Nonparametric K-Sample Tests with Panel Count Data

Ying Zhang
Department of Biostatistics
University of Iowa

November 15, 2005

Abstract

In this manuscript, we study the nonparametric k-sample test problem with panel count data. The asymptotic normality of a smooth functional of the nonparametric maximum pseudo-likelihood estimator (Wellner and Zhang, 2000) is established under some mild conditions. We construct a class of easy-to-implement nonparametric tests for comparing mean functions of k populations based on this asymptotic normality. We conduct various simulations to validate and compare the tests. The simulations show that the tests perform quite well and generally have a good power to detect difference of the mean functions. The method is illustrated with two real data examples.

Some key words: Counting Process; Empirical Process; Interval censored data; Monte-Carlo; Isotonic Regression.

1. Introduction

Suppose that $\mathbb{N} = \{\mathbb{N}(t) : t \geq 0\}$ is a counting process with mean function $E\mathbb{N}(t) = \Lambda_0(t)$. In applications, recurrence of some undesirable repeated events such as tumor and stroke can be modelled as realization of a counting process. Continuous observation over a counting process up to a time point constitutes so called *event history data*. There is a large amount of researches in literatures for event history data; for example, Prentice, Williams, and Peterson (1981), Andersen and Gill (1982), Pepe and Cai (1993), Lawless and Nadeau (1995), Cook, Lawless, and Nadeau (1996), and Lin, Wei, Yang, and Ying (2000).

In this manuscript, we consider incomplete observation on the counting process, in which the event process (counting process) is only observed at a sequence of random times $0 < T_{K,1} < T_{K,2} < \cdots < T_{K,K}$, where the total number of observations K is also assumed to be an integer-valued random variable. The numbers of a recurrent event up to these times, $0 \leq \mathbb{N}(T_{K,1}) \leq \mathbb{N}(T_{K,2}) \leq \cdots \leq \mathbb{N}(T_{K,K})$, are observed accordingly. The observed data for the counting process consist of $X = (K, \underline{T}, \underline{\mathbb{N}})$, where $\underline{T} = (T_{K,1}, T_{K,2}, \cdots, T_{K,K})$ and $\underline{\mathbb{N}} = \{\mathbb{N}(T_{K,1}), \mathbb{N}(T_{K,2}), \cdots, \mathbb{N}(T_{K,K})\}$. This type of data is referred to as *panel count data*.

Panel count data occur frequently in fields such as demographic studies, reliability, clinical trials, and health service researches. Statistical methodology researches on this type of data can be found in Gaver and O'Muircheataigh (1987), Thall and Lachin (1988), Thall (1988), and Kalbfleisch and Lawless (1985), in which some relatively simplified versions of panel count data were studied. Sun and Kalbfleisch (1995) appeared to be the first studying the nonparametric estimation of the mean function with panel count data aforementioned. They used the isotonic regression technique to estimate the mean function of the counting process. Wellner and Zhang (2000) studied likelihood-based nonparametric estimation methods based on a "working model" of nonhomogeneous Poisson process. They showed that the nonparametric maximum pseudo-likelihood estimator is exactly the one studied in Sun and Kalbfleisch (1995) and they also studied the asymptotic properties of both nonparametric maximum pseudo-likelihood and likelihood estimators. The basic picture is that both estimators are consistent but pointwisely converge to the true mean function in a lower rate ($n^{-1/3}$). The maximum likelihood estimator, while more efficient than the maximum pseudo-likelihood estimator, demands much more computationally and its property is hard to study.

More recently, researchers have developed some semiparametric regression methods for the proportional mean model with panel count data, namely, $\Lambda(t|Z) = \Lambda_0(t) \exp(\beta^T Z)$. Sun and Wei (2000) and Hu, Sun, and Wei (2003) devised some simple estimation procedures to estimate regression parameters using generalized estimating equation technique. Their methods, however, rely on extra assumptions on the observation scheme and thus are restricted in practice. Zhang (2002), Wellner, Zhang, and Liu (2004) studied semiparametric likelihood-based methods for estimating both the regression parameters and baseline mean function. The properties of the estimators are comparable with those in nonparametric estimation. The semiparametric maximum likelihood estimator is more efficient than the semiparametric maximum pseudo-likelihood estimator but is much more involved in computation. Numerical evidence of above is given by Wellner and Zhang (2005) through several simulation studies. Motivated by the asymptotic theorem developed by Huang (1996) in studying interval censored data, they also derived the asymptotic normality for the estimators of the regression parameters. Their methods, while having the advantage in independence of the observation scheme, face a challenge in estimating the asymptotic variance. Wellner and Zhang (2005) used the bootstrap procedure to make inference with a quite bit effort in computing, especially for the likelihood case.

This manuscript concerns about comparison of the mean functions of k populations. We assume that data consist of k independent samples of panel count data randomly drawn from k populations correspondingly. Our goal is to construct a test statistic for testing the null hypothesis:

$$H_0 \quad \Lambda_1(t) = \Lambda_2(t) = \cdots = \Lambda_k(t) = \Lambda_0(t), \tag{1.1}$$

where $\Lambda_i(t)$ is the mean function of the ith population for $i = 1, 2, \ldots, k$. Although the semiparametric methods as mentioned above can be used to make the inference, they all have some shortcomings, either on computation complexity or requiring extra assumptions on the observation scheme that may not be realistic in practice. Sun and Fang (2003) developed a simple nonparametric procedure to test the equity of mean functions for all counting processes in a random sample. Their method, though maybe applicable to k-sample test, requires an extra assumption that each process has equal chance to be assigned to any of the k samples, which again may not be realistic in practice.

The rest of paper is organized as follows: In Section 2, we derive the convergence rate of the nonparametric maximum pseudo-likelihood estimator in a specified L_2 normal and establish the asymptotic normality of a smooth functional of the estimator. In Section 3, we construct a class of easy-to-implement test statistics and describe the asymptotic property of the test statistics. In Section 4, we conduct various Monte-Carlo simulation studies to validate the method and compare the power of these tests. We also illustrate our method with two real life data sets, one from the famous bladder tumor study and other from a HIV study. In Section 5, we summarize our results and discuss further development in this area. Finally, we include all the technical proofs of the theorems in the Appendix.

2. The Nonparametric Maximum Pseudo-Likelihood Estimator

Wellner and Zhang (2000) proposed a simple nonparametric estimation method for the mean function of the counting process with panel count data. Assuming the underline counting process being a nonhomogeneous Poisson process and ignoring completely correlations of the cumulative counts $\underline{\mathbb{N}}$, they established the log pseudo-likelihood by omitting the parts irrelevant to the mean function Λ as follows,

$$l_n(\Lambda;\underline{X}) = \sum_{i=1}^{n}\sum_{j=1}^{K_i}\left\{\mathbb{N}^{(i)}(T_{K_i,j})\log\Lambda(T_{K_i,j}) - \Lambda(T_{K_i,j})\right\}, \tag{2.1}$$

where data $\underline{X}$ are the collection of i.i.d copies of X, $X_i = (K_i, \underline{T}_i, \underline{\mathbb{N}}^{(i)})$ for $i = 1, 2, \cdots, n$ with $\underline{T}_i = (T_{K_i,1}, T_{K_i,2}, \cdots, T_{K_i,K_i})$ and $\underline{\mathbb{N}}^{(i)} = \{\mathbb{N}^{(i)}(T_{K_i,1}), \mathbb{N}^{(i)}(T_{K_i,2}), \cdots, \mathbb{N}^{(i)}(T_{K_i,K_i})\}$.

The nonparametric maximum pseudo-likelihood estimator of Λ, $\hat{\Lambda}_n$ is defined to be the step function that has jumps only possibly at the collection of observed times and maximizes (2.1). Wellner and Zhang (2000) have shown that this estimator is easy to calculate and, in fact, it is exactly the isotonic regression estimator proposed by Sun and Kalbfleisch (1995). Wellner and Zhang (2000) also studied asymptotic properties of this estimator. They concluded that the estimator is strongly consistent in a L_2-norm with the measure defined by Sichck and Yu (2000) as

$$\mu(t) = \sum_{k=1}^{\infty} P(K=k)\sum_{j=1}^{K} P(T_{K,j} \le t | K=k) \tag{2.2}$$

under some mild regularity conditions. That is

$$d^2(\hat{\Lambda}_n, \Lambda_0) = \left[\int_{S[T]} \left\{\hat{\Lambda}_n(t) - \Lambda_0(t)\right\}^2 d\mu(t)\right]^{1/2} \to_{a.s} 0, \tag{2.3}$$

where $S[T] = \{t : \quad 0 < t \leq \tau\}$ for some $\tau > 0$. The τ can be viewed as the termination time in a clinical follow-up study. With more mild regularity conditions, they also derived the pointwise asymptotic distribution of the estimator. That is,

$$n^{1/3}\left\{\hat{\Lambda}_n(t_0) - \Lambda_0(t_0)\right\} \to_d 2\left\{\frac{\sigma^2(t_0)\Lambda_0'(t_0)}{2\mu'(t_0)}\right\}^{1/3} \text{argmax}_h \left\{\mathbb{Z}(h) - h^2\right\}, \tag{2.4}$$

for any $t_0 \in S[T]$, where $\sigma^2(t) \equiv Var\{\mathbb{N}(t)\}$ and $\mathbb{Z}$ is a two-sided Brownian motion process started from zero.

The asymptotic properties (2.3) and (2.4) are, however, not directly applicable in construction of test statistics for comparing the mean functions. We expand the study on asymptotic properties of the estimator initiated by Wellner and Zhang (2000) for the purpose of k-sample comparison. The following regularity conditions (in addition to those given in Wellner and Zhang, 2000) are sufficient for the forthcoming theorems.

