The Random Partition Masking Model For Interval-Censored And Masked Competing Risks Data

Jiaping Wang[1], Qiqing Yu[2] and George Y. C. Wong[3]

1. Department of Biostatistics, UNC at Chapel Hill, NC 27514, USA

2. Department of Mathematical Science, SUNY at Binghamton, NY 13902, USA

3. Strang Cancer Prevention Center, 428 E 72nd Street, NY 10021, USA

Current Version: 02/10/2011

Short Title: Interval Censored Data and RPM Model

Key Words: Interval-censored, masked competing risks data, self-consistent algorithm, nonparametric MLE, maximal intersections, RPM model.

Summary: We study the nonparametric maximum likelihood estimate (NPMLE) of the cdf or sub - distribution functions of the failure time for the failure causes in a series system. The study is motivated by a cancer research data (from the Memorial Sloan-Kettering Cancer Center) with interval-censored time and masked failure cause. The NPMLE based on this data set suggests that the existing masking models are not appropriate. We propose a new model called the random partition masking (RPM) model, which does not rely on the commonly used symmetry assumption (namely, given the failure cause, the probability of observing the masked failure causes is independent of the failure time) (see Flehinger *et al.* (1996)). The RPM model is easier to implement in simulation studies than the existing models. We discuss the algorithms for computing the NPMLE and study its asymptotic properties. Our simulation and data analysis indicate that the NPMLE is feasible for a moderate sample size.

§**1. Introduction.** We study the nonparametric maximum likelihood estimate (NPMLE) of the sub-distribution function (scdf) of the failure time T for the failure cause C $F^s_{j0}(t) = P(T \le t, C = j)$ (or equivalently the joint cumulative distribution function (cdf) $F_{T,C}(t, c) = P(T \le t, C \le c)$) in a J-component series system when T is subject to interval censoring and C is subject to masking. A series system means it stops functioning as soon as one of its J constituent components fails. The competing risks (CR) observation on (T, C) can be described as follows. Let the random variable X_j denote the lifetime of the j^{th} component, $j = 1, ..., J$, by the definition of a series system, $T = \min\{X_1, ..., X_J\}$. It is assumed that the probability of a system failure due to simultaneous failures of two or more distinct components is 0, thus there exists a unique positive integer C associated with T, say $X_C = T$.

The study is motivated by a cancer research data set (from the Memorial Sloan-Kettering Cancer Center between 1985 and 1990) that we are analyzing. In this study, 375 women with stages I - III unilateral invasive breast cancer are surgically treated and followed periodically. Here T is the time to relapse of cancer and C is the type of cancer at relapse, which can be either one of the total seven (i.e., $J = 7$). When the relapse is diagnosed at a follow-up time for a patient, there may be one or more types of cancer cells detected. In such cases, sometimes it is impossible to determine when the cancer occurs and which type of cancer relapses first. Thus the relapse time T is only known to lie within two consecutive follow-up time points, say L and R ($L < T \le R$), that means it is interval censored or right censored by the end of the follow-up. Moreover, the type of cancer that occurs first is only known to belong to $\mathcal{M}$, the set of all types of cancer cells detected at the follow-up time. Then C is said to be masked by $\mathcal{M}$ (called the minimum random set (MRS) in the literature, see Guess *et al.* (1991)). There are no second stage data in our breast cancer data set. In fact, in this cancer study, further investigation of $\mathcal{M}$ would not give a clue on which type of cancer occurred first. We call such data interval-censored (IC) and masked competing risks (ICMCR) data. Moreover, if T is right censored, it is assumed that "its MRS is unknown or missing" (Mukhopadhyay (2006, p.806[13])). Thus the ICMCR observation has the form $(L, R, \mathcal{M}^o)$, where $\mathcal{M}^o = \mathcal{M}$ if T is not right censored and $\mathcal{M}^o = \mathcal{C}_r$ $(= \{1, 2, \cdots, J\})$ otherwise.

The study of the estimation of the cdf or scdf in the competing risk literature can date back to 1970's (*e.g.*, Peterson (1977)). In the current literature, most researchers study the inference on the right-censored (RC) competing risks data in the parametrical context (see Miyakawa (1984), Guess *et al.* (1991), Reiser *et al.* (1995), Lin *et al.* (1996), Mukhopadhyay and Basu (1997, 2007), Sen *et al.* (2001), Flehinger *et al.* (2001), Basu *et al.* (1999), Dewanji and Sengupta (2003), and Craiu and Reiser (2006) etc.). They do not work on the masked failure cause and use the EM algorithm or Bayesian approach to find the estimators. Dinse (1982) and (1986) study the nonparametric estimation of $F_{T,C}$ with RC competing risks data. Flehinger *et al.* (1998) provide a semi-parametric approach with masked causes of failure. Some people study the inference on the RC data with masked causes of failure (see *e.g.*, Craiu and Reiser (2006)) or with the second stage data (see *e.g.*, Craiu and Duchesne (2004)). Not so many people study estimation problem with the interval-censored competing risks data. Groeneboom *et al.* (2008b) as well as Hudgens *et al.* (2001) study the NPMLE with IC data but without masking on C. Basu *et al.* (2003) study the IC data with masked causes of failure in the parametric context using Bayesian approach and Gibbs sampling in which they assume the competing risks are independent. Maathuis and Wellner (2008) relate their bivariate estimation problem to estimation problem with the competing risk data without masking. The nonparametric estimation of the scdf with ICMCR data has not been studied in the literature.

Some researchers in the CR data literature consider estimating the cause-specific hazard functions (see Dewanji and Sengupta (2003), Crau and Duchesne (2004)). Notice that if the hazard function exists, it is equivalent to the cdf (or scdf). While the hazard function may not exist, the cdf or scdf always exists.

The nonparametric estimation of the scdf with ICMCR data is quite different from the univariate censored data. The data studied in Hudgens *et al.* (2001) and Groeneboom *et al.* (2008b) do not involve masking. Even though Dinse (1982) proposes the NPMLE with right-censored and masked CR data long time ago, its asymptotic properties has not been established due to the technical difficulties.

One may think the nonparametric estimation of the scdf with the ICMCR data is a special case of the bivariate IC data since they both estimate the bivariate distribution functions. In fact, it is the opposite. In some sense, the non-parametric estimation with bivariate IC data is a special case of the one with ICMCR data. The reason is as follows. In both cases the NPMLE puts weights only to the maximum intersections (MIs, see the definition in Example 1.1), but the MIs with bivariate IC data are of the form of one rectangle set, while the MIs with ICMCR data are of the form of a rectangle or the union of several disjoint rectangles, as explained in the following example.

Example 1.1. First consider a bivariate IC data set with $(L_{1i}, R_{1i}] \times (L_{2i}, R_{2i}]$ (denoted by I_is): $(2,5]\times(1,6]$, $(1,3]\times(1,3]$, $(4,7]\times(1,4]$, $(2,3]\times(5,7]$. A MI B is a non-empty intersection of some I_is such that $I_i \cap B$ either equals B or $\emptyset$ (the empty set) for each I_i. For these 4 data, there are 3 MIs: $\begin{matrix} (2,3]\times(5,6] & \\ (2,3]\times(1,3] & (4,5]\times(1,4] \end{matrix}$. Even though we arrange them as a matrix, They are all of the form of a single rectangle.

Next consider a set of ICMCR data with $J = 4$. $(L_i, R_i] \times \mathcal{M}_i$s (denoted by I_is) are $(1,3]\times\{1,3,4\}$, $(1,2]\times\mathcal{C}_r$, $(2,5]\times\{1,2,4\}$ and $(1,6]\times\{2\}$. For this data set, there are 3 MIs: $(1,2]\times\{1,3,4\}$, $(2,3]\times\{1,4\}$, $(1,2]\times\{2\}$ and $(2,5]\times\{2\}$. Only the last two MIs $(1,2]\times\{2\}$ and $(2,5]\times\{2\}$ are of the form of a rectangle, like the case in the bivariate IC data. If all MIs in an ICMCR data happen to be of the form of a rectangle, the estimation reduces to the bivariate IC data. Notice that for competing risks data without masking, each MI is of the form of a single rectangle as in the bivariate IC data.

It is well known that with censored data one needs to find all the MI's before computing the NPMLE. In view of Example 1.1, since the MIs with ICMCR data are more complicated than the MIs with bivariate IC data, the algorithm for finding all MIs for ICMCR data must be different from the algorithm for finding all MIs for bivariate IC data such as the algorithm HeightMap of Maathuis (2005) and the support reduction algorithm of Groeneboom *et al.* (2008a). The latter algorithms can't be applied to the ICMCR data. Thus we need develop new algorithms to find the maximum intersections and the NPMLE. It is not surprised that the NPMLE of the scdf with ICMCR data and its asymptotic properties have not been investigated so far.

Nonparametric estimation of the scdf is quite important in analyzing real data. First, it is distribution free and is appropriate if it is difficult to find a suitable parametric model to fit the data. Secondly, if one tries to apply a parametric model to a data set, the parametric assumption needs to be validated by a nonparametric estimator. Otherwise, the statistical analysis based on the parametric assumption is nonsense.

The rest of this paper is organized as follows. In §2, we introduce the NPMLE with its algorithm, the random partition masking (RPM) model for ICMCR data and the related assumptions. Since the ICMCR data are first studied in our paper, this model is the first model ever proposed for the ICMCR data in the literature. One can construct a different model (called the Conditional masking probability (CMP) model) which relies on the commonly used symmetry assumption for masking (see Flehinger *et al.* (1996)):

S1 $f_{\mathcal{M}|T,C}(A|t,c) = P(\mathcal{M} = A|T = t, C = c) = f_{\mathcal{M}|C}(A|c)\ \forall\ t.$

S1 means that given the failure cause, the probability of observing the masked failure causes is somewhat independent of the failure time. In addition to the critic that the symmetry assumption "is done purely for mathematical convenience without practical justification" (see Mukhopadhyay and Basu (2007, p.331[15])), it is often not satisfied in practice (see Example 2.1). In Remarks 2.1 and 3.2, we show that the RPM model has the following advantages over the CMP model:

(1) The CMP model relies on the symmetry assumptions but the RPM model does not;

(2) It is much simpler for implementation in simulation studies under the RPM model.

In §3 we establish the asymptotic properties of the NPMLE of the underlying scdf under the discrete assumption on inspection times. We also present the simulation results of the asymptotic properties of the NPMLE under various continuous assumptions on the inspection times. Moreover, we present the empirical rates of convergence based on the simulations. In §4, we present a data analysis on a cancer research data of size 375. The simulation study and data analysis suggest that the NPMLE procedure is feasible. We use the self-consistency algorithm to obtain the estimates on the maximal intersections(MI's, see §2). A key step in this algorithm is to find the exact MI's, which are described in Appendix I. A simpler algorithm for finding the pseudo MI's is given in §3.2, where the pseudo MI's are not the exact ones but cover the exact MI's, thus in some situations, to save the computing time, we can use pseudo MI's to replace the exact ones. In §5, we present the conclusion and discussion.

§**2. The NPMLE and the RPM Model.** The formal definition of the nonparametric likelihood function can be easily given without a proper model for the ICMCR data. However, in order to validate the nonparametric likelihood function and establish asymptotic properties of the NPMLE, we need set up a probability model and determine the identifiability conditions under which the NPMLE can be consistent.

First we introduce some notations. Define $\mathbf{F}_0^s(t) = (F_{10}^s(t), F_{20}^s(t), \cdots, F_{J0}^s(t))'$, where B' is the transpose of the matrix B. Then the true bivariate cumulative distribution function, denoted as F_0 satisfies $F_{0j}^s(t) =$

$F_0(t,j) - F_0(t,j-1)$ and $F_0(t,y) = \sum_{j=1}^{[y]} F_{j0}^s(t)$ where $[y] = \max\{c \in \mathcal{C}_r : c \le y\}$ and $\sum_{j=1}^J = F_{j0}^s(\infty) = 1$. For each $\mathbf{F}^s \in \mathcal{F}^s = \prod_1^J \mathcal{F}$ with $\sum_{j=1}^J F_j^s(\infty) = 1$ where $\mathcal{F}$ is the collection of all non-decreasing and right-continuous functions, define $\mu_{\mathbf{F}^s}(I_i) = \begin{cases} \sum_{j\in M_i}(F_j^s(R_i) - F_j^s(L_i)) & \text{if } I_i = (L_i, R_i] \times M_i,\ L_i < R_i \\ \sum_{j\in M_i} f_j^s(L_i) & \text{if } I_i = [L_i, R_i] \times M_i,\ L_i = R_i \end{cases}$ with $i = 1, \cdots, n$ where $f_j^s(t) = F_j^s(t) - F_j^s(t-)$. If $\hat{F}_{jn}^s(t)$, the NPMLE of $F_{j0}^s(t)$, is consistent for $j = 1, 2, \cdots, J$, then the NPMLE of F_0, denoted as $\hat{F}_n(t,c)$ is also consistent for $c = 1, 2, \cdots, J$. Thus we can find the NPMLE of F_0 by obtaining the NMPLE of its scdf. Though the main focus in on the estimation of the scdf, we still will use the bivariate cumulative distribution function to interpret something if necessary, especially comparison between RPM and CMP models.

