The Poisson likelihood we defined in previous section 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.

This do not matter 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 we shall be testing has a bit 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 functions. Since the maximization of the El will force the distribution to be discrete (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). 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 $EL(\hat{\Lambda}_{NA})$.

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.

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

where $\mu = (\mu_1, \cdots, \mu_k)^T$ is a vector of k constants. We note that the θ_r are linear functionals of the cumulative hazard function. The constraints we shall impose on the 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 interms 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 $F(t)$.

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. 2). 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 maximizing $\log EL(\Lambda_x, \Lambda_y) = L_1 + L_2$ under the constraints in (7). Then, the two-sample test statistic is given as follows:

$$W_2^* = -2\{\log \max EL(\Lambda_x, \Lambda_y)(\text{with constraint (7)}) - \log \max EL(\Lambda_x, \Lambda_y)(\text{without constraint}) \ ,$$

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.*

2 Examples and Simulations

We provide Monte Carlo simulation results that empirically confirm the chi-squared limiting distributions of W_2 and W_2^*. Simulation study 3 compares the small and moderate sample size behaviors of the proposed combined test with other existing procedures for a hypothesis test of two-sample survival data. Two real data examples are provided to illustrate the proposed method for combining the log-rank and Gehan-Wilcoxon tests for one and two samples. We present R code for the real data Example 2 in the appendix. All the computations have been carried out using version 0.9-1 of the 'emplik' package in R.

Simulation 1

This simulation study examines the distribution of W_2 when the constraints are non-random functions such that $g_1(t) = \exp(-t)$, $g_2(t) = \frac{1}{2}t \cdot I_{[t \leq 1]}$, and $g_3(t) = I_{[t \leq 0.9]}$. We use the following distributions to generate the random variables.

$$X \sim \exp(1) \ , \qquad C \sim \exp(0.5), \tag{8}$$

and the censored observations are created via (1). The Q-Q plot (Figure 1) is based on 5,000 runs. The distribution of W_2 agrees well with the theoretically derived χ^2_3-distribution.

Simulation 2

This simulation study examines the distribution of W_2^* where the constraints are random functions. We choose random functions corresponding to the test statistics of the log-rank and Gehan-Wilcoxon tests and obtain the value of W_2^* as described in Section **??**. We also examine the size of the such combined test of the log-rank and Gehan-Wilcoxon tests. In each of 10,000 runs, two identically distributed equal sized random samples are generated from the simulation setup in (8). Figure **??** confirms that the distribution of W_2^* agrees well with χ^2_2. The distribution deviates in the tail area when the sample sizes are $n = m = 30$, but the deviation is in the extreme end of the tail, above the 99th percentile (the higher horizontal line) of the theoretical χ^2-distribution. Results in Table 1 agree: the proposed combined test attains the type I error at the nominal levels.

Example 1. Iowa Psychiatric Patient Data

We apply the combined test of the log-rank and Gehan-Wilcoxon tests to a sample of survival times of 26 psychiatric inpatients to compare with the survival time distribution of the general population in Iowa. The data is part of a larger study of psychiatric inpatients admitted to the University of Iowa hospital during the years 1935-1948 (for more information on the data, see Tsuang and Woolson, 1977). Klein and Moeschberger (1997, p. 189) use the data to illustrate the one-sample log-rank test. The test statistics of the log-rank and Gehan-Wilcoxon tests are adjusted to accommodate the delayed entries and used as $g_r(t)$, $r = 1, 2$, as in (**??**). When applied individually, the log-rank and Gehan-Wilcoxon tests both reject the null with p-values < 0.001 and 0.0432. The combined test statistic reaches the same conclusion with the p-value 0.00088.

Example 2. Kidney Dialysis Patient Data

We apply the combined test of the log-rank and Gehan-Wilcoxon tests to re-analyze the kidney dialysis data of Klein and Moeschberger (1997, p. 197). The test statistics of the log-rank and Gehan-Wilcoxon tests are used as $g_{11} = g_{21}$ and $g_{12} = g_{22}$ as in (**??**). Out of a total of 119 patients, 43 had a catheter surgically placed and 76 percutaneously (for a detailed description of the data, see Nahman *et al.*, 1992). The plot of the estimated survival functions (Figure **??**) shows that the curves cross each other at about 6 months and suggests that the survival experience of the two groups is different. However, as indicated in the introduction, the log-rank test and its weighted versions make different decisions. Both the log-rank and Gehan-Wilcoxon tests, two of the most popular ones, fail to reject the null hypothesis with p-values 0.1032 and 0.9603 respectively, while tests of the $G^{\rho,\gamma}$ family with emphasis on the later time period reject the null. Electing to apply such a $G^{\rho,\gamma}$ family class test, though, is usually a post hoc decision. When our proposed method of combining the tests is applied, it rejects the null with a p-value of 0.0013. This indicates that the combined test can be much more powerful than either one of the individual tests. R code for this example is included in the appendix.