C1. For some interval $O[T] = [\sigma, \tau]$ with $\sigma > 0$ and $\Lambda_0(\sigma) > 0$, $P(\cap_{j=1}^{K}\{T_{K,j} \in [\sigma, \tau]\}) = 1$.

C2. There exists a positive integer k_0 such that $P(K \leq k_0) = 1$, i.e. the number of observations is finite.

C3. $E\left\{e^{C\mathbb{N}(t)}\right\}$ is uniformly bounded above for $t \in S[T]$ and some constant C.

C4. The true baseline mean function Λ_0 is differentiable and the derivative has a positive lower bound in the observation interval, i.e. there exists a constant $f_0 > 0$ such that $\Lambda_0'(t) \geq f_0$ for $t \in O[T]$.

C5. There are some functions η such that for every such η, $\eta \circ \Lambda_0^{-1}$ is a bounded monotone Lipschitz function.

Conditions C1 and C2 are easily justified in many applications. C3 is true if $\mathbb{N}(t)$ is uniformly bounded, which usually holds in view of clinical studies, or if $\mathbb{N}(t)$ is a Poisson or mixed Poisson process. C4 and C5 are mainly used in technical development. C4 indicates that our results are valid when the true mean function satisfies some smoothness conditions. With C4, C5 automatically holds for $\eta \equiv 1$. These conditions, in general, are mild in view of applications.

Theorem 2.1 (Rate of Convergence) Suppose that C1-C3 hold, then

$$n^{1/3} d(\hat{\Lambda}_n, \Lambda_0) = O_p(1).$$

Next, we study a smooth functional of $\hat{\Lambda}_n$, $\nu(\hat{\Lambda}_n) = \int_{S[T]} \eta(t)\hat{\Lambda}_n(t)d\mu(t)$ with a weight function $\eta(t)$. In practice, one may choose to use $\eta_1(t) \equiv 1$ (equal weight on the observations), $\eta_2(t) = P(T_{K,K} \geq t)$ (heavy weight on the earlier observations), or $\eta_3(t) = P(T_{K,K} < t)$ (heavy weight on the later observations).

Theorem 2.2 (Asymptotic Normality) Suppose that C1-C5 hold, then

$$n^{1/2} \left\{ \nu(\hat{\Lambda}_n) - \nu(\Lambda_0) \right\} = W + o_p(1), \tag{2.5}$$

where $W \sim N(0, \Omega)$ with

$$\Omega = E\left[\sum_{j=1}^{K} \eta(T_{K,j}) \left\{ \mathbb{N}(T_{K,j}) - \Lambda_0(T_{K,j}) \right\} \right]^2.$$

This theorem can be viewed as a generalisation of the result given by Huang and Wellner (1995), since Wellner and Zhang (2000) found out that the characteristics of the nonparametric maximum pseudo-likelihood estimator of the mean function with panel count data are structurally similar to those of the nonparametric maximum likelihood estimator of the distribution function with current status data. The proofs of the theorems are given in the Appendix.

3. A Class of Nonparametric K-Sample Test Statistics

In order to construct a test statistic for H_0 with $k \geq 3$, we first estimate the difference of the functional ν for any two samples under the null hypothesis (1.1). Without loss of generality, we consider

$$\begin{aligned}
\nu_n(\hat{\Lambda}_{n_1}, \hat{\Lambda}_{n_2}) &= \sqrt{\frac{n_1 n_2}{n}} \int_{S[T]} \hat{\eta}_n(t) \left\{ \hat{\Lambda}_{n_1}(t) - \hat{\Lambda}_{n_2}(t) \right\} d\hat{\mu}_n(t) \\
&= \sqrt{\frac{n_1 n_2}{n}} \sum_{i=1}^{n} \sum_{j=1}^{K_i} \left[\hat{\eta}_n(T_{K,j}) \left\{ \hat{\Lambda}_{n_1}(T_{K,j}) - \hat{\Lambda}_{n_2}(T_{K,j}) \right\} \right], \qquad (3.1)
\end{aligned}$$

where $\hat{\Lambda}_{n_1}$ and $\hat{\Lambda}_{n_2}$ are the nonparametric maximum pseudo-likelihood estimators of Λ_0 using the individual sample data, respectively; $\hat{\eta}_n$ is the sample estimate of the weight function $\eta(t)$ (For example $\hat{\eta}_{2,n}(t) = 1/n \sum_{i=1}^{n} 1_{[T_{K_i,K_i} \geq t]}$ for $\eta_2(t)$) and $\hat{\mu}_n(t)$ is the empirical estimate of the measure $\mu(t)$ given by $\hat{\mu}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K_i} 1_{[T_{K,j} \leq t]}$ using the pooled data.

Theorem 3.1 In addition to C1-C5, we also assume that

i. The distribution of panel observations $(K, \underline{T})$ is the same across the samples.

ii. For a bounded monotone function η, there is a consistent estimate $\hat{\eta}_n$ such that $n^{1/6} d(\hat{\eta}_n, \eta) = o_p(1)$.

iii. $n_1/n \to p_1$ and $n_2/n \to p_2$ as $n \to \infty$.

Then, $\nu_n(\hat{\Lambda}_{n_1}, \hat{\Lambda}_{n_2}) = \sqrt{p_2} W_1 - \sqrt{p_1} W_2 + o_p(1)$, where $W_1, W_2 \sim N(0, \Omega)$ and are independent.

The proof of this theorem is also given in the Appendix. Based on this theorem, we can now easily construct a class of test statistics for H_0. In a sequel, we suppress $S[T]$ in the integral for clarity. Let

$$\begin{aligned}
\nu_n^{1,2}(\hat{\Lambda}_{n_1}, \hat{\Lambda}_{n_2}) &= \sqrt{\tfrac{n_1 n_2}{n}} \int \hat{\eta}_n(t) \left\{ \hat{\Lambda}_{n_1}(t) - \hat{\Lambda}_{n_2}(t) \right\} d\hat{\mu}_n(t) \\
\nu_n^{1,3}(\hat{\Lambda}_{n_1}, \hat{\Lambda}_{n_3}) &= \sqrt{\tfrac{n_1 n_3}{n}} \int \hat{\eta}_n(t) \left\{ \hat{\Lambda}_{n_1}(t) - \hat{\Lambda}_{n_3}(t) \right\} d\hat{\mu}_n(t) \\
&\vdots \\
\nu_n^{1,k}(\hat{\Lambda}_{n_1}, \hat{\Lambda}_{n_k}) &= \sqrt{\tfrac{n_1 n_k}{n}} \int \hat{\eta}_n(t) \left\{ \hat{\Lambda}_{n_1}(t) - \hat{\Lambda}_{n_k}(t) \right\} d\hat{\mu}_n(t).
\end{aligned}$$

Suppose that $n_1/n \to p_1$, $n_2/n \to p_2$, $\cdots$, and $n_k/n \to p_k$, as $n \to \infty$. Then based on the result of Theorem 3.1, under the null hypothesis (1.1) we have

$$\begin{aligned}
\nu_n^{1,2}(\hat{\Lambda}_{n_1}, \hat{\Lambda}_{n_2}) &= \sqrt{p_2} W_1 - \sqrt{p_1} W_2 + o_p(1) \\
\nu_n^{1,3}(\hat{\Lambda}_{n_1}, \hat{\Lambda}_{n_3}) &= \sqrt{p_3} W_1 - \sqrt{p_1} W_3 + o_p(1) \\
&\vdots \\
\nu_n^{1,k}(\hat{\Lambda}_{n_1}, \hat{\Lambda}_{n_k}) &= \sqrt{p_k} W_1 - \sqrt{p_1} W_k + o_p(1),
\end{aligned}$$

where $W_1, W_2, \cdots, W_k \sim N(0, \Omega)$ and they are all mutually independent.

Denote $\nu_n = \left\{\nu_n^{1,2}(\hat{\Lambda}_{n_1}, \hat{\Lambda}_{n_2}), \nu_n^{1,3}(\hat{\Lambda}_{n_1}, \hat{\Lambda}_{n_3}), \cdots, \nu_n^{1,k}(\hat{\Lambda}_{n_1}, \hat{\Lambda}_{n_k})\right\}^T$, we have that $\nu_n = \underline{W} + o_p(1)$, where $\underline{W} \sim N(0, B)$ with

$$B = \Omega \begin{bmatrix} p_1 + p_2 & \sqrt{p_2 p_3} & \cdots & \sqrt{p_2 p_k} \\ \sqrt{p_2 p_3} & p_1 + p_3 & \sqrt{p_3 p_4} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \sqrt{p_2 p_k} & \sqrt{p_3 p_k} & \cdots & p_1 + p_k \end{bmatrix}.$$

B can be consistently estimated by

$$\hat{B}_n = \hat{\Omega}_n \begin{bmatrix} \frac{n_1+n_2}{n} & \sqrt{\frac{n_2 n_3}{n^2}} & \cdots & \sqrt{\frac{n_2 n_k}{n^2}} \\ \sqrt{\frac{n_2 n_3}{n^2}} & \frac{n_1+n_3}{n} & \sqrt{\frac{n_3 n_4}{n^2}} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \sqrt{\frac{n_2 n_k}{n^2}} & \sqrt{\frac{n_3 n_k}{n^2}} & \cdots & \frac{n_1+n_k}{n} \end{bmatrix}$$

with

$$\hat{\Omega}_n = \frac{1}{n} \sum_{i=1}^{n} \left[\sum_{j=1}^{K_i} \hat{\eta}_n(T_{K,j}) \left\{ \mathbb{N}^{(i)}(T_{K_i,j}) - \hat{\Lambda}_n(T_{K_i,j}) \right\} \right]^2,$$

in which $\hat{\eta}_n$ and $\hat{\Lambda}_n$ are obtained using the pooled data. Therefore a reasonable test statistic for H_0 can be easily constructed by the quadratic form $T_n = \nu_n^T \hat{B}_n^{-1} \nu_n$, which has the sampling distribution χ^2_{k-1} asymptotically. Owing to ease in computing the nonparametric maximum pseudo-likelihood estimator of the mean function with panel count data, it is also very easy to calculate $\hat{\Omega}_n$, a consistent estimate of Ω. Hence the implementation of the test is very little demanding in computation as compared to the Wald test based on the hypothesis of proportional mean model studied in Wellner and Zhang (2005).