§**2.1. The NPMLE.** Let $(L_i, R_i, \mathcal{M}_i)$, $i = 1, \ldots, n$, be i.i.d. copies of $(L, R, \mathcal{M}^o)$. Then the generalized likelihood function, according to the idea of Kiefer and Wolfowitz (1956), is

$$\Lambda_n(\mathbf{F}^s) = \Pi_{i=1}^n \mu_{\mathbf{F}^s}(I_i). \tag{2.1}$$

Our goal is to compute the NPMLE of $\mathbf{F}_0^s$ where $\Lambda_n(\hat{\mathbf{F}}_n^s) = \max_{\mathbf{F}^s\in\mathcal{F}^s} \Lambda_n(\mathbf{F}^s)$. Let A_j, $j = 1, \cdots, m$, be the maximal intersections A_js (see Example 1.1) induced by I_i's. It can be shown that the NPMLE of $\mathbf{F}_0^s$ assigns weights only to MI's (see §1 in Appendix III). The rigorous algorithm for finding all the exact MI's is given in Appendix I.

Denote the weight assigned by the NPMLE to A_j by s_j. It follows that (2.1) can be expressed as $\Lambda_n(\mathbf{s}) = \prod_{i=1}^n [\sum_{j=1}^m \delta_{ij} s_j]$, where $\mathbf{s} = (s_1, s_2, \cdots, s_{m-1})$, $\mathbf{s} \in D_s$, $D_s = \{\mathbf{s} : s_j \ge 0, s_1 + s_2 + \cdots + s_{m-1} \le 1\}$, $s_m = 1 - \sum_{j=1}^{m-1} s_j$ and $\delta_{ij} = \mathbf{1}_{(A_j \subset I_i)}$. Let $\hat{\mathbf{s}}$ be the NPMLE of $\mathbf{s}$, a NPMLE of $\mathbf{F}_0^s$ can be given by

$$\hat{F}_{cn}^s(t) = \hat{F}_n(t,c) - \hat{F}_n(t,c-1) \text{ and } \hat{F}_n(t,c) = \sum_{j:\ A_j \subset (-\infty,t]\times(-\infty,c]} \hat{s}_j,\ c = 1, 2, \cdots, J \tag{2.2}$$

(see, for example, Wong and Yu (1999)).

For the ICMCR data, a MI A_j has the form $(a_j, b_j] \times W_j$, where $W_j \subset \mathcal{C}_r$ and $W_j \ne \emptyset$, and thus it is not a singleton. If A_j is not a singleton, the distribution of the weight s_j over A_j is not uniquely determined. The $\hat{\mathbf{F}}_n^s$ in (2.2) is a standard approach, which assigns the weight $\hat{s}_j$ to the point (b_j, c), where $c = \max_{h\in W_j} h$. It turns out that the non-uniqueness of the NPMLE would not affect its consistency under certain regularity conditions (see Theorem 3.1 in §3.1 and the simulation results in §3.2). In general, the NPMLE does not have an explicit solution. An algorithm to be implemented is

the self-consistent (SC) algorithm which is described as follows:

At step 1, let $s_j^{(1)} = 1/m$ for $j = 1, 2, \cdots, m$.

At step h ($h \ge 2$),

$$s_j^{(h)} = \sum_{i=1}^n \frac{1}{n} \frac{\delta_{ij} s_j^{(h-1)}}{\sum_{k=1}^m \delta_{ik} s_k^{(h-1)}}, j = 1, 2, \cdots, m \text{ and } h \ge 2. \tag{2.3}$$

Stop at convergence, which means $||\mathbf{s}^{(h)} - \mathbf{s}^{(h-1)}||$ is small enough.

By the similar way as in Turnbull (1976), it can be shown that when $h \to \infty$, $\mathbf{s}^{(h)}$ converges to the NPMLE $\hat{\mathbf{s}}$ which maximizes Λ_n. Then an NPMLE of $\mathbf{F}_0^s$ can be computed through (2.2).

§**2.2. The RPM Model for the ICMCR Data.** An important issue is to specify a probability model so that the likelihood function $\Lambda_n(\mathbf{F}^s = \prod_{i=1}^n \mu_{\mathbf{F}^s}(I_i)$ is valid. Basu *et al.* (2003) point out that almost all existing models for competing risk data with masking are based on the symmetry assumption S1.

Remark 2.1. It is shown in §6 of Appendix III that under certain assumptions, the RPM model and the CPM model are equivalent. Moreover, S1 is valid iff S2 holds.

S2 For each t, $f_T(t) > 0$ implies $f_{T,C}(t,c) > 0\ \forall\ c \in \mathcal{C}$.

Example 2.1. Consider a discrete case. Suppose that $J = 2$, the survival time $T \in \{1, 2\}$ and the failure cause $C \in \{1, 2\}$. There are two partitions denoted as $P_1 = (\{1\}, \{2\})$ and $P_2 = (\{1, 2\})$. Suppose that $P(T = 1) = P(T = 2) = 0.5$, $f_{T,C}(1,1) = 0$, $f_{T,C}(2,1) > 0$ and $P(\Delta = 1) = P(\Delta = 2) = 1/2$ where Δ denotes the random index of the partitions. Assume a model with two censoring variables Y_1 and Y_2 such that $P(Y_1 = 1, Y_2 = 2) = 1$. Then $f_{\mathcal{M}|T,C}(\mathcal{C}_r|t,c)$ is not constant in t if $c = 1$. That is, the symmetry assumption S1 fails. But it can be shown that the NPMLE of F_0 is consistent for $t \in \{1\} \cup ([2,\infty)$ and is asymptotically normally distributed (see the proof in §2 of Appendix III).

The NPMLE in Figure 1 in §4 suggests that S2 is not valid for our cancer research data. Thus we shall propose a new model to formalize these data forms, which consists of an interval censorship model for generating (L, R) and a masking model for generating $\mathcal{M}^o$.

First, we introduce the mixed case interval censorship model (see Schick and Yu (2000)) for specifying (L, R). Let K be a positive random integer representing the number of inspection times. Let $\mathbf{Y}$ ($\stackrel{def}{=}\{Y_{k,j} \ : \ j = 1, 2, \cdots, k, \ \ k = 1, 2, \cdots\}$) be an array of random variables such that $-\infty < Y_{k,1} < Y_{k,2} < \cdots < Y_{k,k} < \infty$, which are the inspection times when an individual is subject to total of k inspections. Then given the event $\{K = k\}$, let (L, R) denote the endpoints of the random interval among $(Y_{k,0}, Y_{k,1}], (Y_{k,1}, Y_{k,2}], \cdots, (Y_{k,k}, Y_{k,k+1})$ which contains T, where $Y_{k,0} = -\infty$, $Y_{k,k+1} = \infty$.

Second, we propose a masking model. Let $\mathcal{P}$ be the collection of all possible partitions of $\mathcal{C}_r$. A partition $P_h = (P_{h1}, P_{h2}, \cdots, P_{hk_h})$ satisfies that $P_{hi} \subset \mathcal{C}_r$, $P_{hi} \neq \emptyset$, $\bigcup_{i=1}^{k_h} P_{hi} = \mathcal{C}_r$ and $P_{hi} \bigcap P_{hj} = \emptyset \ \forall \ i \neq j$. By the definition, for each given partition P_h and given value of C, say $C = c$, there exists a unique i such that $C = c \in P_{hi}$ (the interpretation of a partition is explained in Remark 2.3). The number of possible partitions, denoted by $n_{\mathcal{P}}$, is finite. Thus there are $n_{\mathcal{P}}!$ different ways to order these partitions. Let $P_1, P_2, \cdots, P_{n_{\mathcal{P}}}$ be one such ordering on the partitions. Then we can define a random variable, say Δ, taking values on $\{1, 2, \cdots, n_{\mathcal{P}}\}$ which are the indexes on the partitions, with the density function f_Δ. Once a partition $(P_{h1}, P_{h2}, \cdots, P_{hk_h})$ is chosen after the failure occurs, $\mathcal{M}^o$ can be uniquely determined through the formula

$$\mathcal{M}^o = \begin{cases} \mathcal{C}_r & \text{if } T > Y_{k,k} \\ P_{hi} & \text{if } C \in P_{hi}, \ T \leq Y_{k,k}, \end{cases} \quad \text{conditioning on } K = k \text{ and } \Delta = h.$$

The RPM model for ICMCR data can be summarized as follows by assuming that (T, C) and $(K, \Delta, \mathbf{Y})$ are independent. Conditioning on $K = k$ and $\Delta = h$,

$$(L, R, \mathcal{M}^o) = \begin{cases} (Y_{k,0}, Y_{k,1}, P_{hj}) & \text{if } C \in P_{hj}, \ P_{hj} \in P_h, \ T \leq Y_{k,1}, \\ (Y_{k,i-1}, Y_{k,i}, P_{hj}) & \text{if } C \in P_{hj}, \ P_{hj} \in P_h, \ Y_{k,i-1} < T \leq Y_{k,i}, \ i = 2, \cdots, k, \\ (Y_{k,k}, Y_{k,k+1}, \mathcal{C}_r) & \text{if } T > Y_{k,k}. \end{cases}$$

Then the full likelihood function is

$$\begin{aligned} \mathcal{L}_{full} =& \prod_{i=1}^{n} \sum_{k=1}^{\infty} \sum_{h: \ M_i \in P_h} \sum_{j=0}^{k} f_{Y_{K,j}, Y_{K,j+1}, K, \Delta}(L_i, R_i, k, h) \sum_{c \in M_i} \int_{L_i}^{R_i} dF_{T,C}(t, c) \\ =& [\prod_{i=1}^{n} \sum_{k} \sum_{h: \ M_i \in P_h} \sum_{j=0}^{k} f_{Y_{K,j}, Y_{K,j+1}, K, \Delta}(L_i, R_i, k, h)] \prod_{i=1}^{n} \sum_{c \in M_i} \int_{L_i}^{R_i} dF_{T,C}(t, c) \\ \propto& \prod_{i=1}^{n} \mu_{F_{T,C}}(I_i). \end{aligned} \tag{2.4}$$

This justifies the formal likelihood function in (2.1).

The model we considered here is a brand new one in the following senses.

(1) The nonparametric model for both interval censoring and masking has not been studied so far.

(2) The masking model is different from all existing approaches in the literature.

Verify that the derivation in (2.4) does not rely on the symmetry assumption S1. Thus the new model does not rely on the symmetry assumption at all. We call the old model the conditional masking probability (CMP) model as they depend on the CMP (see S1) and we call the new model the RPM model for the obvious reason. In view of Remark 2.1, the CMP model can be viewed as a special case of the RPM model. This is a significant contribution to the theory for analyzing the CR data.

Remark 2.2. Even though nobody has done it in the literature, one can construct a model for the ICMCR data based on the symmetric assumption S1. Let's call it Model 2. Under Model 2 and the RPM model, it makes no difference in obtaining the MLE or NPMLE, but it does make a difference in the proofs of asymptotic properties of the estimates. By Remark 2.1 and Example 2.1, the consistency cannot be established under Model 2 but it can be established under the RPM model. In applications, the symmetry assumption is often not satisfied. Thus the new model is less restrictive than the old one. Moreover, it is much easier to carry out simulation studies under the new model.

There is a realistic interpretation to the partition in industrial data. When a TV set breaks down, there is often an procedure to find out the failure cause. Suppose that there are J causes and $P_2 = (\{1\}, \{2\}, \{3, 4, ..., J\})$ is a partition. The partition P_2 can be interpreted as follows: In the process of determining the cause of failure in a J-component series system, exactly two steps will be taken. Steps 1 and 2 can determine whether the failure is due to causes 1 and 2, respectively. If the failure is not due to these two causes, no further investigation will be taken for cost saving. However, it is only one of the six examination schemes corresponding to P_2 and each has two steps. The first step can be either of the three inspections:

(1) whether the cause is due to part 1;

(2) whether the cause is due to part 2;

(3) whether the cause is not due to parts 1 and 2.

There is also a realistic interpretation to the partition in medical data. For instance, in the breast cancer data to be analyzed in the paper, a relapse can be classified to 7 types of cancer. If the length between two follow-ups are short enough, it is likely that there is just one type of cancer and the partition can be viewed as $(\{1\}, \{2\}, ..., \{J\})$. Otherwise, there are several types of cancer that occur between L and R. For instance, a patient has breast cancer (type 1) and lung cancer (type 3) at relapse, then the partition can be viewed as $(\{1, 3\}, \{j\}, j = 2, 4, 5, 6, 7)$, as well as $(\{1, 3\}, \{2, 4, 5, 6, 7\})$, etc.. It is not important to precisely identify the partition, as far as the NPMLE is concerned.