A Appendix

Appendix 1

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

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

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 $\sup_{t \le T_n} |g_r(t) - g_{r0}(t)| = o_p(1)$ and $\sup_{t \le T_n} |\frac{g_r(t)}{g_{r0}(t)}| = O_p(1)$ for $r = 1, \ldots, k$ as $n \to \infty$.

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. 9) 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

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

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

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

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

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

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

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

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. 8) 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

$$0 = \tilde\lambda^T h(\lambda) = A + B,$$

$$\text{where} \quad A = \sum_i \tilde\lambda^T G_1(t_i) \log\{1 - \frac{d_{1i}}{R_{1i}}\} - \sum_j \tilde\lambda^T G_2(s_j) \log\{1 - \frac{d_{2j}}{R_{2j}}\} - \tilde\lambda^T \mu = O_p(n^{-1/2})$$

$$\text{and} \quad B = \sum_i \tilde\lambda^T G_1(t_i) \log\left[\frac{1 - d_{1i}/\{R_{1i} + \min(m,n)\rho\tilde\lambda^T G_1(t_i)\}}{1 - d_{1i}/R_{1i}}\right] - \sum_j \tilde\lambda^T G_2(s_j) \log\left[\frac{1 - d_{2j}/\{R_{2j} + \min(m,n)\rho\tilde\lambda^T G_2(s_j)\}}{1 - d_{2j}/R_{2j}}\right] .$$

A similar calculation as in the proof of Lemma 2 yields

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

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 $N(0, I)$.

Appendix 2

R Code for Example 2

```
## R code for the two sample combined test of log-rank and Gehan-Wilcoxon test
## with Kidney Dialysis Patient Data from Klein and Moeschberger (1997, p.197).

## load the library and data ##
> library(KMsurv)
> data(kidney)
> library(emplik)
```

```
> names(kidney)
[1] "time"  "delta" "type"
> sum(kidney[,3]==1)
[1] 43
> sum(kidney[,3]==2)
[1] 76

### define weight functions fR1(t) and fR2(t) ###
### these functions count the risk set size at time t, so delta=1 always ###
> temp1 <- Wdataclean3(z=kidney$time[kidney[,3]==1],d=rep(1,43) )
> temp2 <- DnR(x=temp1$value, d=temp1$dd, w=temp1$weight)
> fR1 <- approxfun(x=temp2$times,y=temp2$n.risk,method="constant",yright=0,rule=2,f=1)

> temp1 <- Wdataclean3(z=kidney$time[kidney[,3]==2],d=rep(1,76) )
> temp2 <- DnR(x=temp1$value, d=temp1$dd, w=temp1$weight)
> fR2 <- approxfun(x=temp2$times,y=temp2$n.risk,method="constant",yright=0,rule=2,f=1)

### weight function for two sample Gehan-Wilcoxon test: g_{12}=g_{22} in (11) ###
> funWX <- function(t){ fR1(t)*fR2(t)/((76*43)*sqrt(119/(76*43)) )}

### Here comes the test: ###
> emplikHs.test2(x1=kidney[kidney[,3]==1,1],d1=kidney[kidney[,3]==1,2],
                 x2=kidney[kidney[,3]==2,1],d2=kidney[kidney[,3]==2,2],
                 theta=0, fun1= funWX, fun2=funWX)
$`-2LLR`
[1] 0.002473070
$lambda
[1] -0.1713749

### p-value ###
> 1-pchisq(0.002473070,df=1)
[1] 0.9603376

### the weight function for log-rank test: g_{11}=g_{21} in (11) ###
> funlogrank <- function(t){sqrt(119/(76*43))*fR1(t)*fR2(t)/(fR1(t)+fR2(t))}

### Now the log-rank test ###
> emplikHs.test2(x1=kidney[kidney[,3]==1,1],d1=kidney[kidney[,3]==1,2],
```

```
                 x2=kidney[kidney[,3]==2,1],d2=kidney[kidney[,3]==2,2],
                 theta=0, fun1=funlogrank, fun2=funlogrank)
$`-2LLR'
[1] 2.655808
$lambda
[1] 3.568833

### p-value ###
> 1-pchisq(2.655808, df=1)
[1] 0.1031723

### the weight function for both type tests ###
> funBOTH <- function(t) { cbind( funlogrank(t), funWX(t) ) }

### The test that combines both tests
> emplikHs.test2(x1=kidney[kidney$type==1,1],d1=kidney[kidney$type==1,2],
                 x2=kidney[kidney$type==2,1],d2=kidney[kidney$type==2,2],
                 theta=c(0,0), fun1=funBOTH, fun2=funBOTH)
$`-2LLR'
[1] 13.25476
$lambda
[1]  14.80228 -21.86733

### p-value ###
> 1-pchisq(13.25476, df=2)
[1] 0.001323626
```