4. Numerical Results

4.1. Simulation Studies

We evaluate the performance of the tests for case $k = 3$ (three-sample comparison) in terms of size and power using the simulation settings described in Wellner and Zhang (2000).

Scenario 1 (Data from Poisson processes). $\{(K_i, \underline{T}_i): \; i = 1, 2, \cdots, n\}$ ($n = n_1+n_2+n_3$) is a random sample, where K_i is drawn uniformly in $\{1, 2, 3, 4, 5, 6\}$ for $i = 1, 2, \ldots, n$. Given K_i, the panel observation times $\underline{T}_i = (T_{K_i,1}, T_{K_i,2}, \ldots, T_{K_i,K_i})$ are made from the order statistics of K_i random observations, generated from the distribution, **Unif(0,10)**. The panel counts are made from the following counting processes: $\mathbb{N}^{(i)}(t) \sim Poisson(\lambda t)$ for $i = 1, 2, \ldots, n_1$; $\mathbb{N}^{(i)}(t) \sim Poisson\{(1+\theta_1)\lambda t\}$ for $i = n_1+1, n_1+2, \ldots, n_1+n_2$; $\mathbb{N}^{(i)}(t) \sim Poisson\{(1+\theta_2)\lambda t\}$ for $i = n_1 + n_2 + 1, n_1 + n_2 + 2, \ldots, n_1 + n_2 + n_3$. We compare the test statistics with the weight functions $\eta_n(t) \equiv 1, 1/n \sum_{i=1}^{n} 1_{[T_{K_i,K_i} \geq t]}$, and $1/n \sum_{i=1}^{n} 1_{[T_{K_i,K_i} < t]}$, respectively, and we call them $T_n^{(1)}$, $T_n^{(2)}$, and $T_n^{(3)}$ accordingly.

We choose $\lambda = 1$ and consider all possible (θ_1, θ_2) resulted from the combination of 0, $0 \cdot 1, 0 \cdot 25$, and $0 \cdot 5$. We conduct Monte-Carlo simulation studies with $n_1 = n_2 = n_3 = 50$ and $n_1 = n_2 = n_3 = 100$, respectively. For each study, 1000 repeated samples are generated and the corresponding test statistics are calculated. The proportion of the test statistics beyond $\chi^2_{2,0\cdot 95}$ (rejection region with significance level $0 \cdot 05$) is computed for each case and reported in Table 1.

This study shows that all three test statistics $T_n^{(1)}$, $T_n^{(2)}$, and $T_n^{(3)}$ perform well with moderate sample size based on the large sample property. The empirical sizes of the tests are all close to their nominal value $0 \cdot 05$. These tests, in general, have a satisfactory power to detect a departure from the null hypothesis. When the three mean functions differ by at most 50%, (Cases 4, 7, 9, and 10) the test power is virtually 100% for the small study (n=150). For the larger study (n=300), the test power is at least 90%, if not 100%, for the mean functions differing by at most 25%.

For all the alternative hypotheses, the test statistics $T_n^{(1)}$ and $T_n^{(3)}$ tend to have a better power than $T_n^{(2)}$. This is due to the fact that $T_n^{(2)}$ weights earlier observations more than later observations, in which the mean functions differ the greatest if they are different. While the gain in power for test $T_n^{(3)}$ over $T_n^{(2)}$ is ascribed to the aforementioned reason, we notice that the equally weighted test $T_n^{(1)}$ performs just as good as $T_n^{(3)}$ and even slightly better in many occasions. $T_n^{(3)}$ puts a zero weight on the early but crowded observations, so the difference of the estimators in the early times does not contribute to the difference in the functionals. Replacing zero weight by the smallest non-zero weight for the early observations may potentially improve the power.

Scenario 2 (Data from one-jump processes). $\{(K_i, \underline{T}_i) : \quad i = 1, 2, \cdots, n\}$ $(n = n_1 + n_2 + n_3)$ are generated exactly the same as described in Scenario 1. Denote V_i the event time generated from **Exponential** distributions: $V_i \sim \exp(\lambda)$ for $i = 1, 2, \ldots, n_1$; $V_i \sim \exp\{\lambda(1 + \theta_1)\}$ for $i = n_1+1, n_1+2, \ldots, n_1+n_2$; and $V_i \sim \exp\{\lambda(1+\theta_2)\}$ for $i = n_1+n_2+1, n_1+n_2+2, \ldots, n_1+n_2+n_3$. The panel counts $\underline{\mathbb{N}}_i = \left(\mathbb{N}_{K_i,1}^{(i)}, \mathbb{N}_{K_i,2}^{(i)}, \ldots, \mathbb{N}_{K_i,K_i}^{(i)}\right)$ are made by $\mathbb{N}_{K_i,j}^{(i)} = 1_{[V_i \leq T_{K_i,j}]}$ for $j = 1, 2, \cdots, K_i$. This type of data, as a special case of panel count data, was referred to as mixed case interval censored data by Schick and Yu (2000). The mean function of the process is just the distribution function of the event time.

In this study, we select $\lambda = 0 \cdot 2$ and consider all possible (θ_1, θ_2) resulted from the combination of $0, 0{\cdot}25, 0{\cdot}50$, and $1{\cdot}00$. Monte-Carlo simulation studies with 1000 repetitions are conducted for $n_1 = n_2 = n_3 = 50$ and $n_1 = n_2 = n_3 = 100$, respectively. Results are given in Table 2. The results show again that all these tests have the empirical size close to the nominal level 0·05 with moderate sample size. In this case, $T_n^{(3)}$, however, performs noticeably inferior to the first two tests. This is due to the fact that the maximum difference of the three mean functions appears in an early time of the interval [0,10]. For example, the biggest difference of the mean functions occurs at $t = 3 \cdot 466$ in Case 4. Clearly, putting small weight around this area as $T_n^{(3)}$ does diminishes the difference in the functionals.

The behavior of the test statistics are expected to be similar to that of the weighted log rank test statistics. In general, they have a good power to detect the difference when the mean functions do not cross over. The power may reduce dramatically otherwise. For illustration, we design the third set of simulations which incorporates the scenario that the mean functions cross over.

Scenario 3 We consider the mixed case interval censored data as described in Scenario 2. But the event time is generated from the following: $V_i \sim \text{Wei}(1, 1)$ for $i = 1, 2, \ldots, n_1 + n_2$ and $V_i \sim \text{Wei}(a, a)$ for $i = n_1 + n_2 + 1, 2, \ldots, n_1 + n_2 + n_3$. In this scenario, the distribution function of the event time is $1 - \exp(-t)$ for the first two groups and $1 - \exp\left\{-\left(\frac{t}{a}\right)^a\right\}$ for the third group. Figure 1 plots these functions with various a. Table 3 shows the simulation results based on 1000 repetitions with $n_1 = n_2 = n_3 = 50$ and $n_1 = n_2 = n_3 = 100$, respectively. It is quite apparent that all three tests will yield small power in detecting the difference for $a = 0 \cdot 75$ or $0 \cdot 50$ since the powers do not increase as sample size doubles. That the positive difference of the mean functions in the test statistics between group 1 and group 3 offsets the negative difference as shown in Figure 1 is the legitimate explanation for the lower power. When $a = 0 \cdot 25$, the three tests have a reasonable power mainly because the positive part of the difference appears dominating the negative part. Test statistics $T_n^{(3)}$ has the largest power as it weights the positive difference more than other two test statistics.

Selection of the weight function is a subjective matter. In practice, having a prior knowledge about the shapes of the mean functions will help select the weight function to improve the power. Our simulations indicate that when the mean functions do not cross over, the equally weighted test tends to have a good power. However, when they cross over, the test with a right weight function may have a significantly improved power but the degree of improvement depends on how the mean functions cross over.

4.2. Two Real-Life Examples

Example 1 (Bladder Tumor Trial) We apply our method to the well-known data set extracted from a bladder tumor study conducted by the Veterans Administration Cooperative Urological Research Group (VACURG). (Andrews and Herzberg, 1985) Previous studies of this data set can be found in Byar, Blackard, and the VACURG (1977), Byar (1980), Wei, Lin and Weissfeld (1989), Wellner and Zhang (2000), Sun and Wei (2000), and Zhang (2002).

The data set consists of 116 patients in a randomized clinical trial in which patients experienced superficial bladder tumor when entering the trial and were assigned randomly into one the three arms: placebo (47 patients), pyridocine pills (31 patients), and periodic instillation of a chemotherapeutic agent (38 patients), thiotepa. The follow-up number and time vary greatly from patient to patient. At each follow-up visit, any tumors noticed were counted, measured and then removed transurethrally, and the treatments continued. This data set fits our panel count data framework perfectly. In this manuscript, we compare the tumor recurrence under the three arms based on the nonparametric maximum pseudo-likelihood estimator of the mean function of tumor counts. The estimators are plotted in Figure 2. We select placebo as the reference group in our analysis that results $T_n^{(1)} = 4 \cdot 928141$, $T_n^{(2)} = 3 \cdot 868270$ and $T_n^{(3)} = 4 \cdot 952747$ with p-value=0·0851, 0·1445 and 0·0840, respectively. The tests based on $T_n^{(1)}$ and $T_n^{(3)}$ reject the null hypothesis at level 0·1 but fail to reject at level 0·05. The rational that both $T_n^{(1)}$ and $T_n^{(3)}$ outperform $T_n^{(2)}$ in terms of test power can be easily understood. From Figure 2, we can see clearly that the greatest difference of the three samples appears in later times of the trial, so having close to zero weights for the observations near the end discourages the difference between the functionals of the estimators. We also conduct the multiple pairwise comparisons using Bonferroni technique and we only detect the difference between the pyridocine and thiotepa treatments at overall level 0·1 in the tests $T_n^{(1)}$ and $T_n^{(3)}$ with p-value=0·0326 and 0·0267, respectively.