Remark 2.3. It is worth mentioning that in the aforementioned engineering applications, one may choose the value of Δ randomly, but in the medical applications like our cancer research, the nature chooses the value of Δ.

Remark 2.4. The new masking model is akin to the case 1 interval censorship (C1) model. In the C1 model, there is a random partition $\{(-\infty, Y], (Y, \infty)\}$ (denoted by $\{I_1, I_2\}$) of the space $(-\infty, \infty)$ based on the random inspection time Y. The observation on the survival time T is I_j where $T \in I_j$. In our masking model, there is a random partition $\{P_{\Delta,1}, ..., P_{\Delta,k_\Delta}\}$ of the space $\{1, ..., J\}$. The observation on C is $\mathcal{M} = P_{\Delta,j}$ where $C \in P_{\Delta,j}$. Thus the new masking model is more reasonable than the CMP model.

§**2.3. Identifiability Assumptions.** As to be shown in Example 2.3, without suitable identifiability condition, the NPMLE is not consistent. One may think of a quite strong identifiability condition as follows.

A1 For each t in the range of finite L or R, let $W_{t,j}$, $j = 1, \cdots, m_t$ be all the values of $\mathcal{M}^o$ satisfying $P(\mathcal{M}^o = W_{t,m} | L = t \text{ or } R = t) > 0$, then the rank of the matrix $(\phi(W_{t,1})', \cdots, \phi(W_{t,m_t})')$ is J, where $\phi(W_{i,m}) = (\mathbf{1}_{(1 \in W_{i,m})}, \mathbf{1}_{(2 \in W_{i,m})}, \cdots, \mathbf{1}_{(J \in W_{i,m})})$.

For discrete L and R, given a random sample from the ICMCR data, if n is large enough then one can check whether A1 holds through the sample data by the condition as follows.

A1* For each $t \in X$, where X denotes the collection of all distinct values from L_1,..., L_n, R_1,...,R_n, let $W_{t,1}$,..., W_{t,m_t} be all the distinct values of M_i's with $L_i = t$ or $R_i = t$, then $rank(\phi(W_{t,1})', \cdots, \phi(W_{t,m_t})')$ is J.

It is obvious that if A1* is false, A1 might still be true, as it is hard to say whether n is indeed large enough. For continuous L and R, A1 is difficult to verify through the sample. In fact given t, with probability 1 there is just 1 observation (L_i, R_i, M_i) such that $L_i = t$ or $R_i = t$, then $m_t = 1$ and $rank(\phi(W_{t,1})') = 1 < J$.

Since the inspection times may depend on the failure components, the consistency region for each NPMLE of the scdf may vary component by component.

Remark 2.5. A1 ensures that the NPMLE of $\mathbf{F}^s_0(t)$ is consistent for each t in the range of L or R, provided $t < \tau$ for some τ. Example 2.2 below suggests that under a weaker condition, the boundary point τ should depend on the component c, say τ_c.

Example 2.2. Let the domain of (T, C) be $\{0, 1, 2\} \times \{1, 2, 3\}$. Consider the case 1 model with the censoring variable $Y_{1,1} \sim Bin(1, p)$ for some $p \in (0, 1)$, define $P_1 = (\{1\}, \{2\}, \{3\})$ and $P_2 = (\{1, 3\}, \{2\})$. Let $P(Y_{1,1} = 0, \Delta = 1) = P(Y_{1,1} = 1, \Delta = 2) = 1/2$. It can be easily verified that the NPMLE of $F^s_{20}(1)$ is consistent but not the NPMLE's of $F^s_{10}(1)$ and $F^s_{30}(1)$. So we need to define a bound, say τ_c such that the NPMLE of the scdf $F^s_{c0}(t)$ is not consistent for each $t \in (\tau_c, \infty)$.

In view of Example 2.2, we shall give a weaker identifiability condition. For a better presentation, we first consider it under discrete assumptions. In fact, if one assumes that the censoring vector (L, R) takes on finitely many values, since $\mathcal{M}$ also takes on finitely many values, the current estimation problem reduces to a multinomial distribution problem, and it can be shown that under certain regularity conditions and under the RPM model, the NPMLE of $\mathbf{F}_0^s$ is consistent and asymptotically normally distributed, and the asymptotic covariance matrix can be estimated by the inverse of the Fisher information matrix.

Since a follow-up study often lasts for a certain period of time and carries out the inspections at some fixed times, we can assume that the inspection times $Y_{k,i}$'s only takes values from the set $\mathcal{S} = \{t_1, \cdots, t_d\}$ with $-\infty = t_0 < t_1 < \cdots < t_d$, where $d < \infty$ with $P(Y_{k,i} = t_j) > 0$ for some i and j, then the observations are of the form $I_l = (t_{i_l}, t_{j_l}] \times M_l$ for some $t_{i_l} < t_{j_l}$ and $M_l \in \mathcal{J}(\stackrel{def}{=} \{A \neq \emptyset : A \subset \mathcal{C}_r\})$, $l = 1, \cdots, n$. Moreover, when the sample size n is large enough, each values of $\mathcal{S}$ will be observed. Thus, under the finite discrete assumption and by the definition of MI, if n is large enough, the sample MI's will be same as the population MI's which have finitely many distinct values and the form is $(t_{j-1}, t_j] \times M$ with $M \in \mathcal{J}$ for $j = 1, \cdots, d$ or $(t_d, \infty) \times \mathcal{C}_r$.

Let $\mathcal{A} = \{A_j : j = 1, \cdots, m\}$ be the collection of all the MI's, then the weight assigned by an $\mathbf{F}^s \in \mathcal{F}^s$ to each A_j, denote $s_j = \mu_{\mathbf{F}^s}(A_j) = \sum_{c \in M_j} (F_c^s(R_j) - F_c^s(L_j))$ where $(L_j, R_j) \in \{(t_0, t_1), (t_1, t_2), \cdots, (t_{d-1}, t_d), (t_d, \infty)\}$ and $M_j \in \mathcal{J}$. The log likelihood function can be expressed as $\mathcal{L}_n(\mathbf{s}) = \sum_{i=1}^n \log[\sum_{j=1}^m \delta_{ij} s_j]$. Denote the NPMLE of $s_j^o = \mu_{\mathbf{F}_0^s}(A_j)$ by $\hat{s}_j$ for $j = 1, \cdots, m$. We can show that each NPMLE of s_j's in Example 2.2 is consistent (by Theorem 3.1 in §3). However, this is not so in Example 2.3.

Example 2.3. Let $J = 4$, $T \in \{1, 2\}$. Consider the case 1 model, that is, $K = 1$ w.p.1 and the censoring variable $U = Y_{1,1} \in \{1, 2\}$. Order the partitions as $P_1 = (\{1\}, \{2\}, \{3\}, \{4\})$, $P_2 = (\{1, 2\}, \{3, 4\})$, and $P_3 = (\{1, 3\}, \{2, 4\})$. Let the conditional density of Δ given U be $f_{\Delta|U}(1|1) = 1$, $f_{\Delta|U}(2|2) = f_{\Delta|U}(3|2) = 1/2$. When n is large enough, the possible observations are $(-\infty, 1] \times \{1\}$, $(-\infty, 1] \times \{2\}$, $(-\infty, 1] \times \{3\}$, $(-\infty, 1] \times \{4\}$, $(1, \infty) \times \{1, 2, 3, 4\}$, $(-\infty, 2] \times \{1, 2\}$, $(-\infty, 2] \times \{3, 4\}$, $(-\infty, 2] \times \{1, 3\}$ and $(-\infty, 2] \times \{2, 4\}$ with sizes N_1, ... N_9, respectively. Then the MI's are $(-\infty, 1] \times \{1\}$, $(-\infty, 1] \times \{2\}$, $(-\infty, 1] \times \{3\}$, $(-\infty, 1] \times \{4\}$, $(1, 2] \times \{1\}$, $(1, 2] \times \{2\}$, $(1, 2] \times \{3\}$ and $(1, 2] \times \{4\}$ with weights s_1, ... s_8, respectively. The NPMLE of $(s_1, ..., s_8)$ is

$$\begin{array}{llll} \hat{s}_1 = \frac{N_1}{W_1}, & \hat{s}_3 = \frac{N_3}{W_1}, & \hat{s}_5 = W_3 - \frac{N_1+N_2}{W_1} - \alpha, & \hat{s}_7 = \alpha, \\ \hat{s}_2 = \frac{N_2}{W_1}, & \hat{s}_4 = \frac{N_4}{W_1}, & \hat{s}_6 = W_2 - W_3 + \frac{N_3-N_2}{W_1} + \alpha, & \hat{s}_8 = 1 - \hat{s}_1 - \cdots - \hat{s}_7, \end{array} \tag{2.5}$$

$$\text{where } \max\{0, \frac{N_2 - N_3}{W_1} + W_3 - W_2\} \leq \alpha \leq \min\{W_3 - \frac{N_1 + N_3}{W_1}, W_2 + W_3 + \frac{N_2 - N_3}{W_1}\},$$

$W_1 = \sum_{i=1}^5 N_i$, $W_2 = \frac{N_6}{N_6+N_7}$, and $W_3 = \frac{N_8}{N_8+N_9}$ (see the derivation in §3 of Appendix III). It is easy to verify that $\hat{s}_i$ are consistent for $i = 1, 2, 3, 4$ but not for $i = 5, 6, 7, 8$.

Verify Examples 2.1 and 2.2 satisfy A2 below, but Example 2.3 does not.

A2 Let $A_{i,j} = (t_{i-1}, t_i] \times B_{i,j}$, $j = 1, \cdots, k_i$, $i = 1, \cdots, d$ be all the distinct population MI's, where $B_{i,j}$'s are induced by $W_{i,m}$, $m = 1, \cdots, w_i$ with $P(\mathcal{M}^o = W_{i,m}, L = t_{i-1}) > 0$ and $P(\mathcal{M}^o = W_{i,m}, R = t_i) > 0$. Then for each possible (i, j), there exists some constant $g_{i,j,m}$'s such that $\phi(B_{i,j}) = \sum_{m=1}^{w_i} g_{i,j,m} \phi(W_{i,m})$.

A2 is weaker than A1. Verify that A1 and A2 imply that for a given random sample,

A3 if n is large enough then the matrix $(\phi(M_1)', \cdots, \phi(M_n)')$ is of rank J.

If A3 does not hold, it is likely that there is no consistent estimator of $\mathbf{F}_0^s$. Thus A1* and A3 can be used to check whether A2 holds or not.

§3. Asymptotic Properties of the NPMLE.

We shall study the asymptotic properties under discrete or continuous assumptions. Under the discrete assumptions, we shall establish its properties by rigorous proofs. Under continuous assumptions, we are also working on the consistency proof of the NPMLE in a paper under preparation, and in this section, we shall present simulation studies that suggest that the NPMLE is consistent.

3.1. Theorems. Under discrete assumptions, we have the following theorems.

Theorem 3.1 (Consistency). Suppose that the inspection time takes values from $\mathcal{S} = \{t_1, \cdots, t_d\}$ with $0 \leq t_1 < \cdots < t_d < \infty$, and A2 holds. Then $\hat{\mathbf{s}} \to \mathbf{s}^o$ almost surely.

Theorem 3.2 (Asymptotic Normality). Assume that the inspection time takes values from $\mathcal{S} = \{t_1, \cdots, t_d\}$ with $0 \leq t_1 < \cdots < t_d < \infty$, A2 holds and denote $\hat{s}_j = \mu_{\hat{\mathbf{F}}_n^s}(A_j)$ and $s_j^o = \mu_{\mathbf{F}_0^s}(A_j) > 0$ for each $A_j \in \mathcal{A}$, $j = 1, \cdots, m$. Then $\sqrt{n}\begin{pmatrix} \hat{s}_1 - s_1^o \\ \vdots \\ \hat{s}_{m-1} - s_{m-1}^o \end{pmatrix}$ is asymptotically normal distributed with mean $\mathbf{0}$ and dispersion matrix $\mathcal{I}^{-1}$ where $\mathcal{I} = -E(\frac{\partial^2 \mathcal{L}_n(\mathbf{F}_0^s)}{\partial \mathbf{s} \partial \mathbf{s}'})$.