References

Andersen, P. K., Borgan, O., Gill, R., and Keiding, N., 1993. *Statistical Models Based on Counting Processes.* Springer, New York.

Fang, H. and Zhou, M., 2000. On empirical likelihood ratio method with k-sample problems. Technical Report 380, Department of Statistics, University of Kentucky.

Fleming, T. R. and Harrington, D. P., 1991. *Counting Processes and Survival Analysis.* Wiley, New York.

Gill, R., 1980. *Censoring and Stochastic Integrals.* Mathematical Centre Tracts 124. Mathematisch Centrum, Amsterdam.

Gill, R., 1983. Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.* 11 (1), 49–58.

Harrington, D. P. and Fleming, T. R., 1982. A class of rank test procedures for censored survival data. *Biometrika* 69, 553–566.

Kaplan, E. and Meier, P. (1958), *Non-parametric Estimator From Incomplete Observations* J. Amer. Statist. Assoc. 53, 457–481.

Klein, J. P. and Moeschberger, M. L., 1997. *Survival Analysis – Techniques for Censored and Truncated Data.* Springer, New York.

Lai, T. L. and Ying, Z., 1991. Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* 19, 531–556.

Li, G., 1995. On nonparametric likelihood ratio estimation of survival probabilities for censored data. *Statist. Probab. Lett.* 25, 95–104.

Li, G. and Wang, Q. H., 2003. Empirical likelihood regression analysis for right censored data. *Statistica Sinica* 13, 51–68.

Murphy, S. A., 1995. Likelihood ratio based confidence intervals in survival analysis. *J. Amer. Statist. Assoc.* 90, 1399–1405.

Murphy, S. and Van der Vaart, 1997. Semiparametric likelihood ratio inference. *Ann. Statist.* 25, 1471–1509.

Naik-Nimbalkar, U. V. and Rajarshi, M. B., 1997. Empirical likelihood ratio test for equality of k medians in censored data. *Statist. Probab. Lett.* 34, 267–273.

Owen, A., 1988. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 75, 237–249.

Owen, A., 1990. Empirical likelihood ratio confidence regions. *Ann. Statist.* 18, 90–120.

Owen, A., 2001. *Empirical Likelihood.* Chapman & Hall, London.

Pan, X. R. and Zhou, M., 1999. Using 1-parameter sub-family of distributions in empirical likelihood ratio with censored data. *J. Statist. Plann. Inference* 75 (2), 379–392.

Pan, X. R. and Zhou, M., 2002. Empirical likelihood in terms of cumulative hazard function for censored data. *J. Multivariate Analysis* 80 (1), 166–188.

Thomas, D. R. and Grunkemeier, G. L., 1975. Confidence interval estimation of survival probabilities for censored data. *J. Amer. Statist. Assoc.* 70, 865–871.

Tsuang, M. T. and Woolson, R. F., 1977. Mortality in patients with schizophrenia, mania and depression. *British J. Psychiatry* 130, 162–166

A Tables and Figures

n	α				
	0.01	0.05	0.1	0.15	0.2
30	0.0144	0.0550	0.1032	0.1511	0.1981
50	0.0103	0.0524	0.1057	0.156	0.2052
100	0.0102	0.0476	0.1	0.1508	0.1999

Table 1: Estimated type I errors at various significance levels α (two samples).