Sun and Wei (2000) and Zhang (2002) have detected the difference between thiotepa treatment and placebo at significance level 0·05 using semiparametric regression analysis. Their conclusions, however, are made based on the proportional mean model, which may not be realistic in this application. This assumption was questioned by Wellenr and Zhang (2005) as the three estimators shown in Figure 2 cross over in earlier times of the trial. In addition, the crossing of these estimators may contribute to a lower power in detecting the difference as indicated through simulation scenario 3.

Example 2 (HIV Clinical Trial) This a an ongoing randomized AIDS clinical trial for comparing three antiretroviral regimens. A total of 513 HIV-1-infected patients were randomized to this study in US. Arm A (166 patients) of the study is a standard 3-drug regimen serving as a control group, while Arms B (171 patients) and C (176 patients) are new 4-drug regimens. Although several primary endpoints are used to compare the long-term effectiveness of the three regimens in this study, we focus on comparing time-to-detection-limit of HIV-1 RNA assays (50 copies per ml plasma) which is a measurement of time-to-response of antiviral therapies. A shorter time-to-response of a regimen indicates that the regimen is more potent. HIV-1 RNA levels were scheduled to be monitored at weeks 4, 8 and every 8 weeks thereafter during treatment. However, individual patients may not exactly follow this schedule. So it is reasonable to assume probability distributions for the follow-up monitoring times. Drop-out from the monitoring is very common in the study. The number of monitoring varies greatly from subject to subject, ranging from 1 to 28. The time-to-detection-limit is only known in an interval of two consecutive monitoring times, so the data fit the framework of the mixed case interval censored data. Zhang, Liu and Hu (2003) constructed a simple nonparametric two-sample test for comparing the difference of the distribution function of time-to-response between the two regimens under Arms A and B.

Applying the proposed method to this example, the three tests result in $T_n^{(1)} = 12 \cdot 04938$, $T_n^{(2)} = 12 \cdot 47060$ and $T_n^{(3)} = 9 \cdot 311337$ with p-value=0·0024, 0·0020, and 0·0095, respectively. We reject the null hypothesis using any of these tests at level 0·05. We, however, notice that the test $T_n^{(2)}$ appears to be the most powerful and $T_n^{(3)}$ is the least powerful for this example among the three tests. This can be easily understood since the biggest difference of the three estimators occurs at an earlier time of the study as shown in Figure 3 and these estimators tend to converge to one in the end of study.

Using Bonferroni approach, at overall significance level 0·05, we only found pairwise difference between Arms B and C with p-value=0·0006, 0·0005 and 0·0034, respectively. Pairwise difference between Arms A and B can be detected only at overall significance level 0·1 with p-value=0·0259, 0·0278 and 0·0272, respectively, though the difference is quite transparent in Figure 3.

5. Summary

This manuscript introduces a class of nonparametric tests for comparing the mean functions of k samples with panel count data. The test statistics are constructed based on a smooth functional of the nonparametric maximum pseudo-likelihood estimator of the mean function. Large sample properties of the test statistics are thoroughly studied using empirical process theory. Various simulation studies with moderate sample size are conducted to evaluate the performance of the test statistics. The simulation results show that the tests, in general, have a good power to detect difference in mean functions of multiple samples, especially when the mean functions do not cross over.

In this study, we select the first sample as the reference sample for the comparison. Actually, the selection of the reference sample does not affect the value of the test statistic, since any pair difference of the functionals is a linear combination of those used in the construction of the test statistics.

The tests we conducted in this manuscript, however, rely on the assumption that the observation scheme with respect to the total number of observation K and panel observation times $\underline{T}$ are the same across different samples. Obviously, this assumption may not be easily verified in practice in view of many applications in clinical trials. To avoid this assumption, one can modify the two-sample test statistic using

$$\nu_n(\hat{\Lambda}_{n_1}, \hat{\Lambda}_{n_2}) = \sqrt{\frac{n_1 n_2}{n}} \left\{ \int \hat{\Lambda}_{n_1}(t)\hat{\eta}_{n_1}(t) d\hat{\mu}_{n_1}(t) - \int \hat{\Lambda}_{n_2}(t)\hat{\eta}_{n_2}(t) d\hat{\mu}_{n_2}(t) \right\},$$

where $\hat{\eta}_{n_i}(t)$ $(i = 1, 2)$ can be chosen as the inverse of the kernel smoothing estimate of $\mu_i'(t)$. Here μ_i is the measure μ for sample i, and $\hat{\mu}_{n_i}$ is the empirical estimate of μ_i defined earlier. It is still relatively easy to compute this test statistic with some extra effort for getting a consistent estimate $\hat{\eta}_{n_i}$ $(i = 1, 2)$. The quantity $\nu_n(\hat{\Lambda}_{n_1}, \hat{\Lambda}_{n_2})$ can be easily shown to have the asymptotic normality with mean zero and asymptotic variance given by

$$p_2 E\left[\sum_{j=1}^{K} \eta_1(T_{K,j}) \left\{\mathbb{N}(T_{K,j}) - \Lambda_0(T_{K,j})\right\}\right]^2 + p_1 E\left[\sum_{j=1}^{K} \eta_2(T_{K,j}) \left\{\mathbb{N}(T_{K,j}) - \Lambda_0(T_{K,j})\right\}\right]^2,$$

which is again easily estimable.

Wellner and Zhang (2000) have shown that the nonparametric maximum likelihood estimator has a better asymptotic property in terms of estimation efficiency compared to the nonparametric maximum pseudo-likelihood estimator. It is not clear whether this advantage is transformable into the performance of the test statistics. It is of interest to develop similar test statistics using the nonparametric maximum likelihood estimators and compare them with the tests proposed in this manuscript. The task remains for the future research.

ACKNOWLEDGEMENT

The authors is grateful to two referees for their comments and suggestions that help improve the manuscript.

6. Appendix

In this section, we sketch the proofs for the three theorems. We mainly use the modern empirical process theory justifying our arguments. Throughout this section, we adopt the empirical process notations used in van der Vaart and Wellner (1996). We denote $Pf = \int f dP$ and $\mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^n f(X_i)$. We let C be a constant, which may represent different values at various places.

Proof of Theorem 2.1:

Let $m_\Lambda(X) = \sum_{j=1}^K \{\mathbb{N}(T_{K,j}) \log \Lambda(T_{K,j}) - \Lambda(T_{K,j})\}$ and define $\mathbb{M}(\Lambda) = Pm_\Lambda(X)$. Then the log pseudo-likelihood function (2.1) can be represented by $l_n(\Lambda; \underline{X}) = n\mathbb{P}_n m_\Lambda(X)$. We also restrict the mean function Λ to be inside the class

$$\mathcal{F} = \{\Lambda : \quad (0, \tau] \rightarrow [0, M] | \;\; \Lambda \;\; \text{is monotone nondecreasing and} \;\; \Lambda(0) = 0\},$$

for some $M < \infty$. Hence $\mathbb{M}$ defines a smooth functional for the class $\mathcal{F}$.

We derive the rate of convergence based on Theorem 3.2.5 of van der Vaart and Wellner (1996). First, let $h(x) = x\log(x) - x + 1$. It can be easily shown by Taylor expansion that

$h(x) \geq \frac{1}{4}(x-1)^2$ for x in a neighborhood of $x = 1$ and thus for any Λ in a neighborhood of Λ_0, a simple algebraic calculation yields $\mathbb{M}(\Lambda_0) - \mathbb{M}(\Lambda) \geq Cd^2(\Lambda_0, \Lambda)$. So the separation condition of Theorem 3.2.5 of van der Vaart and Wellner (1996) satisfies. Second, we consider the classes $\mathcal{F}_\delta = \{\Lambda \in \mathcal{F} : \quad d(\Lambda, \Lambda_0) \leq \delta\}$ for $\delta > 0$ and $\mathcal{M}_\delta = \{m_\Lambda(X) - m_{\Lambda_0}(X) : \Lambda \in \mathcal{F}_\delta\}$. Since $\mathcal{F}_\delta$ is a class of monotone nondecreasing functions, by Theorem 2.7.5 of van der Vaart and Wellner (1996), for any $\epsilon > 0$, there exists a set of brackets: $\{[\Lambda_i^l, \Lambda_i^r] : \quad i = 1, 2, \cdots, q\}$ with $q \leq \exp(C/\epsilon)$, such that for any $\Lambda \in \mathcal{F}_\delta$, $\Lambda_i^l(t) \leq \Lambda(t) \leq \Lambda_i^r(t)$ for all $t \in O[T]$ and some $1 \leq i \leq q$, and $d^2(\Lambda_i^r, \Lambda_i^l) = \int \left\{\Lambda_i^r(t) - \Lambda_i^l(t)\right\}^2 d\mu(t) \leq \epsilon^2$. (Here we use the fact that μ is a finite measure with our hypotheses and therefore can be normalized to be a probability measure) Hence, we can construct a set of brackets for $\mathcal{M}_\delta$: $\{[m_i^l(X), m_i^r(X)] : i = 1, 2, \cdots, q\}$ by letting

$$m_i^l(X) = \sum_{j=1}^{K} \left\{\mathbb{N}(T_{K,j}) \log \Lambda_i^l(T_{K,j}) - \Lambda_i^r(T_{K,j})\right\} - m_{\Lambda_0}(X)$$

and

$$m_i^r(X) = \sum_{j=1}^{K} \left\{\mathbb{N}(T_{K,j}) \log \Lambda_i^r(T_{K,j}) - \Lambda_i^l(T_{K,j})\right\} - m_{\Lambda_0}(X).$$