For a better presentation, the proofs of the two theorems are given in §9 and §10 of Appendix III. From the last two theorems, the consistent estimator of $\mathcal{I}$ can be given by $\hat{\mathcal{I}} = -\frac{\partial^2 \mathcal{L}_n(\hat{\mathbf{F}}_n^s)}{\partial \mathbf{s} \partial \mathbf{s}'}$. Furthermore, for $c = 1, \cdots, J$, $\sqrt{n}(\hat{F}_{cn}^s(t) - F_{c0}^s(t))$ is asymptotically normally distributed with $\hat{F}_{cn}^s$ being defined by equation (2.2). A consistent estimator of the asymptotic variance of $\hat{F}_{cn}(t)$ is $\frac{1}{n}\mathbf{u}'\hat{\mathcal{I}}^{-1}\mathbf{u}$ where $\mathbf{u}$ is a $(m-1) \times 1$ vector with ith entry $u_i = \mathbf{1}(A_i \subset (-\infty, t] \times [c, c])$.

Remark 3.1. We may construct confidence regions or testing hypotheses based on the asymptotic normality established in Theorem 3.2 under the discrete assumption. In particular, in the literature of competing risks data with masking, most people focus on parametric analysis, but none of them check the validity of their parametric assumptions, because the asymptotic properties of the nonparametric NMLE of $\mathbf{F}_0^s$ have not been established so far. Our results in Theorem 3.2 can be applied to test H_0: $F_c^s = F_{co}^s$ for $c = 1, \cdots, J$, where F_{co}^s is a given scdf with or without parameters. Due to the length of the paper, we shall address this issue in a future project.

§**3.2. Simulation.** We shall first present some simulation results which suggest that with ICMCR data the NPMLE is feasible at least for sample sizes up to 800 and $J = 4$, even if we use the alternative simpler but slower algorithm (see §3.2.2). Our simulation results also suggest that under A1 the NPMLE of $\mathbf{F}_0^s$ is consistent if T is continuous.

The implementation of the RPM model in the simulation can be described as follows:

First choose densities f_C, $f_{T|C}$, f_Δ, $f_{K|\Delta}$, and $f_{(Y_{k,1},...,Y_{k,k})|\Delta,K}$ with $f_{K|\Delta}(k|\cdot) > 0$.

1. Generate $C \sim f_C$, say, $C = j$.
2. Conditional on $C = j$, generate $T \sim f_{T|C}(\cdot|j)$, say, $T = t$, now $(T, C) = (t, j)$.
3. Generate $\Delta \sim f_\Delta$, say, $\Delta = h$. Then $\mathcal{M} = P_{hl}$ where $j \in P_{hl}$.
4. Generate $K \sim f_{K|\Delta}(\cdot|h)$, say, $K = k$, then $(Y_{k,1}, ..., Y_{k,k}) \sim f_{(Y_{k,1},...,Y_{k,k})|\Delta,K}(\cdot|h, k)$, say, $(Y_{k,1}, Y_{k,2}, \cdots, Y_{k,k}) = (y_{k,1}, y_{k,2}, \cdots, y_{k,k})$ with $y_{k,1} < \cdots < y_{k,k}$.
5. Conditional on $\Delta = h$, $K = k$, $(T, C) = (t, j)$, and $Y_{k,i} = y_{k,i}$, find i such that $t \in (y_{k,i-1}, y_{k,i}]$. If $i \leq k$, then $(L, R, \mathcal{M}^o) = (y_{k,i-1}, y_{k,i}, P_{hl})$ where $j \in P_{hl}$ for some $1 \leq l \leq ||P_h||$, the number of elements in P_h. Otherwise, let $(L, R, \mathcal{M}^o) = (y_{k,k}, \infty, \mathcal{C}_r)$.

Repeat this procedure n times, we obtain the needed observations.

In the foregoing implementation scheme, it is assumed that Δ, K and $\mathbf{Y}$ are dependent. In our simulation, we let Δ, K and $\mathbf{Y}$ be independent for simplicity.

Remark 3.2. It is more complicated to implement a simulation study with the CMP model than with the RPM Model. Under the CMP Model, in order to generate $\mathcal{M}^o$, one has to solve for $f_{\mathcal{M}|C}$'s subject to constraints S1 and the hidden constraints $\sum_{A: A \in \mathcal{J}} f_{\mathcal{M}|C}(A|c) = 1$. Then generate $\mathcal{M}$ based on $f_{\mathcal{M}|C}(\cdot|j)$, replacing aforementioned Step 3. This can be done but is not as simple as the procedure for generating $\mathcal{M}$ under the RPM model(see Step 3 above).

§**3.2.1. Simulation Results.** Based on the RPM model, we generate observations under various distributions, compute the NPMLE of the scdf at four different quantiles of the conditional distribution $F(t|c) = P(T \leq t|C = c)$, and compare them to the true values. In our simulation studies, we let $J = 4$, and allow K to take values up to 16. We consider both continuous Y_i's (in Examples 3.1, 3.2 and 3.3) and discrete Y_i's (in Example 3.4), which are specified in their examples. In all examples, we assume that $f_C(1) = 1/40$, $f_C(2) = 1/40$, $f_C(3) = 36/40$, $f_C(4) = 2/40$ and assume that $\mathcal{M}$ is generated through (3.1) with $P(\Delta = 1) = \frac{8}{10}$, $P(\Delta = 2) = P(\Delta = 3) = \frac{1}{10}$, and $P_1 = (\{1\}, \{2\}, \{3\}, \{4\})$, $P_2 = (\{1, 2\}, \{3, 4\})$, $P_3 = (\{1, 3\}, \{2, 4\})$. Verify that these partitions satisfy assumption A3 at least. However, $f_{T|C}$ and $f_{\mathbf{Y}}$ are specified in each example. We computed the estimators through the SC-algorithm with the error bound 0.0001 for sample sizes n = 100, 200, 400 and 800. The simulation results in Examples 3.1-3.4 are given in Tables 1-4 in Appendix II, respectively.

Example 3.1 (Log-normal Distribution). By $T \sim LN(\mu, \sigma)$, we mean $\log T \sim N(\mu, \sigma^2)$. Let $T|(C = j) \sim LN(5, \frac{5-j}{2})$ and $Y_i \sim U(2(i-1), 2i)$ for $i = 1, \cdots, K$.

Example 3.2 (Exponential Distribution). Let $P(T \le t|C = j) = 1 - e^{-t/2^{(j-1)}}$, $t > 0$, and $Y_i \sim U(0, 10)$, for $i = 1, \cdots, K$.

Example 3.3 (Uniform Distribution). Let $f_{T|C}(t|1) = \frac{1}{8}\mathbf{1}(t \in (0, 8))$; $f_{T|C}(t|2) = \frac{1}{7}\mathbf{1}(t \in (0, 7))$; $f_{T|C}(t|3) = \frac{1}{6}\mathbf{1}(t \in (0, 6))$; $f_{T|C}(t|4) = \frac{1}{5}\mathbf{1}(t \in (0, 5))$, and $Y_i \sim U(2(i-1), 2i)$ for $i = 1, \cdots, K$.

Example 3.4 (Discrete inspection times). Let (T, C) have a distribution as in Example 3.1, but the distribution of $(\mathbf{\Delta}, K, \mathbf{Y})$ is specified as follows. Let the domain of the inspection times be B, where $B = \{2.155, 3.012, 3.283, 3.419, 3.881, 4.405, 5, 5.675, 6.442, 6.499, 7.312, 7.616, 8.299, 8.447, 10.98, 14.27\}$. Let $f_K(k) = \begin{cases} 0.1 & \text{if } k \in \{3, 5, 7, 9, 13\} \\ 0.5 & \text{if } k = 16. \end{cases}$ Conditional on $K = k$, we choose the first k values from the 16 ordered numbers in B.

From the results in Tables 1 through 4 under different distributions, we notice that for each risk component, the sample mean of the estimated probability is around the true value, and the standard error is getting to zero as n increases. Thus these simulated results suggest that $|\hat{F}^s_{jn}(t) - F^s_{j0}(t)| \to 0$ as $n \to \infty$ for the relevant values of t.

Remark 3.3. In view of Remark 2.1, the conditions in the 4 examples in our simulation satisfy the symmetry assumption S1. However, it is much easier to generate the pseudo observation under the RPM model than under the old CMP model.

§**3.2.2. An alternative approach for finding the MI's.** A key process in computing the NPMLE through (2.3) is to find all the MI's. A rigorous algorithm is given in Appendix I, which is somewhat cumbersome and complicated. There is an alternative simpler approach for finding a collection of pseudo MI's, which can replace the role of all MI's. The algorithm for finding the pseudo MI's is as follows.

Assume that the observations are $I_i = (l_i, r_i] \times M_i$, $i = 1, 2, \cdots, n$. For each $j \in \mathcal{C}_r$,

1. let $G_j = \{(l_{j,i}, r_{j,i}] \times \{j\} : i = 1, 2, \cdots, n_j\}$ be the collection of sets of the form $(l_i, r_i] \times \{j\}$ where $j \in M_i$, $i \in \{1, ..., n\}$;
2. find the MI's induced by the aforementioned $(l_{j,i}, r_{j,i}]$'s, denoted by $(a_k, b_k]$'s;
3. for each (k, j), treat $(a_k, b_k] \times \{j\}$ as a true MI induced by I_i's.

The aforementioned approach yields more "MI's" than the approach in Appendix I, but both of them lead to the same estimator of F^s_{c0}. In Tables 5 and 6 in Appendix II, we display ratio(t, j) ($\stackrel{def}{=} \frac{2 \cdot |\hat{F}_{j1}(t) - \hat{F}_{j2}(t)|}{(\hat{F}_{j1}(t) + \hat{F}_{j2}(t))}$), where $\hat{F}_{j1}$ and $\hat{F}_{j2}$ are the estimators of the j^{th} scdf at time t under the first approach and the alternative approach, respectively, under assumptions in Examples 3.1 and 3.4. The ratio(t, j)'s are very small ($< 1\%$). Thus they suggest that the two estimators from the different approaches can give us the same result under the tolerance given in the self-consistency algorithm.

The alternative approach is simple to remember and easy to implement, but is slower than the rigorous one as it creates more pseudo MI's to replace the true MI's. When the sample size is large enough, their difference in computing time is quite small (see Table 7 in Appendix II, where Ratio $= \frac{\text{Time used in Approach 1}}{\text{Time used in Approach 2}}$). However, these two approaches will never be the same unless the largest observation is not right censored.

§**3.2.3. Empirical Rates of Convergence.** In Appendix III we show that the NPMLE of the ICMCR data under the discrete inspection times is asymptotically normal. Example 3.4 satisfies the assumptions that the inspection times are discrete and by comparing the rates given in Table 11 with the value 0.7071, we can find the empirical rate of convergence is close to $n^{1/2}$, which is exactly as shown in the theorem. For the univariate case, there is the same result for the discrete inspection times, see Yu *et al.* (1998a,b), which means there are some connections between the ICMCR data and univariate interval-censored data under the discrete inspection times. Similarly, the results on the rate of convergence r of the NPMLE of the cdf F_T of a univariate random variable T under the case 2 model can be the references for our further study on the rate of convergence of the NPMLE for the continuous inspection times.

(1) Groeneboom and Wellner (1992) conjecture that $r = (n \ln n)^{1/3}$ under the assumption that F_T and $F_{U,V}$ have strictly positive and continuous derivatives at $T = x$ and $(U, V) = (x, x)$ where (U, V) are two random inspection times in the case 2 model.

(2) Groeneboom (1996) proves $r = n^{1/3}$ under the assumptions that F_T is continuous with bounded derivative $f_T(x) \ge c_0 > 0, x \in (0, M)$, for some constant c_0 and M. (U, V) are two continuous random

inspection times in the case 2 model with the following additional conditions:

(2.1) f_U and f_V are continuous with $f_U(x) + f_V(x) > 0$ for $\forall\ x \in [0, M]$;

(2.2) $f_{U,V}$ is continuous with uniformly bounded partial derivatives, except at a finite number of points, where left and right (partial) derivatives exist;

(2.3) $P(V - U < \epsilon_0) = 0$ for some $\epsilon_0 \in (0, \frac{1}{2M}]$.

Tables 8 through 11 given in Appendix II show the empirical convergent rates based on the data from Examples 3.1-3.4. Notice that $(\frac{100\ln 100}{200\ln 200})^{1/3} = 0.7575$, $(\frac{200\ln 200}{400\ln 400})^{1/3} = 0.7618$, $(\frac{400\ln 400}{800\ln 800})^{1/3} = 0.7653$, $\lim_{n\to\infty}(\frac{n\ln n}{2n\ln(2n)})^{1/3} = (\frac{n}{2n})^{1/3} = (\frac{1}{2})^{1/3} = 0.7973$. We consider $f^s_{30}(t)$ mainly.

