{1/2}(-D)^{1/2}\lambda_{xy} + o_p(1) \ ,$$

and noticing that $\lambda_{xy}^T(-D)^{1/2}$ converges in distribution to $N(0, I)$.