Using C2, C3 and C4, it can be easily shown that $\|m_i^r(X) - m_i^l(X)\|_{P,B}^2 \leq C\epsilon^2$, where $\|\cdot\|_{P,B}$ is the "Bernstein norm" defined to be $\|f\|_{P,B} = \left\{2P\left(e^{|f|} - 1 - |f|\right)\right\}^{1/2}$. (see var der Vaart and Wellner, 1996, p324)

Moreover, for any $m_\Lambda(X) - m_{\Lambda_0}(X) \in \mathcal{M}_\delta$, the same techniques as used for arguing the preceding statement yield that $\|m_{\Lambda(X)} - m_{\Lambda_0}(X)\|_{P,B}^2 \leq C\delta^2$. Hence, by Lemma 3.4.3 of van der Vaart and Wellner (1996), we have

$$(6.1) \qquad E_P\|\sqrt{n}(\mathbb{P}_n - P)\|_{\mathcal{M}_\delta} \leq C\tilde{J}_{[\,]}(\delta, \mathcal{M}_\delta, \|\ \ \|_{P,B})\left\{1 + \frac{\tilde{J}_{[\,]}(\delta, \mathcal{M}_\delta, \|\ \ \|_{P,B})}{\delta^2\sqrt{n}}\right\},$$

where

$$\tilde{J}_{[\,]}(\delta, \mathcal{M}_\delta, \|\ \ \|_{P,B}) = \int_0^\delta \sqrt{1 + \log N_{[\,]}(\epsilon, \mathcal{M}_\delta, \|\ \ \|_{P,B})} d\epsilon \leq C\delta^{1/2}.$$

This yields the right hand side of (6.1), $\phi_n(\delta) \leq C\left(\delta^{1/2} + \delta^{-1}/\sqrt{n}\right)$. It is easy to see that $\phi_n(\delta)/\delta$ is a decreasing function of δ and $n^{2/3}\phi_n(n^{-1/3}) = n^{2/3}\left(n^{-1/6} + n^{1/3}n^{-1/2}\right) = 2\sqrt{n}$.

This suffices to conclude $n^{1/3}d(\hat{\Lambda}_n, \Lambda_0) = O_p(1)$ by Theorem 3.2.5 of van der Vaart and Wellner (1996).

Proof of Theorem 2.2: The proof of this theorem is closely related to the arguments used in Huang and Wellner (1995). First, we rewrite Equation (2.5) in Wellner and Zhang (2000) as $\sum_{i=1}^{n}\sum_{j=1}^{K_i}\left[\mathbb{N}^{(i)}(T_{K_i,j}) - \hat{\Lambda}_n(T_{K_i,j})\right] = 0$. Using the same block argument as originally described in Groeneboom and Wellner (1992), p43, we have

$$\mathbb{P}_n\left[\sum_{j=1}^{K}\left\{\mathbb{N}(T_{K,j}) - \hat{\Lambda}_n(T_{K,j})\right\}\eta\circ\Lambda_0^{-1}\left\{\hat{\Lambda}_n(T_{K,j})\right\}\right] = 0. \tag{6.2}$$

for any function η.

Second, we can express

$$\begin{aligned} n^{1/2}\left\{\nu(\hat{\Lambda}_n) - \nu(\Lambda_0)\right\} &= -n^{1/2}P\left[\sum_{j=1}^{K}\left\{\mathbb{N}(T_{K,j}) - \hat{\Lambda}_n(T_{K,j})\right\}\eta(T_{K,j})\right] \\ &= \Delta_{1n} + \Delta_{2n} + \Delta_{3n} + \Delta_{4n}, \end{aligned} \tag{6.3}$$

where

$$\begin{aligned} \Delta_{1n} &= n^{1/2}(\mathbb{P}_n - P)\left[\textstyle\sum_{j=1}^{K}\{\mathbb{N}(T_{K,j}) - \Lambda_0(T_{K,j})\}\eta(T_{K,j})\right] \\ \Delta_{2n} &= -n^{1/2}(\mathbb{P}_n - P)\left[\textstyle\sum_{j=1}^{K}\left\{\hat{\Lambda}_n(T_{K,j}) - \Lambda_0(T_{K,j})\right\}\eta(T_{K,j})\right] \\ \Delta_{3n} &= n^{1/2}(\mathbb{P}_n - P)\left(\textstyle\sum_{j=1}^{K}\left\{\mathbb{N}(T_{K,j}) - \hat{\Lambda}_n(T_{K,j})\right\}\left[\eta\circ\Lambda_0^{-1}\left\{\hat{\Lambda}_n(T_{K,j})\right\} - \eta\circ\Lambda_0^{-1}\{\Lambda_0(T_{K,j})\}\right]\right). \\ \Delta_{4n} &= -n^{1/2}P\left(\textstyle\sum_{j=1}^{K}\left\{\mathbb{N}(T_{K,j}) - \hat{\Lambda}_n(T_{K,j})\right\}\left[\eta\circ\Lambda_0^{-1}\{\Lambda_0(T_{K,j})\} - \eta\circ\Lambda_0^{-1}\left\{\hat{\Lambda}_n(T_{K,j})\right\}\right]\right) \end{aligned}$$

By the Central Limit Theorem, we have that $\Delta_{1n} \to_d N(0, \Omega)$. By Theorem 2.1 and C5, we have that

$$\begin{aligned} |\Delta_{4n}| &= n^{1/2}\left|\int\left\{\Lambda_0(t) - \hat{\Lambda}_n(t)\right\}\left[\eta\circ\Lambda_0^{-1}\{\Lambda_0(t)\} - \eta\circ\Lambda_0^{-1}\left\{\hat{\Lambda}_n(t)\right\}\right]d\mu(t)\right| \\ &\le Cn^{1/2}\int\left\{\Lambda_0(t) - \hat{\Lambda}_n(t)\right\}^2 d\mu(t) \to_p 0 \end{aligned}$$

Hence, to prove the theorem, it suffices to show that both Δ_{2n} and Δ_{3n} are $o_p(1)$.

Third, to show $\Delta_{2n} = o_p(1)$, we denote $\phi_1(\Lambda) = \sum_{j=1}^{K}\{\Lambda(T_{K,j}) - \Lambda_0(T_{K,j})\}\eta(T_{K,j})$ and define a class, $\Phi_1(\delta) = \{\phi_1(\Lambda) : \ \Lambda \in \mathcal{F}_\delta\}$. Using the set of brackets for $\mathcal{F}_\delta$ constructed in the proof of Theorem 2.1, we can form a set of brackets for $\Phi_1(\delta)$, $\left\{[\phi_{1i}^l, \phi_{1i}^r]: \quad i = 1, 2, \cdots, q\right\}$ by letting

$$\begin{aligned} \phi_{1i}^l(X) &= \textstyle\sum_{j=1}^K \left\{\Lambda_i^l(T_{K,j}) - \Lambda_0(T_{K,j})\right\} \eta(T_{K,j}) \\ \phi_{1i}^r(X) &= \textstyle\sum_{j=1}^K \left\{\Lambda_i^r(T_{K,j}) - \Lambda_0(T_{K,j})\right\} \eta(T_{K,j}). \end{aligned}$$

A simple algebraic calculation yields that $P\left\{\phi_{1i}^r(X) - \phi_{1i}^l(X)\right\}^2 \le C\epsilon^2$ by the boundedness of η. This indicates that the total number of ϵ-brackets associated with $L_2(P)$ norm for $\Phi_1(\delta)$ will be in the order of $\exp(C/\epsilon)$ and hence $\Phi_1(\delta)$ is a P-Donsker class for any $\delta > 0$. Similarly, for any $\phi_1 \in \Phi_1(\delta)$, it is easy to see that

$$P\phi_1^2(X) \le C \int \{\Lambda(t) - \Lambda_0(t)\}^2 d\mu(t) \le C\delta^2 \to 0 \quad \text{as} \quad \delta \to 0.$$

Therefore, by the uniformly asymptotic equicontinuity of the empirical process resulting from a Donsker property (Corollary 2.3.12 of van der Vaart and Wellner, 1996), we can conclude that $\Delta_{2n} = o_p(1)$.

Finally, to show $\Delta_{3n} = o_p(1)$, we denote

$$\phi_2(\Lambda) = \sum_{j=1}^K \{\mathbb{N}(T_{k,j}) - \Lambda(T_{K,j})\} \left[\eta \circ \Lambda_0^{-1}\{\Lambda(T_{K,j})\} - \eta \circ \Lambda_0^{-1}\{\Lambda_0(T_{K,j})\}\right]$$

and define a class $\Phi_2(\delta) = \{\phi_2(\Lambda): \quad \Lambda \in \mathcal{F}_\delta\}$.