In Examples 3.1 and 3.3, $f^s_{30}(t)$ is positive, continuous and bounded on $[0, M]$ for some constant M such that the quantile $t \in [0, M]$, obviously it is bigger than some c_0 for $t \in (0, M)$. In both examples $P(Y_{k,i} - Y_{k,i-1} < \eta) = 0$ for some $\eta \in (0, \frac{1}{2M}]$. Thus, Example 3.1 and 3.3 satisfy the assumptions given in (2). By comparing the rates given in Tables 8 and 10 with the computed value 0.7973, we can find that almost all of the rates have a trend to be away from 0.7071 and oscillating increase to some value which is maybe 0.7973, thus it suggests that the convergent rate is $n^{1/3}$.

In Example 3.2, f^s_{30} is the density function of the exponential distribution which is strictly positive continuous. The inspection times are also positive continuous and the joint distribution of Y_i's are positive in the same domain. Thus at each quantile t for risk 3, $f^s_{30}(t)$ satisfies the conditions given in (1). By comparing the rates given in the three rows of each block of Table 9 to the values 0.7575, 0.7618 and 0.7653, respectively, we can find most of the rates have a trend to deviate from 0.7071 and be around 0.7575, 0.7618 and 0.7653 respectively for each n, thus it suggests that the convergent rate is $(n \log n)^{1/3}$.

§**4. Data Analysis.** Our proposed procedure is applied to a standard breast cancer relapse follow-up study based on data from 375 women with stages I - III unilateral invasive breast cancer surgically treated at Memorial Sloan-Kettering Cancer Center between 1985 and 1990. The median follow-up duration was 46 months. If a patient did not relapse toward the end of the study, then her relapse time was right censored. Of the 375 observations, 288 were right censored (no relapse), the other 87 patients had relapses. When cancer relapses, there may be several types of cancer diagnosed, and it was not clear which cancer relapsed first. These data consist of 7 risk components described as follows.

Cause	(1)	(2)	(3)	(4)	(5)	(6)	(7)
name	*Breast*	*Lung*	*Bone*	*Liver*	*Mediastinum*	*Supraclavicular*	*OTHER*

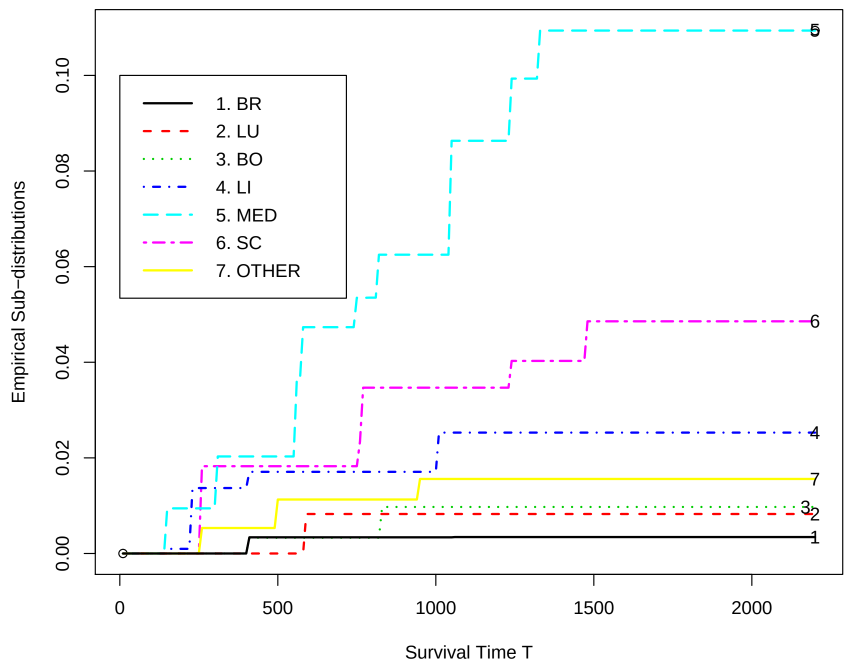

Figure 1. Estimation of the sub−distribution function for each type of cancer

Figure 1. Empirical scdf for each type of cancer.

The observed masked competing risks contain $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{5\}$, $\{6\}$, $\{7\}$, $\{1,3\}$, $\{1,5\}$, $\{4,5\}$, $\{4,6\}$, $\{5,7\}$, $\{5,6,7\}$, $\{1,2,3,4,5,6,7\}$. The observation $\{1\}$ means that we know the breast cancer relapse first; $\{4,6\}$ means that type of cancer that relapse first is either the liver cancer or supraclavicular. Thus it contains masking and it can be verified that the data satisfy assumption A3 at least. The empirical scdfs are given in Figure 1. According to Figure 1, the relapse rate for the Breast cancer is the smallest. Notice that these patients are all breast cancer patients after surgery. This analysis suggests that the relapse is mainly not due to breast cancer. During the study time, the highest relapse rate is from the fifth type of cancer.

§5. Conclusion and Discussion.

This paper proposes a RPM model to estimate the scdfs based on the ICMCR data. We also compare the proposed model with the existed CMP model to obtain that our model is easier implemented in the simulation and the real data analysis and less restrictive.

We developed an algorithm to find the Mis and use a self-consistency algorithm to estimate the scdfs and is applied to the simulations and real data analysis. In the simulations we find when the sample size increases, the estimates are closer and closer to the true value. Also we compute the empirical rate of convergence. In the theoretical part, we give the proofs of the consistency and the asymptotical normality under the discrete inspection times. We will give the proof for the continuous inspection times in the future work. Also in the future work, we can provide some hypothesis to test whether a parametric model is valid or not.

Appendix I. Algorithm For Finding MI's.

Recall that $(l_i, r_i, \mathcal{M}_i)$, $i = 1, 2, \cdots, n$, are the observations, and I_i is as in (2.1).

Step 1. For each $j = 1, 2, \cdots, J$, construct G_j, which is a collection of all I_i's with $j \in \mathcal{M}_i$. Notice that G_j's are not a partition.

Step 2. For each $j = 1, 2, \cdots, J$, find the MI's induced by the I_i's in G_j. In particular, denote $G_j = \{I_{j,i}, i = 1, 2, \cdots, n_j\}$, where $I_{j,i} = \begin{cases} (l_{j,i}, r_{j,i}] \times \mathcal{M}_{j,i} & \text{if } l_{j,i} < r_{j,i} \\ [l_{j,i}, r_{j,i}] \times \mathcal{M}_{j,i} & \text{if } l_{j,i} = r_{j,i} \end{cases}$, and proceed in two sub-steps as follows.

Step 2.1 (find the MI's induced by the interval part of $I_{j,i}$'s (with j fixed)). Let $Ends[j]$ be a collection of $2n_j$ elements such that each $I_{j,i}$ induces two elements $\mathbf{x}_{j,2i-1}$ and $\mathbf{x}_{j,2i}$ of $Ends[j]$, namely, $\mathbf{x}_{j,2i-1} = (l_{j,i}, \mathcal{M}_{j,i}, 0, \mathrm{EI})$ and $\mathbf{x}_{j,2i} = (r_{j,i}, \mathcal{M}_{j,i}, 1, \mathrm{EI})$, where $EI = \mathbf{1}(l_{j,i} = r_{j,i})$. Notice that the third component, say SI, in $\mathbf{x}_{j,k}$ is the side indicator whether its first component is a right end-point of $I_{j,i}$.

Let $OrderedEnds[j]$ be the set of ordered statistics in the first coordinates of the elements in the set $Ends[j]$. The ordering is in the sense as follows. For each pair of elements in $Ends[j]$, if their first coordinates are not equal, then order them in an obvious way; if they are equal but SI's are not, then check the exact indicators: if both EI $= 1$, then the element with SI $= 1$ is assumed to be bigger; otherwise, the element with SI $= 1$ is assumed to be smaller; finally, if x and y in the set $Ends[j]$ are the same, order them according to their indices. Denote $OrderedEnds[j] = \{(e_{j,i}, \mathcal{M}_{j,i}, k_i) : i = 1, 2, \cdots, 2n_j, k_i \in \{0, 1\}\}$, where the masking parts $\mathcal{M}_{j,i}$ and the side indicators k_i go with the ordering.

Then we construct a new collection MG_j as follows.

1. Initialize $i = 1$ and $MG_j = \emptyset$.

2. If i reaches $2n_j$, then stops; otherwise, pick the pair $\{(e_{j,i},\ \mathcal{M}_{j,i},\ k_i),\ (e_{j,i+1},\ \mathcal{M}_{j,i+1},\ k_{i+1})\}$ from $OrderedEnds[j]$.

3. Check whether $(k_i, k_{i+1}) = (0, 1)$. If it is true, then go to next step; otherwise, go back to 2 with i replaced by $i + 1$.

4. Let $\mathcal{M}^* = \mathcal{M}_{j,i} \bigcap \mathcal{M}_{j,i+1}$, add the intersection element $(e_{j,i}, e_{j,i+1}] \times \mathcal{M}^*$ to MG_j, replace i by $i+1$, then go back to 2.

The elements in MG_j's are candidates of MI's.

Step 2.2 (narrow down the masking part for each element in MG_j). First denote $MG_j = \{(e_{l,i}, e_{r,i}] \times \mathcal{M}_i^*, i = 1, 2, \cdots, m_j\}$,

0. Initialize $i = 1$ and $IMG_j = \emptyset$.

1. If i reaches $m_j + 1$, then stop; otherwise, pick the i^{th} element, say $(e_{l,i}, e_{r,i}] \times \mathcal{M}_i^*$ from MG_j. If $|\mathcal{M}_i^*| > 1$, go to next step to compare with the n_j elements $(l_k, r_k] \times \mathcal{M}_k$ in G_j; otherwise, add $(e_{l,i}, e_{r,i}] \times \mathcal{M}_i^*$ to IMG_j, replace i by $i + 1$ and go back to 1. Here $|.|$ denotes the size of the set.

2. Initialize $k = 1$.

3. If k reaches $n_j + 1$, then add $(e_{l,i}, e_{r,i}] \times \mathcal{M}_i^*$ to IMG_j, go back to 1 with i replaced by $i+1$; otherwise, pick an element $(l_k, r_k] \times \mathcal{M}_k$ from G_j, if $(e_{l,i}, e_{r,i}] \subset (l_k, r_k]$, then go to next step; otherwise, go back to 3 with k replaced by $k + 1$.

4. If $|\mathcal{M}_i^* \bigcap \mathcal{M}_k| < |\mathcal{M}_i^*|$, replace $\mathcal{M}_i^*$ by $\mathcal{M}_i^* \bigcap \mathcal{M}_k$, then go to next step; otherwise, go back to 3 with k replaced by $k + 1$.

5. If $|\mathcal{M}_i^*| = 1$, then add $(e_{l,i}, e_{r,i}] \times \mathcal{M}_i^*$ to IMG_j, go back to 1 with i replaced by $i + 1$; otherwise, go to 3 with k replaced by $k + 1$.

Now IMG_j is the set of all MI's induced by I_i's in G_j (see §4 in Appendix III), $j = 1, ..., J$. Notice that these MI's are not the true MI's induced by all I_i's.

Step 3 (find the true MI's). The true MI's induced by I_i's can be obtained through the following sub-steps.

0. Fix a $j \in \{1, ..., J\}$, initialize $k = 1$, $i = 1$.

1. If k reaches $|IMG_j| + 1$, then stops. Otherwise, pick an element $(e_{l,k}, e_{r,k}] \times \mathcal{M}_k^*$ from IMG_j. If $|\mathcal{M}_k^*| = 1$ or $\mathcal{M}_k^* = \emptyset$, then go back to 1 with k replaced by $k + 1$; otherwise, go to next step.

2. If i reaches $|\bigcup_{s \in \mathcal{M}_k^* \setminus \{j\}} IMG_s| + 1$, then go back to 1 with k replaced by $k + 1$ and i replaced by 1. Otherwise, pick an element $(e_{l,i}, e_{r,i}] \times \mathcal{M}_i^*$ from $\bigcup_{s \in \mathcal{M}_i^* \setminus \{j\}} IMG_s$ with $(e_{l,i}, e_{r,i}] \times \mathcal{M}_i^* \subset IMG_s$ for some $s \in \mathcal{M}_i^* \setminus \{j\}$. If $e_{l,k} < e_{l,i} \le e_{r,k} \le e_{r,i}$ or $e_{l,k} \le e_{l,i} \le e_{r,k} < e_{r,i}$ or $e_{l,i} < e_{l,k} \le e_{r,i} \le e_{r,k}$ or $e_{l,i} \le e_{l,k} \le e_{r,i} < e_{r,k}$, then replace $(e_{l,k}, e_{r,k}] \times \mathcal{M}_k^*$ by $(e_{l,k}, e_{r,k}] \times \emptyset$, replace $(e_{l,i}, e_{r,i}] \times \mathcal{M}_i^*$ by $(e_{l,i}, e_{r,i}] \times \emptyset$, and go back to 1 with k replaced by $k + 1$ and i replaced by 1; otherwise, go to next step.