With the same bracket set for $\mathcal{F}_\delta$, we can easily form a bracket set for $\Phi_2(\delta)$, $\{[\phi_{2i}^l, \phi_{2i}^r]: i = 1, 2, \cdots, q\}$ by letting

$$\begin{aligned} \phi_{2i}^l(X) = \textstyle\sum_{j=1}^K \big[&\mathbb{N}(T_{K,j})\eta \circ \Lambda_0^{-1}\{\Lambda_i^l(T_{K,j})\} + \Lambda_i^l(T_{K,j})\eta \circ \Lambda_0^{-1}\{\Lambda_0(T_{K,j})\} \\ &- \Lambda_i^r(T_{K,j})\eta \circ \Lambda_0^{-1}\{\Lambda_i^r(T_{K,j})\} - \mathbb{N}(T_{K,j})\eta \circ \Lambda_0^{-1}\{\Lambda_0(T_{K,j})\}\big] \\ \phi_{2i}^r(X) = \textstyle\sum_{j=1}^K \big[&\mathbb{N}(T_{K,j})\eta \circ \Lambda_0^{-1}\{\Lambda_i^r(T_{K,j})\} + \Lambda_i^r(T_{K,j})\eta \circ \Lambda_0^{-1}\{\Lambda_0(T_{K,j})\} \\ &- \Lambda_i^l(T_{K,j})\eta \circ \Lambda_0^{-1}\{\Lambda_i^l(T_{K,j})\} - \mathbb{N}(T_{K,j})\eta \circ \Lambda_0^{-1}\{\Lambda_0(T_{K,j})\}\big]. \end{aligned}$$

By C3 and C5, we can show through a simple algebraic calculation that $P\left\{\phi_{2i}^r(X) - \phi_{2i}^l(X)\right\}^2 \le C\epsilon^2$. So $\Phi_2(\delta)$ is a P-Donsker class. We can also similarly show that for for any $\phi_2(\Lambda) \in \Phi_2(\delta)$, $P\phi_2^2(\Lambda) \le \int \{\Lambda(t) - \Lambda_0(t)\}^2 d\mu(t) = C\delta^2$ by C3 and C5. Hence $\Delta_{3n} = o_p(1)$ by Corollary 2.3.12 of van der Vaart and Wellner (1996) again and the proof is complete.

Proof of Theorem 3.1: First, we express

$$\nu_n\left(\hat{\Lambda}_{n_1}, \hat{\Lambda}_{n_2}\right) \quad = \quad \sqrt{\frac{n_2}{n}} n_1^{1/2} \left\{\nu(\hat{\Lambda}_{n_1}) - \nu(\Lambda_0)\right\} - \sqrt{\frac{n_1}{n}} n_2^{1/2} \left(\nu(\hat{\Lambda}_{n_2}) - \nu(\Lambda_0)\right)$$

$$(6.4) \qquad +\sqrt{\frac{n_1 n_2}{n^2}} A_n^{(1)} + \sqrt{\frac{n_1 n_2}{n^2}} A_n^{(2)} + \sqrt{\frac{n_1 n_2}{n^2}} A_n^{(3)},$$

where

$$\begin{aligned}
A_n^{(1)} &= n^{1/2} \int \left\{ \hat{\Lambda}_{n_1}(t) - \hat{\Lambda}_{n_2}(t) \right\} \{\hat{\eta}_n(t) - \eta(t)\} \, d\mu(t) \\
A_n^{(2)} &= n^{1/2} \left(\mathbb{P}_n - P \right) \left[\sum_{j=1}^K \left\{ \hat{\Lambda}_{n_1}(T_{K,j}) - \hat{\Lambda}_{n_2}(T_{k,j}) \right\} \eta(T_{K,j}) \right] \\
A_n^{(3)} &= n^{1/2} \left(\mathbb{P}_n - P \right) \left[\sum_{j=1}^K \left\{ \hat{\Lambda}_{n_1}(T_{K,j}) - \hat{\Lambda}_{n_2}(T_{k,j}) \right\} \{\hat{\eta}_n(T_{K,j}) - \eta(T_{K,j})\} \right].
\end{aligned}$$

Since $\hat{\Lambda}_{n_1}$ and $\hat{\Lambda}_{n_2}$ are obtained from two independent samples, the difference of the first two terms in (6.4) is asymptotically equivalent to $\sqrt{p_2} W_1 - \sqrt{p_1} W_2$ with W_1 and W_2 drawn independently from $N(0, \Omega)$ by Theorem 2.1 and Assumptions (i) and (iii).

Next, we show that $A_n^{(1)}$, $A_n^{(2)}$, and $A_n^{(3)}$ are all $o_p(1)$. Note that, using the Cauchy-Schwarz inequality along with the result of Theorem 2.1 and Assumption (ii), we have

$$\left| A_n^{(1)} \right|^2 \leq n \int \left\{ \hat{\Lambda}_{n_1}(t) - \hat{\Lambda}_{n_2}(t) \right\}^2 d\mu(t) \int \{\hat{\eta}_n(t) - \eta(t)\}^2 \, d\mu(t) = n O_p(n^{-2/3}) o_p(n^{-1/3}) = o_p(1)$$

To verify other two terms, we let

$$\begin{aligned}
\psi_1(\Lambda_1, \Lambda_2) &= \sum_{j=1}^K \{\Lambda_1(T_{K,j}) - \Lambda_2(T_{K,j})\} \, \eta_0(T_{K,j}) \\
\psi_2(\Lambda_1, \Lambda_2, \eta) &= \sum_{j=1}^K \{\Lambda_1(T_{K,j}) - \Lambda_2(T_{K,j})\} \{\eta(T_{K,j}) - \eta_0(T_{K,j})\}
\end{aligned}$$

and define two classes

$$\begin{aligned}
\Psi_1(\delta) &= \{\psi_1(\Lambda_1, \Lambda_2) : \ \ \Lambda_1, \Lambda_2 \in \mathcal{F}_\delta\} \\
\Psi_2(\delta) &= \{\psi_1(\Lambda_1, \Lambda_2, \eta) : \ \ \Lambda_1, \Lambda_2 \in \mathcal{F}_\delta \ \text{ and } \ d(\eta, \eta_0) \leq \delta\}.
\end{aligned}$$

Because we consider η to be a bounded monotone function, we can similarly argue that both $\Psi_1(\delta)$ and $\Psi_2(\delta)$ are P-Donsker for any $\delta > 0$ based on the same bracketing number arguments as used in the proof of Theorem 2.2. Moreover, as $\delta \to 0$, we can also similarly show that

$$\begin{aligned}
P\psi_1^2(\Lambda_1, \Lambda_2) &\to 0 \ \text{ for any } \ \psi_1 \in \Psi_1(\delta) \\
P\psi_2^2(\Lambda_1, \Lambda_2, \eta) &\to 0 \ \text{ for any } \ \psi_2 \in \Psi_2(\delta).
\end{aligned}$$

Hence, $A_n^{(2)} = o_p(1)$ and $A_n^{(3)} = o_p(1)$ are established according to Corollary 2.3.12 of van der Vaart and Wellner (1996).

REFERENCES

Andersen, P. K. and Gill, R. D. (1982). "Cox's regression model for counting processes: a large sample study," *The Annals of Statistics*, **10**, 1100–1120.

Andrews, D. F. and Herzberg, A. M. (1985). *Data; A Collection of Problems from Many Fields for the Student and Research Worker.* Springer-Verlag, New York.

Byar, D. P. (1980). "The Veterans Administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparisons of placebo, pyridoxine, and topical thiotepa." In *Bladder Tumors and Other Topics in Urological Oncology*, pp. 363 - 370, edited by M. pavone-Macaluso, P. H. Smith, and F. Edsmyr. Plenum, New York

Byar, D. P., Blackard, C. and the VACURG. (1977). "Comparisons of placebo, pyridoxine, and topical thiotepa in preventing recurrence of stage I bladder cancer." *Urology* **10**, 556 - 561.

Cook, R. J., Lawless, J. F., and Nadeau, C. (1996). "Robust tests for treatment comparisons based on recurrent event responses," *Biometrics*, **52**, 557 - 571.

Gaver, D. P., and O'Muircheartaigh, I. G. (1987). "Robust Empirical Bayes analysis of event rates," *Technometrics*, **29**, 1-15.

Groeneboom, P. and Wellner, J. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* Birkhäuser, Basel.

Hu, X.J., Sun, J., and Wei, L.J. (2003). "Regression analysis of panel count data." *Scandinavian Journal of Statistics* **25**, 25-43.

Huang, J. (1996). "Efficient estimation for the Cox model with interval censoring," *Ann. Statist.*, **24**, 540-568.

Huang, J. and Wellner, J.A. (1995). "Asymptotic normality of the NPMLE of linear functionl for interval censored data, case 1." *Statist. Neer.*, **49**, 153-163.

Kalbfleisch, J. D. and Lawless, J. F. (1985). "The analysis of panel count data under a Markov assumption," *J. Amer. Statist. Assoc.*, **80**, 863 - 871.

Lawless, J. F. and Nadeau, C. (1995). "Some simple robust methods for the analysis of recurrent events," *Technometrics*, **37**, 158 - 168.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000). "Semiparametric regression for the mean and rate functions of recurrent events," *J. Roy. Statist. Soc. B*, 711-730.

Pepe, M. S. and Cai, J. (1993). "Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates," *Journal of the American Statistical Association*, **88**, 811-820.

Prentice R. L., Williams, B. J., and Peterson, A. V. (1981). "On the regression analysis of multivariate failure time data," *Biometrika*, **68**, 373-379.

Schick, A. and Yu, Q. (2000). "Consistency of the GMLE with mixed case interval-censored data," *Scand. J. Statist.* **27**, 45-55.

Sun, J. and Kalbfleisch, J. D. (1995). "Estimation of the mean function of point processes based on panel count data," *Statistica Sinica*, **5**, 279 - 290.

Sun, J. and Wei, L. J. (2000). "Regression analysis of panel count data with covariate-dependent observation and censoring times," *J. R. Stat. Soc. Ser. B*, **62**, 293 - 302.

Sun, J. and Fang, H-B. (2003). "A nonparametric test for panel count data," *Biometrika*, **90**, 199-208.