3. If $[(e_{l,i}, e_{r,i}] \times \mathcal{M}_i^*] \subset (e_{l,k}, e_{r,k}] \times \mathcal{M}_k^*$, replace $(e_{l,k}, e_{r,k}] \times \mathcal{M}_k^*$ by $(e_{l,k}, e_{r,k}] \times \emptyset$, then go back to 1 with k replaced by $k+1$ and i replaced by 1; otherwise, go to next step.

4. If $[(e_{l,k}, e_{r,k}] \times \mathcal{M}_k^*] \subset (e_{l,i}, e_{r,i}] \times \mathcal{M}_i^*$, replace $(e_{l,i}, e_{r,i}] \times \mathcal{M}_i^*$ by $(e_{l,i}, e_{r,i}] \times \emptyset$, then go back to 2 with i replaced by $i+1$; otherwise, go to next step.

5. The only possible condition now is that these two elements are disjoint, thus go back to 2 with i replaced by $i+1$.

Repeat this procedure for $j = 1, ..., J$, set $FMG = \bigcup_{j=1}^{J} \{(e_{l,i}, e_{r,i}] \times \mathcal{M}_i^* \in IMG_j : \mathcal{M}_i^* \neq \emptyset\}$, which is the set of true MI's induced by all I_i's (see §5 in Appendix III). □

Example I. We shall explain this algorithm by a simple example. Assume that $J = 3$, and there are 5 observations: $O_1 = (2, 4] \times \{1, 2\}$, $O_2 = (1, 5] \times \{1\}$, $O_3 = (3, 5] \times \{2, 3\}$, $O_4 = (-\infty, 2] \times \{1, 2\}$, and $O_5 = (5, +\infty) \times \{1, 2, 3\}$.

Step 1. Construct G_j's as follows.

$G_1 = \{(2, 4] \times \{1, 2\}, (1, 5] \times \{1\}, (-\infty, 2] \times \{1, 2\}, (5, \infty) \times \{1, 2, 3\}\}$;

$G_2 = \{(2, 4] \times \{1, 2\}, (3, 5] \times \{2, 3\}, (-\infty, 2] \times \{1, 2\}, (5, \infty) \times \{1, 2, 3\}\}$;

$G_3 = \{(3, 5] \times \{2, 3\}, (5, \infty) \times \{1, 2, 3\}\}$.

Step 2. Find the MI's induced by I_i's in each group G_j. Here we consider group G_1 only for illustration purpose. This step has two sub-steps.

Step 2.1 Find the MI's w.r.t. the interval part.

1. Construct $Ends[1]$ from G_1: $Ends[1] = \{(2, \{1, 2\}, 0), (4, \{1, 2\}, 1), (1, \{1\}, 0), (5, \{1\}, 1), (-\infty, \{1, 2\}, 0), (2, \{1, 2\}, 1), (5, \{1, 2, 3\}, 0), (\infty, \{1, 2, 3\}, 1)\}$.

2. Order the endpoints in $Ends[1]$ as follows.

$OrderEnds[1] = \{(-\infty, \{1, 2\}, 0), (1, \{1\}, 0), (2, \{1, 2\}, 1), (2, \{1, 2\}, 0), (4, \{1, 2\}, 1), (5, \{1\}, 1), (5, \{1, 2, 3\}, 0), (\infty, \{1, 2, 3\}, 1)\}$,

3. Obtain the set of initial MI's induced by $I_{1,i}$'s. $MI_1 = \{(1, 2] \times \{1\}, (2, 4] \times \{1, 2\}, (5, \infty) \times \{1, 2, 3\}\}$

Step 2.2. Narrow down the masked competing risks in MI_1 to obtain:

$IMG_1 = \{(1, 2] \times \{1\}, (2, 4] \times \{1\}, (5, \infty) \times \{1, 2, 3\}\}$ since $(2, 4] \subset (1, 5]$ and $\{1\} \subset \{1, 2\}$.

By the same way, we can find the other two MI's for G_1 and G_2 as the following:

$IMG_2 = \{(-\infty, 2] \times \{1, 2\}, (3, 4] \times \{2\}, (5, \infty) \times \{1, 2, 3\}\}$,

$IMG_3 = \{(3, 5] \times \{2, 3\}, (5, \infty) \times \{1, 2, 3\}\}$.

Step 3. Find the true MI's induced by all I_i's.

Pick each elements from $IMG_1 \bigcup IMG_2 \bigcup IMG_3$, if the masked competing risks (MCR) part has only one element, then put it into the final MI set denoted by FMG; otherwise check this element with other elements in $IMG_1 \bigcup IMG_2 \bigcup IMG_3$ to see whether it is smallest with rules given in the algorithm. Then IMG_j becomes

$IMG_1 = \{(1, 2] \times \{1\}, (2, 4] \times \{1\}, (5, \infty) \times \emptyset\}$,

$IMG_2 = \{(-\infty, 2] \times \emptyset, (3, 4] \times \{2\}, (5, \infty) \times \emptyset\}$,

$IMG_3 = \{(3, 5] \times \emptyset, (5, \infty) \times \{1, 2, 3\}\}$.

Then by picking up the elements from $\bigcup_{j=1}^{3} IMG_j$ with $MCR \neq \emptyset$, we can obtain the MI's: $FMG = \{(1, 2] \times \{1\}, (2, 4] \times \{1\}, (3, 4] \times \{2\}, (5, \infty) \times \{1, 2, 3\}\}$. □

Appendix II. Tables.

Table 1. Continuous inspection times with log-normal survival distribution (Example 3.1).

Risk 1	t=3.0124	t=5	t=8.2990	t=14.271
True Value	0.01	0.0125	0.015	0.175
n=100	0.0097(0.0114)	0.0124(0.0127)	0.0162(0.0155)	0.0220(0.0265)
n=200	0.0095(0.0082)	0.0120(0.0089)	0.0153(0.0106)	0.0197(0.0173)
n=400	0.0098(0.0058)	0.0122(0.0062)	0.0154(0.0075)	0.0188(0.0113)
n=800	0.0100(0.0043)	0.0122(0.0045)	0.0148(0.0051)	0.0182(0.0082)
Risk 2	t=3.419	t=5	t=7.3116	t=10.98
True Value	0.01	0.0125	0.015	0.0175
n=100	0.0082(0.0112)	0.0115(0.0126)	0.0142(0.0143)	0.0191(0.0227)
n=200	0.0097(0.0084)	0.0125(0.0089)	0.0153(0.01)	0.019(0.0138)
n=400	0.0096(0.0058)	0.0119(0.0063)	0.0144(0.0072)	0.0184(0.0103)
n=800	0.01(0.0044)	0.0121(0.0044)	0.0145(0.0051)	0.0178(0.0069)
Risk 3	t=3.881	t=5	t=6.442	t=8.447
True Value	0.36	0.45	0.54	0.63
n=100	0.3665(0.0647)	0.4466(0.0673)	0.5315(0.0609)	0.6236(0.0648)
n=200	0.365(0.0492)	0.4489(0.0479)	0.5337(0.0485)	0.6234(0.0486)
n=400	0.365(0.0351)	0.4506(0.0359)	0.5345(0.0347)	0.6259(0.0359)
n=800	0.3638(0.0266)	0.4503(0.0259)	0.5352(0.0257)	0.6255(0.0276)
Risk 4	t=4.405	t= 5	t=5.675	t=6.499
True Value	0.02	0.025	0.03	0.035
n=100	0.0172(0.0184)	0.0228(0.0204)	0.03(0.0218)	0.0338(0.0224)
n=200	0.0186(0.0135)	0.0246(0.0154)	0.0304(0.0161)	0.0342(0.0168)
n=400	0.0184(0.0099)	0.0246(0.0109)	0.0309(0.0111)	0.0343(0.0113)
n=800	0.019(0.0073)	0.0249(0.0078)	0.0309(0.0087)	0.0341(0.0092)

Table 2. Continuous inspection times with exponential survival distribution (Example 3.2).

Risk 1	t=0.5108	t=0.6931	t=0.9163	t=1.204
True Value	0.01	0.0125	0.015	0.0175
n=100	0.0013(0.0068)	0.0023(0.0088)	0.0038(0.0113)	0.0054(0.013)
n=200	0.0017(0.0069)	0.0035(0.0096)	0.0059(0.0121)	0.0083(0.0137)
n=400	0.0035(0.0084)	0.0055(0.0103)	0.0088(0.0123)	0.0126(0.0129)
n=800	0.0044(0.0084)	0.0075(0.0101)	0.0109(0.0108)	0.0147(0.0108)
Risk 2	t=1.022	t=1.386	t=1.833	t=2.408
True Value	0.01	0.0125	0.015	0.0175
n=100	0.0017(0.0076)	0.0030(0.0097)	0.0047(0.0122)	0.0077(0.0147)
n=200	0.0026(0.008)	0.0047(0.0104)	0.0077(0.0124)	0.0106(0.0136)
n=400	0.0033(0.0076)	0.0061(0.0097)	0.0096(0.0114)	0.0137(0.0117)
n=800	0.0057(0.0079)	0.0092(0.009)	0.0125(0.0091)	0.0163(0.0084)
Risk 3	t=2.0433	t=2.7726	t=3.6652	t=4.8159
True Value	0.36	0.45	0.54	0.63
n=100	0.3403(0.1204)	0.4416(0.1118)	0.5390(0.0993)	0.6352(0.0795)
n=200	0.3552(0.0906)	0.4474(0.0824)	0.5382(0.0738)	0.6309(0.0608)
n=400	0.3559(0.0686)	0.4474(0.0612)	0.5419(0.056)	0.6332(0.0442)
n=800	0.3611(0.0543)	0.4510(0.0474)	0.5414(0.041)	0.6319(0.0337)
Risk 4	t=4.087	t=5.5452	t=7.3303	t=9.6318
True Value	0.02	0.025	0.03	0.035
n=100	0.0167(0.0199)	0.02377(0.022)	0.0301(0.0219)	0.0356(0.0216)
n=200	0.0176(0.0152)	0.0243(0.0153)	0.0295(0.0146)	0.0346(0.0143)
n=400	0.0188(0.0114)	0.0246(0.0109)	0.0302(0.0104)	0.0349(0.0106)
n=800	0.0200(0.0079)	0.0250(0.0076)	0.0304(0.0072)	0.0349(0.0073)

Table 3. Continuous inspection times with uniform survival distribution (Example 3.3).

Risk 1	t=1.6	t=3.2	t=4.8	t=6.4
True Value	0.005	0.01	0.015	0.02
n=100	0.0039(0.0077)	0.0097(0.0121)	0.0147(0.0146)	0.0211(0.0172)
n=200	0.0045(0.0065)	0.0097(0.0088)	0.0147(0.0107)	0.0217(0.0125)
n=400	0.0049(0.005)	0.0097(0.0064)	0.0144(0.0076)	0.0212(0.0092)
n=800	0.0051(0.0038)	0.0102(0.0047)	0.0147(0.0056)	0.0205(0.0063)
Risk 2	t=1.4	t=2.8	t=4.2	t=5.6
True Value	0.005	0.01	0.015	0.02
n=100	0.0028(0.0073)	0.0078(0.0116)	0.0132(0.0145)	0.0183(0.0174)
n=200	0.0033(0.0059)	0.0086(0.0088)	0.0145(0.0106)	0.0200(0.0121)
n=400	0.0038(0.0048)	0.0090(0.0061)	0.0146(0.0072)	0.0203(0.0085)
n=800	0.0047(0.0038)	0.0096(0.0047)	0.0145(0.0053)	0.0205(0.0064)
Risk 3	t=1.2	t=2.4	t=3.6	t=4.8
True Value	0.18	0.36	0.54	0.72
n=100	0.1783(0.0657)	0.3387(0.0726)	0.5490(0.0761)	0.7121(0.0762)
n=200	0.1811(0.0494)	0.3532(0.0587)	0.5443(0.057)	0.7131(0.0553)
n=400	0.1774(0.039)	0.3484(0.0439)	0.5463(0.0447)	0.7161(0.0416)
n=800	0.1782(0.0289)	0.3549(0.0333)	0.5441(0.0334)	0.7180(0.0325)
Risk 4	t=1.0	t=2.0	t=3.0	t=4.0
True Value	0.01	0.02	0.03	0.04
n=100	0.0065(0.0121)	0.0181(0.018)	0.0284(0.0215)	0.0409(0.0239)
n=200	0.0081(0.0108)	0.0197(0.0136)	0.0293(0.0158)	0.0417(0.0174)
n=400	0.0087(0.0083)	0.0196(0.0095)	0.0299(0.0116)	0.0421(0.0127)
n=800	0.0094(0.0063)	0.0205(0.007)	0.0297(0.0083)	0.0414(0.0089)

Table 4. Discrete Inspection Times with log-normal survival distribution (Example 3.4).

Risk 1	t=3.0124	t=5	t=8.299	t=14.271
True Value	0.01	0.0125	0.015	0.175
n=100	0.0105(0.0108)	0.0131(0.0125)	0.0164(0.0147)	0.0191(0.016)
n=200	0.0098(0.0075)	0.0122(0.0085)	0.0149(0.0098)	0.0177(0.0111)
n=400	0.0103(0.0056)	0.0130(0.0063)	0.0158(0.007)	0.0184(0.0079)
n=800	0.0098(0.0038)	0.0124(0.0044)	0.0150(0.005)	0.0176(0.0056)
Risk 2	t=3.4192	t=5	t=7.3116	t=10.98
True Value	0.01	0.0125	0.015	0.0175
n=100	0.0093(0.0106)	0.0119(0.0119)	0.0146(0.0136)	0.0168(0.0152)
n=200	0.0096(0.0074)	0.0119(0.0084)	0.0143(0.0094)	0.0166(0.0106)
n=400	0.0098(0.0053)	0.0121(0.0061)	0.0145(0.0069)	0.0168(0.0077)
n=800	0.0101(0.0037)	0.0126(0.0042)	0.0150(0.0049)	0.0176(0.0054)
Risk 3	t=3.881	t=5	t=6.4417	t=8.4472
True Value	0.36	0.45	0.54	0.63
n=100	0.3594(0.0479)	0.4499(0.053)	0.5404(0.0556)	0.6303(0.0593)
n=200	0.3609(0.0346)	0.4503(0.0379)	0.5406(0.0396)	0.6303(0.0409)
n=400	0.3605(0.0244)	0.4499(0.0262)	0.5399(0.0269)	0.6294(0.0283)
n=800	0.3601(0.0173)	0.4498(0.0182)	0.5404(0.0189)	0.6304(0.0194)
Risk 4	t=4.4051	t= 5	t=5.6752	t=6.4989
True Value	0.02	0.025	0.03	0.035
n=100	0.0201(0.0151)	0.0250(0.0172)	0.0297(0.0195)	0.0347(0.0213)
n=200	0.0204(0.0112)	0.0254(0.0126)	0.0304(0.0145)	0.0354(0.0156)
n=400	0.0199(0.008)	0.0248(0.0091)	0.0300(0.01)	0.0348(0.0108)
n=800	0.0199(0.0058)	0.0249(0.0064)	0.0297(0.0071)	0.0345(0.0077)

Table 5. The estimator's difference between the two estimating approaches.

Risk 1	0.01	0.0125	0.015	0.0175
n=100	0.0184%	0.0088%	0.0222%	0.0588%
n=200	0.0095%	0.0073%	0.0076%	0.0409%
n=400	0.0035%	0.0026%	0.0024%	0.0195%
n=800	0.0005%	0.0006%	0.0002%	0.004%
Risk 2	0.01	0.0125	0.015	0.0175
n=100	0.0068%	0.0054%	0.0226%	0.0226%
n=200	0.002 %	0.0025%	0.003%	0.0167%
n=400	0.0007%	0.0016%	0.0026%	0.0187%
n=800	0.001%	0.0003%	0.0004%	0.0053%
Risk 3	0.36	0.45	0.54	0.63
n=100	0.048%	0.0047%	0.0054%	0.0073%
n=200	0.0021%	0.0026%	0.003%	0.0049%
n=400	0.0011%	0.0013%	0.0015%	0.0029%
n=800	0.0001%	0.0001%	0.0001%	0.0002%
Risk 4	0.02	0.025	0.03	0.035
n=100	0.0195%	0.0145%	0.0114%	0.0099%
n=200	0.0092%	0.0059%	0.0061%	0.005%
n=400	0.0029%	0.0037%	0.0029%	0.0028%
n=800	0.0014%	0.0003%	0.0001%	0.0001%

Example 3.1: Continuous Inspection Times with Log-normal Distributions.

Table 6. The estimator's difference between the two estimating approaches.

Risk 1	0.01	0.0125	0.015	0.0175
n=100	0.4765%	2.0849 %	5.2561 %	6.1354 %
n=200	0.0756%	0.2992%	2.0916%	2.173%
n=400	0.0118%	0.1366%	0.2395%	0.3477%
n=800	0.0032%	0.0792%	0.0769%	0.1404%
Risk 2	0.01	0.0125	0.015	0.0175
n=100	1.5985%	2.5269 %	2.9074%	2.9642%
n=200	0.4713%	1.2595%	1.6482%	1.9564%
n=400	0.433%	0.6464%	0.8852%	1.2499%
n=800	0.0922%	0.1632%	0.2996%	0.4908%
Risk 3	0.36	0.45	0.54	0.63
n=100	0.1387%	0.145%	0.1522%	0.3052 %
n=200	0.0281%	0.0238%	0.0225%	0.139%
n=400	0.0022%	0.0016%	0.0012%	0.0447%
n=800	0.0001%	0.0002%	0.0003%	0.0162%
Risk 4	0.02	0.025	0.03	0.035
n=100	0.4002 %	0.3021%	0.1096%	0.211%
n=200	0.261%	0.3662%	0.3703%	0.1812%
n=400	0.2188%	0.2144%	0.2%	0.0456%
n=800	0.0449%	0.0463%	0.0489%	0.0157%

Example 3.4: Discrete Inspection Times with Log-normal Distribution.

Table 7. The time's difference between the two estimating approaches.

Size	Time of 1st Approach	Time of 2nd Approach	Ratio
n=100	85.684	124.692	0.687
n=200	1120.276	1462.347	0.766
n=400	11378.34	13632.84	0.835
n=800	126122.0	138022.9	0.914

Example 3.1: Log-normal Distribution. Tolerance=0.0001.

Size	Time of 1st Approach	Time of 2nd Approach	Ratio
n=100	4.361	7.708	0.566
n=200	9.661	19.894	0.486
n=400	32.953	76.164	0.433
n=800	93.624	162.833	0.575

Example 3.4: Discrete Inspection Times. Tolerance=0.00001.

Table 8. Empirical convergent rates for the continuous inspection times.

Risk 1	t=3.01242	t=5	t=8.298976	t=14.27113
$\frac{\hat{SE}_{200}}{\hat{SE}_{100}}$	0.7142	0.7018	0.6869	0.6510
$\frac{\hat{SE}_{400}}{\hat{SE}_{200}}$	0.7155	0.6922	0.7005	0.6524
$\frac{\hat{SE}_{800}}{\hat{SE}_{400}}$	0.7325	0.7278	0.6895	0.7245
Risk 2	t=3.419236	t=5	t=7.3115745	t=10.97960
$\frac{\hat{SE}_{200}}{\hat{SE}_{100}}$	0.7479	0.7057	0.7017	0.6068
$\frac{\hat{SE}_{400}}{\hat{SE}_{200}}$	0.6942	0.7053	0.7212	0.7458
$\frac{\hat{SE}_{800}}{\hat{SE}_{400}}$	0.7501	0.7000	0.6997	0.6661
Risk 3	t=3.880992	t=5	t=6.441652	t=8.447229
$\frac{\hat{SE}_{200}}{\hat{SE}_{100}}$	0.7601	0.7112	0.7955	0.7494
$\frac{\hat{SE}_{400}}{\hat{SE}_{200}}$	0.7128	0.7501	0.7166	0.7383
$\frac{\hat{SE}_{800}}{\hat{SE}_{400}}$	0.7584	0.7214	0.7404	0.7680
Risk 4	t=4.405106	t=5	t=5.675232	t=6.498934
$\frac{\hat{SE}_{200}}{\hat{SE}_{100}}$	0.7340	0.7569	0.7348	0.7505
$\frac{\hat{SE}_{400}}{\hat{SE}_{200}}$	0.7298	0.7078	0.6933	0.6761
$\frac{\hat{SE}_{800}}{\hat{SE}_{400}}$	0.7371	0.7110	0.7776	0.8145

Example 3.1: Ratios in Log-normal Distributions.

Table 9. Empirical convergent rates for the continuous inspection times.

Risk 1	t=0.5108256	t=0.6931472	t=0.9162907	t=1.2039733
$\frac{\hat{SE}_{200}}{\hat{SE}_{100}}$	1.0169	1.0877	1.0782	1.0493
$\frac{\hat{SE}_{400}}{\hat{SE}_{200}}$	1.2045	1.0722	1.0112	0.9423
$\frac{\hat{SE}_{800}}{\hat{SE}_{400}}$	1.0101	0.9805	0.8835	0.8378
Risk 2	t=1.021651	t=1.386294	t=1.832581	t=2.407946
$\frac{\hat{SE}_{200}}{\hat{SE}_{100}}$	1.0510	1.0638	1.0198	0.9208
$\frac{\hat{SE}_{400}}{\hat{SE}_{200}}$	0.9510	0.9387	0.9177	0.8648
$\frac{\hat{SE}_{800}}{\hat{SE}_{400}}$	1.0434	0.9296	0.8021	0.7113
Risk 3	t=2.043302	t=2.772589	t=3.665163	t=4.815891
$\frac{\hat{SE}_{200}}{\hat{SE}_{100}}$	0.7529	0.7365	0.7430	0.7643
$\frac{\hat{SE}_{400}}{\hat{SE}_{200}}$	0.7568	0.7429	0.7595	0.7273
$\frac{\hat{SE}_{800}}{\hat{SE}_{400}}$	0.7922	0.7750	0.7321	0.7626
Risk 4	t=4.086605	t=5.545177	t=7.330326	t=9.631782
$\frac{\hat{SE}_{200}}{\hat{SE}_{100}}$	0.7629	0.6973	0.6649	0.6610
$\frac{\hat{SE}_{400}}{\hat{SE}_{200}}$	0.7475	0.7103	0.7173	0.7443
$\frac{\hat{SE}_{800}}{\hat{SE}_{400}}$	0.6948	0.6959	0.6888	0.6859

Example 3.2: Ratios in Exponential Distributions.

Table 10. Empirical convergent rates for inspection times with the uniform distribution (Ex. 3.).

Risk 1	t=1.6	t=3.2	t=4.8	t=6.4
$\frac{\hat{SE}_{200}}{\hat{SE}_{100}}$	0.8486	0.7251	0.7348	0.7294
$\frac{\hat{SE}_{400}}{\hat{SE}_{200}}$	0.7733	0.7298	0.7058	0.7385
$\frac{\hat{SE}_{800}}{\hat{SE}_{400}}$	0.7560	0.7327	0.7384	0.6798
Risk 2	**t=1.4**	**t=2.8**	**t=4.2**	**t=5.6**
$\frac{\hat{SE}_{200}}{\hat{SE}_{100}}$	0.8132	0.7614	0.7274	0.6951
$\frac{\hat{SE}_{400}}{\hat{SE}_{200}}$	0.8063	0.6900	0.6836	0.7045
$\frac{\hat{SE}_{800}}{\hat{SE}_{400}}$	0.7978	0.7636	0.7287	0.7517
Risk 3	**t=1.6**	**t=2.4**	**t=3.2**	**t=4.8**
$\frac{\hat{SE}_{200}}{\hat{SE}_{100}}$	0.7514	0.8077	0.7492	0.7257
$\frac{\hat{SE}_{400}}{\hat{SE}_{200}}$	0.7888	0.7492	0.7841	0.7529
$\frac{\hat{SE}_{800}}{\hat{SE}_{400}}$	0.7429	0.7583	0.7468	0.7813
Risk 4	**t=1.0**	**t=2.0**	**t=3.0**	**t=4.0**
$\frac{\hat{SE}_{200}}{\hat{SE}_{100}}$	0.8887	0.7570	0.7318	0.7296
$\frac{\hat{SE}_{400}}{\hat{SE}_{200}}$	0.7703	0.7010	0.7352	0.7266
$\frac{\hat{SE}_{800}}{\hat{SE}_{400}}$	0.7617	0.7356	0.7153	0.7030

Table 11. Empirical Convergent Rates for the discrete inspection times and continuous survival time (Example 3.4).