Thall, P. F., and Lachin, J. M. (1988). "Analysis of Recurrent Events: Nonparametric Methods for Random-Interval Count Data," *J. Amer. Statist. Assoc.*, **83**, 339-347.

Thall, P. F. (1988). "Mixed Poisson likelihood regression models for longitudinal interval count data," *Biometrics*, **44**, 197-209.

Van der Vaart, A.W., and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

Wei, L. J., Lin, D.Y., and Weissfeld, L. (1989). "Regression analysis of multivariate incomplete failure time data by modeling marginal distributions." *J. Amer. Statist. Assoc.* **84**, 1065 - 1073.

WELLNER, J. A. AND ZHANG, Y. (2000). "Two estimators of the mean of a counting process with panel count data," *Ann. Statist.*, **28**, 779 - 814.

WELLNER, J. A. AND ZHANG, Y. (2005). "Likelihood-based semiparametric estimation methods for panel count data," *Technical Report* **488**, Department of Statistics, University of Washington.

WELLNER, J. A., ZHANG, Y., AND LIU, H. (2004). "A semiparametric regression model for panel count data: when do pseudo-likelihood estimators become badly inefficient?" *Proceedings of the Second Seattle Biostatistical Symposium: Analysis of Correlated Data.* (Eds. Lin, D.Y. and P.J., Heagerty). Springer-Verlag, New York, 143-174.

ZHANG, Y. (2002). "A semiparametric pseudo likelihood estimation method for panel count data," *Biometrika*, **89**, 39-48.

ZHANG, Y., LIU, W., AND ZHAN Y-H. (2001). "A nonparametric two-sample test of the failure function with interval censoring case 2," *Biometrika*, **88**, 677-686.

Table 1: *Monte-Carlo simulation results on the percentage of rejection at 5% significance level based on 1000 replications for Scenario 1 with* $\lambda = 1$.

			$n_1 = n_2 = n_3$=50			$n_1 = n_2 = n_3$=100		
Case			$T_n^{(1)}$	$T_n^{(2)}$	$T_n^{(3)}$	$T_n^{(1)}$	$T_n^{(2)}$	$T_n^{(3)}$
	Under H_0							
1	$\theta_1 = 0$	$\theta_2 = 0$	$0 \cdot 061$	$0 \cdot 067$	$0 \cdot 064$	$0 \cdot 063$	$0 \cdot 062$	$0 \cdot 060$
	Under H_a							
2	$\theta_1 = 0$	$\theta_2 = 0 \cdot 10$	$0 \cdot 229$	$0 \cdot 205$	$0 \cdot 229$	$0 \cdot 397$	$0 \cdot 323$	$0 \cdot 370$
3	$\theta_1 = 0$	$\theta_2 = 0 \cdot 25$	$0 \cdot 852$	$0 \cdot 772$	$0 \cdot 835$	$0 \cdot 986$	$0 \cdot 964$	$0 \cdot 984$
4	$\theta_1 = 0$	$\theta_2 = 0 \cdot 50$	$1 \cdot 000$	$1 \cdot 000$	$1 \cdot 000$	$1 \cdot 000$	$1 \cdot 000$	$1 \cdot 000$
5	$\theta_1 = 0 \cdot 10$	$\theta_2 = 0 \cdot 10$	$0 \cdot 219$	$0 \cdot 184$	$0 \cdot 227$	$0 \cdot 380$	$0 \cdot 330$	$0 \cdot 366$
6	$\theta_1 = 0 \cdot 10$	$\theta_2 = 0 \cdot 25$	$0 \cdot 721$	$0 \cdot 647$	$0 \cdot 696$	$0 \cdot 961$	$0 \cdot 918$	$0 \cdot 945$
7	$\theta_1 = 0 \cdot 10$	$\theta_2 = 0 \cdot 50$	$0 \cdot 998$	$0 \cdot 995$	$1 \cdot 000$	$1 \cdot 000$	$1 \cdot 000$	$1 \cdot 000$
8	$\theta_1 = 0 \cdot 25$	$\theta_2 = 0 \cdot 25$	$0 \cdot 843$	$0 \cdot 781$	$0 \cdot 825$	$0 \cdot 984$	$0 \cdot 963$	$0 \cdot 982$
9	$\theta_1 = 0 \cdot 25$	$\theta_2 = 0 \cdot 50$	$1 \cdot 000$	$0 \cdot 994$	$1 \cdot 000$	$1 \cdot 000$	$1 \cdot 000$	$1 \cdot 000$
10	$\theta_1 = 0 \cdot 50$	$\theta_2 = 0 \cdot 50$	$1 \cdot 000$	$0 \cdot 997$	$0 \cdot 999$	$1 \cdot 000$	$1 \cdot 000$	$1 \cdot 000$

Table 2: *Monte-Carlo simulation results on the percentage of rejection at 5% significance level based on 1000 replications for Scenario 2 with $\lambda = 0 \cdot 2$.*

			$n_1 = n_2 = n_3$=50			$n_1 = n_2 = n_3$=100		
Case			$T_n^{(1)}$	$T_n^{(2)}$	$T_n^{(3)}$	$T_n^{(1)}$	$T_n^{(2)}$	$T_n^{(3)}$
	Under H_0							
1	$\theta_1 = 0$	$\theta_2 = 0$	$0 \cdot 066$	$0 \cdot 063$	$0 \cdot 047$	$0 \cdot 063$	$0 \cdot 060$	$0 \cdot 042$
	Under H_a							
2	$\theta_1 = 0$	$\theta_2 = 0 \cdot 25$	$0 \cdot 161$	$0 \cdot 165$	$0 \cdot 094$	$0 \cdot 256$	$0 \cdot 251$	$0 \cdot 169$
3	$\theta_1 = 0$	$\theta_2 = 0 \cdot 50$	$0 \cdot 352$	$0 \cdot 345$	$0 \cdot 240$	$0 \cdot 630$	$0 \cdot 600$	$0 \cdot 515$
4	$\theta_1 = 0$	$\theta_2 = 1 \cdot 00$	$0 \cdot 804$	$0 \cdot 791$	$0 \cdot 594$	$0 \cdot 988$	$0 \cdot 983$	$0 \cdot 945$
5	$\theta_1 = 0 \cdot 25$	$\theta_2 = 0 \cdot 25$	$0 \cdot 144$	$0 \cdot 139$	$0 \cdot 105$	$0 \cdot 237$	$0 \cdot 214$	$0 \cdot 187$
6	$\theta_1 = 0 \cdot 25$	$\theta_2 = 0 \cdot 50$	$0 \cdot 276$	$0 \cdot 256$	$0 \cdot 194$	$0 \cdot 503$	$0 \cdot 460$	$0 \cdot 386$
7	$\theta_1 = 0 \cdot 25$	$\theta_2 = 1 \cdot 00$	$0 \cdot 720$	$0 \cdot 695$	$0 \cdot 506$	$0 \cdot 941$	$0 \cdot 919$	$0 \cdot 855$
8	$\theta_1 = 0 \cdot 50$	$\theta_2 = 0 \cdot 50$	$0 \cdot 363$	$0 \cdot 332$	$0 \cdot 269$	$0 \cdot 648$	$0 \cdot 606$	$0 \cdot 543$
9	$\theta_1 = 0 \cdot 50$	$\theta_2 = 1 \cdot 00$	$0 \cdot 689$	$0 \cdot 659$	$0 \cdot 537$	$0 \cdot 946$	$0 \cdot 921$	$0 \cdot 893$
10	$\theta_1 = 1 \cdot 00$	$\theta_2 = 1 \cdot 00$	$0 \cdot 844$	$0 \cdot 810$	$0 \cdot 715$	$0 \cdot 985$	$0 \cdot 978$	$0 \cdot 964$

Table 3: *Monte-Carlo simulation results on the percentage of rejection at 5% significance level based on 1000 replications for Scenario 3.*

		$n_1 = n_2 = n_3$=50			$n_1 = n_2 = n_3$=100		
Case		$T_n^{(1)}$	$T_n^{(2)}$	$T_n^{(3)}$	$T_n^{(1)}$	$T_n^{(2)}$	$T_n^{(3)}$
1	$a = 0 \cdot 75$	$0 \cdot 095$	$0 \cdot 101$	$0 \cdot 025$	$0 \cdot 099$	$0 \cdot 104$	$0 \cdot 023$
2	$a = 0 \cdot 50$	$0 \cdot 090$	$0 \cdot 104$	$0 \cdot 012$	$0 \cdot 086$	$0 \cdot 103$	$0 \cdot 044$
3	$a = 0 \cdot 25$	$0 \cdot 160$	$0 \cdot 124$	$0 \cdot 256$	$0 \cdot 279$	$0 \cdot 152$	$0 \cdot 835$

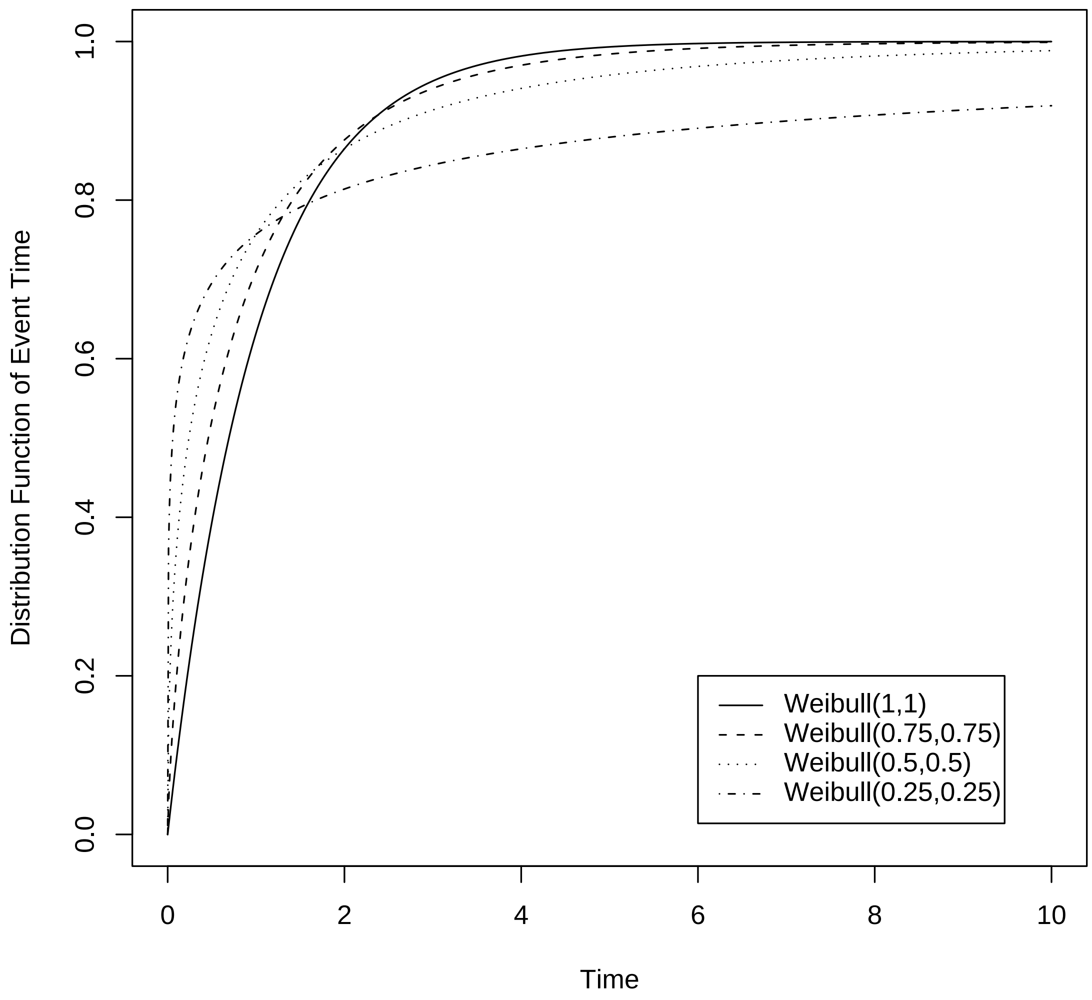

Figure 1: The distribution functions of several Weibull random variables

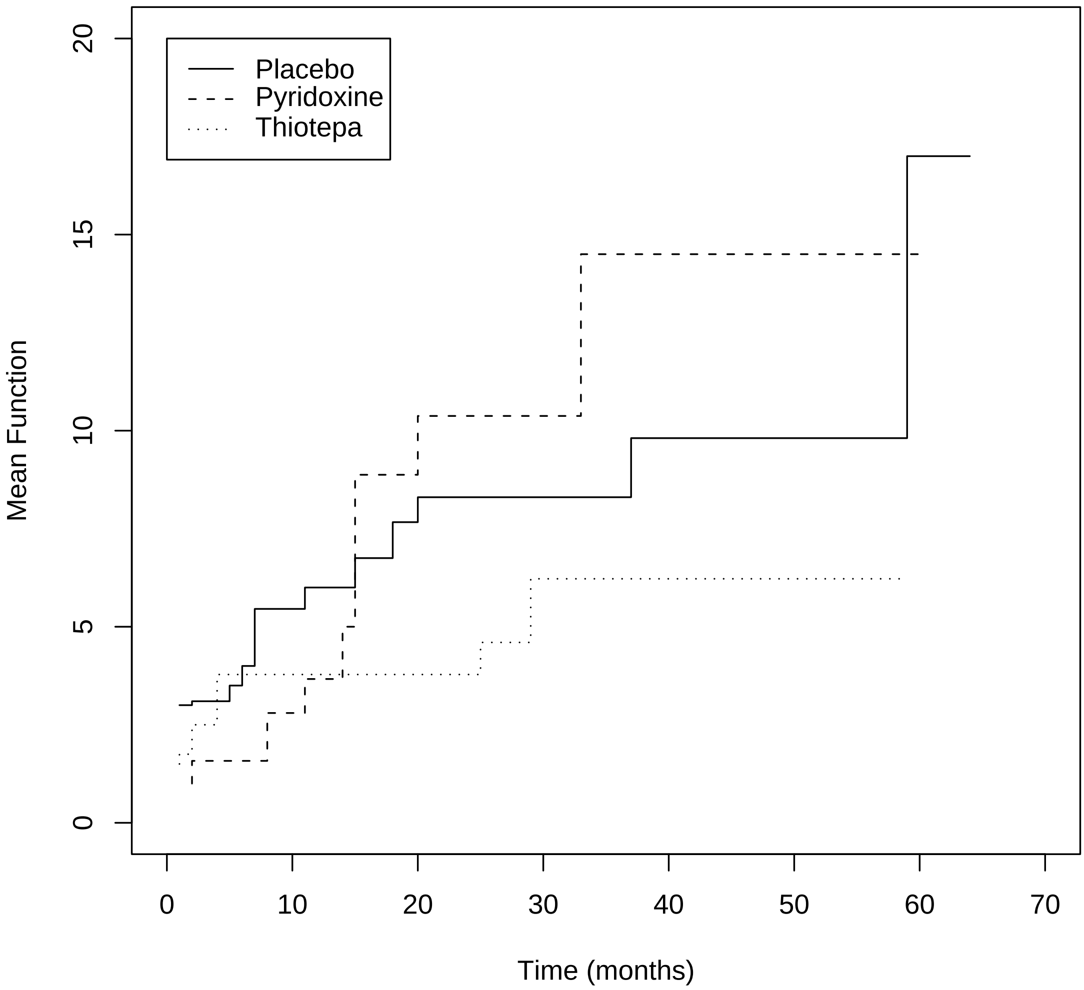

Figure 2: The nonparametric maximum pseudo-likelihood estimators of the mean function of bladder tumor counts under the three treatments

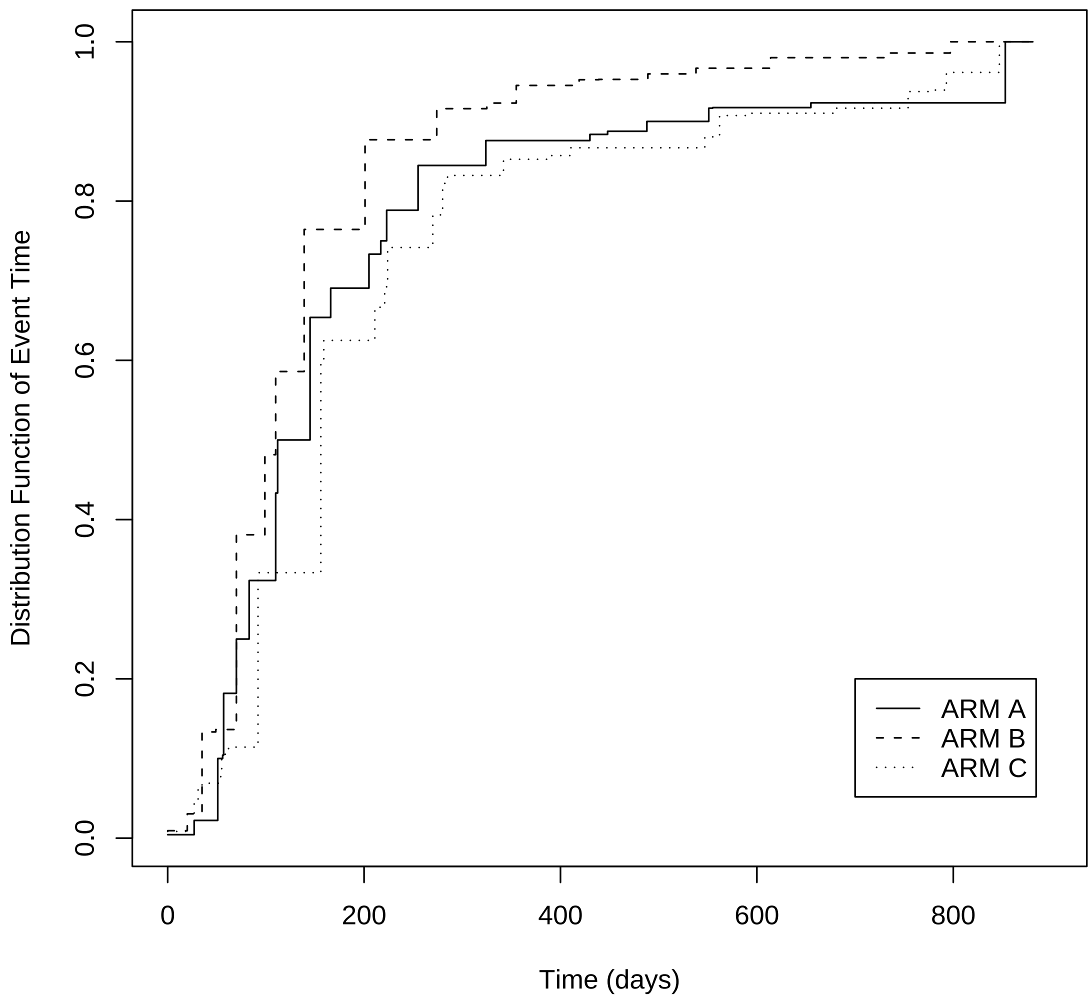

Figure 3: The nonparametric maximum pseudo-likelihood estimators of the cumulative distribution function of time-to-response of antiviral therapies under the three study arms