Risk 1	t=3.01242	t=5	t=8.298976	t=14.27113
$\frac{\hat{SE}_{200}}{\hat{SE}_{100}}$	0.6887	0.6815	0.6659	0.6961
$\frac{\hat{SE}_{400}}{\hat{SE}_{200}}$	0.7465	0.7455	0.7110	0.7078
$\frac{\hat{SE}_{800}}{\hat{SE}_{400}}$	0.6807	0.6971	0.7209	0.7140
Risk 2	t=3.419236	t=5	t=7.3115745	t=10.97960
$\frac{\hat{SE}_{200}}{\hat{SE}_{100}}$	0.6951	0.7026	0.6894	0.6991
$\frac{\hat{SE}_{400}}{\hat{SE}_{200}}$	0.7263	0.7262	0.7312	0.7244
$\frac{\hat{SE}_{800}}{\hat{SE}_{400}}$	0.6845	0.6880	0.7098	0.7064
Risk 3	t=3.880992	t=5	t=6.441652	t=8.447229
$\frac{\hat{SE}_{200}}{\hat{SE}_{100}}$	0.7214	0.7155	0.7128	0.6901
$\frac{\hat{SE}_{400}}{\hat{SE}_{200}}$	0.7050	0.6911	0.6777	0.6915
$\frac{\hat{SE}_{800}}{\hat{SE}_{400}}$	0.7078	0.6944	0.7048	0.6861
Risk 4	t=4.405106	t= 5	t=5.675232	t=6.498934
$\frac{\hat{SE}_{200}}{\hat{SE}_{100}}$	0.7437	0.7329	0.7419	0.7308
$\frac{\hat{SE}_{400}}{\hat{SE}_{200}}$	0.7124	0.7243	0.6898	0.6951
$\frac{\hat{SE}_{800}}{\hat{SE}_{400}}$	0.7264	0.7043	0.7100	0.7096

REFERENCES CITED

* Basu, S., Basu, A. P., Mukhopadhyay, C. (1999). Bayesian analysis for masked system failure data using non-identical Weibull models. *Journal of Statistical Planning and Inference* 78 255-275.
* Basu, S., Sen, A., Banerjee, M. (2003). Bayesian analysis of competing risks with partially masked cause of failure. *Applied Statistics* 52 77-93.
* Craiu, R. V. and Duchesne, T. (2004). Inference based on the EM algorithm for the competing risks model with masked causes of failure. *Biometrika* 91 543-558.
* Craiu, R. V. and Reiser, B. (2006). Inference for the dependent competing risks model with masked causes of failure. *Lifetime Data Analysis* 21-33.
* Dewanji, A. and Sengupta, D. (2003). Estimation of competing risks with general missing pattern in failure types. *Biometrics* 59 1063-1070
* Dinse, G. E. (1982). Nonparametric estimation for partially-complete time and type of failure data. *Biometrics* 38 417-431.
* Dinse, G. E. (1986). Nonparametric prevalence and mortality estimators for animal experiments with incomplete cause-of-death data. *Journal of the American Statistical Association* 81 328-336.
* Flehinger, B. J., Reiser, B., and Yashchin, E. (1996). Inference about defects in the presence of masking. *Technometrics* 38 247-255.
* Flehinger, B. J., Reiser, B., and Yashchin, E. (1998). Survival with competing risks and masked causes of failures. *Biometrika* 85 151-164.
* Flehinger, B. J., Reiser, B., and Yashchin, E. (2001). Statistical analysis for masked data. In: Balakrishnan N, Rao CR (eds) *Handbook of statistics* 20 499-522. Elsevier Science.
* Groeneboom, P. and Wellner, J. A. (1992). Information bounds and nonparametric maximum likelihood estimation. *Birkhäuser Verlag, Basel*.
* Groeneboom, P. (1996). Lecture on inverse problems. In P.Bernard.(Ed.). *Lecture on probability and statistics.* p.157. Berlin, Springer-Verlag.

* Groeneboom, P., Jongbloed, G. and Wellner, J. A. (2008a). The support reduction algorithm for computing nonparametric function estimates in mixture models. *Scandinavian Journal of Statistics* 35 385-399.
* Groeneboom, P., Maathuis, M. H. and Wellner, J. A. (2008b). Current status data with competing risks: consistency and rates of convergence of the MLE. *Annals of Statistics* 36 1031-1063.
* Guess, F. M., Usher, J. S. and Hodgson, T. J. (1991). Estimating system and component reliabilities under partial information on cause of failure. *Journal of Statistical Planning and Inference* 29 75-85.
* Hudgens, M. G., Satten, G. A., and Longini, I. M., Jr. (2001). Nonparametric maximum likelihood estimation for competing risks survival data subject to interval censoring and truncation. *Biometrics* 57 74-80.
* Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Annals of Statistics*, 27 887-906.
* Lin, D. K. J., Usher, J. S., and Guess, F. M. (1996). Bayes estimation of component-reliability from masked system-life data. *IEEE Transactions on Reliability* R-45 233-237.
* Maathuis, M. H. (2005). Reduction algorithm for the MLE for the distribution function of bivariate interval censored data. *Journal of Computing and Graphical Statistics* 14 352-362.
* Maathuis, M.H. and Wellner, J.A.(2008). Inconsistency of the MLE for the joint distribution of interval censored survival times and continuous marks. *Scandinavian Journal of Statistics* 35 83-103.
* Miyakawa, M. (1984). Analysis of incomplete data in competing risk model. *IEEE Transactions on Reliability* 33 293-296.
* Mukhopadhyay, C. (2006). Maximum likelihood analysis of masked series system lifetime data. *Journal of Statistical Planning and Inference* 136 803-838.
* Mukhopadhyay, C., and Basu, A. P. (1997). Bayesian analysis of incomplete time and cause of failure data. *Journal of Statistical Planning and Inference* 59 79-100.
* Mukhopadhyay C., and Basu S. (2007). Bayesian analysis of masked series system lifetime data. *Communications in Statistics - Theory And Methods* 36 329-348.
* Peterson, A. V., Jr. (1977). Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions. *Journal of the American Statistical Association* 72 854-858.
* Reiser, B., Guttman, I., Lin, D. K. J., Usher, J. S. and Guess, F. M. (1995). Bayesian inference for masked system life time data. *Applied Statistics* 44 79-90.
* Schick, A. and Yu, Q. (2000) Consistency of the GMLE with mixed case interval-censored data. *Scandinavian Journal of Statistics* 27 45-55.
* Sen, A., Basu, S. and Benerjee, M. (2001). Analysis of masked failure time data under competing risks. In: Balakrishnan N, Rao CR(eds) *Handbook of statistics* 20 523-540. Elsevier Science.
* Turnbull, B. W. (1976). The empirical distribution function with arbitrary grouped, censored and truncated data. *Journal of the Royal Statistical Society Series B* 38 290-295.
* Wong, G. Y. C. and Yu, Q. (1999). Generalized MLE of a joint distribution function with multivariate interval-censored data. *Journal of Multivariate Analysis* 69 155-166.
* Yu, Q., Schick, A., Li, L. and Wong, G. Y. C. (1998a). Asymptotical properties of the GMLE in the case 1 interval censorship model with discrete inspection times. *Canadian Journal of Statistics* 26 619-627.
* Yu, Q., Schick, A., Li, L. and Wong, G. Y. C. (1998b). Asymptotical properties of the GMLE in the case 2 interval-censored data. *Statistics & Probability Letters* 37 223-228.
* Yu, S., Yu, Q. and Wong, G. Y. C. (2006). Consistency of the generalized MLE with multivariate mixed case interval-censored data. *Journal of Multivariate Analysis* 97 720-732.

This appendix may become a technical report for shortening the paper

Appendix III We shall give the proofs of the statements and theorems in the paper.

§**1. Proposition 1.** The NPMLE of F_0 assigns weights, say, s_1, s_2, $\cdots$, s_m only to the corresponding MI's $A_1, A_2, \cdots, A_m$.

Proof: It suffices to show that if B is a set such that $B \cap (\cup_{k=1}^m A_k) = \emptyset$ and $\mathbf{F}^s$ is a scdf vector such that $\mu_{\mathbf{F}^s}(B) > 0$, then there exists another scdf vector $\mathbf{F}^{s*}$ such that $\mu_{\mathbf{F}^{s*}}(B) = 0$ and $\Lambda_n(\mathbf{F}^{s*}) > \Lambda_n(\mathbf{F}^s)$.

For convenience, write $I_i = (L_i, R_i] \times M_i$, $i = 1, ..., n$. Suppose that there are total of n_1+2 distinct values of L_i's and R_i's, including $-\infty$ and ∞, say $a_0 = -\infty < a_1 < a_2 < \cdots < a_{n_1+1} = \infty$. Then the elements in the set $\{(a_l, c) : \ l = 1, ..., n_1; \ c = 1, ..., J\}$ form a grid that partitions the space $[-\infty, \infty] \times [-\infty, \infty]$ with components $B = (a_{l-1}, a_l] \times \{c\}$ or $(a_{l-1}, a_l] \times (c-1, c)$. Moreover, it can be shown that each element B in the grid satisfies that

1. $B \cap A_j = \emptyset$ or B for each A_j, $j = 1, 2, \cdots, m$.
2. $B \cap I_i = \emptyset$ or B for each I_i, $i = 1, 2, \cdots, n$.

Suppose that $\mathbf{F}^s$ assigns some positive weight to at least one non-MI set B_{lc} (that is, $B_{lc} \cap \bigcup_{j=1}^m A_j = \emptyset$), then one of the following cases must happen:

1. $B_{lc} = (a_{l-1}, a_l] \times (c-1, c)$. Denote $\mathcal{B}_1^*$ the collection of these non-MI's;
2. $B_{lc} = (a_{l-1}, a_l] \times \{c\}$ and $B_{lc} \cap I_i = \emptyset$ for all $i = 1, 2, \cdots, n$. Denote $\mathcal{B}_2^*$ the collection of these non-MI's;
3. $B_{lc} = (a_{l-1}, a_l] \times \{c\}$, but $B_{lc} \cap I_i \neq \emptyset$ for some i. Denote $\mathcal{B}_3^*$ the collection of these non-MI's.

Thus $B_k \bigcap (\cup_{j=1}^m A_j) = \emptyset$ for any $B_k \in \mathcal{B}_1^* \bigcup \mathcal{B}_2^* \bigcup \mathcal{B}_3^*$ by the aforementioned assumptions.

Suppose that the weight assigned by $\mathbf{F}^s$ on B_k is $w_k > 0$ and define an new $\mathbf{F}^{s*}$ as follows.

1. In case 1 or 2, $\mathbf{F}^{s*}$ moves $\frac{w_k}{m}$ to each of the m A_j's, that is,

$$\mu_{\mathbf{F}^{s*}}(B_{lc}) = \begin{cases} \mu_{\mathbf{F}^s}(B_{lc}) & \text{if } B_{lc} \neq B_k \text{ and } B_k \in \bigcup_{i=1}^3 \mathcal{B}_i^*, \\ 0 & \text{if } B_{lc} = B_k, \end{cases}$$

and $\mu_{\mathbf{F}^{s*}}(A_j) = \mu_{\mathbf{F}^s}(A_j) + \frac{w_k}{m} > \mu_{\mathbf{F}^s}(A_j)$. Thus for each $i = 1, 2, \cdots, n$, $\mu_{\mathbf{F}^{s*}}(I_i) = \mu_{\mathbf{F}^s}(I_i) + \sum_{j=1}^m \frac{w_k}{m} \mathbf{1}(A_j \subset I_i) > \mu_{\mathbf{F}^s}(I_i)$. Then we obtain

$$\Lambda_n(\mathbf{F}^{s*}) = \prod_{i=1}^n \mu_{\mathbf{F}^{s*}}(I_i) > \prod_{i=1}^n \mu_{\mathbf{F}^s}(I_i) = \Lambda_n(\mathbf{F}^s).$$

2. In case 3 for $B_k \in \mathcal{B}_3^*$, if $B_k \cap I_i \neq \emptyset$ for some i, then without loss of the generality, we can assume that $B_k \subset I_i$ for $i = 1, 2, \cdots, n_k$, but $B_{lc} \cap I_i = \emptyset$ for the rest i's, and $j = 1, 2, \cdots, m_k$, $m_k < m$ such that $A_j \subset \bigcap_{i=1}^{n_k} I_i$. Then F^* moves $\frac{w_k}{m_k}$ to each of the m_k A_j's, that is,

$$\mu_{\mathbf{F}^{s*}}(B_{lc}) = \begin{cases} \mu_{\mathbf{F}^s}(B_{lc}) & \text{if } B_{lc} \in \mathcal{B}_1^* \bigcup \mathcal{B}_2^*, \\ 0 & \text{if } B_{lc} = B_k, \\ \mu_{\mathbf{F}^s}(B_{lc}) & \text{if } B_{lc} \neq B_k \text{ and } B_{lc} \in \mathcal{B}_3^* \end{cases}$$

and

$$\mu_{\mathbf{F}^{s*}}(A_j) = \begin{cases} \mu_{\mathbf{F}^s}(A_j) + \frac{w_k}{m_k} & \text{for } j = 1, 2, \cdots, m_k, \\ \mu_{\mathbf{F}^s}(A_j) & \text{for } j = m_k + 1, \cdots, m. \end{cases}$$

Thus $\exists$ at least an i such that $\mu_{\mathbf{F}^{s*}}(I_i) = \mu_{\mathbf{F}^s}(I_i) + \sum_{j=1}^{m_k} \frac{w_k}{m_k} \mathbf{1}(A_j \subset I_i) > \mu_{\mathbf{F}^s}(I_i)$. Thus we have

$$\Lambda_n(\mathbf{F}^{s*}) = \prod_{i=1}^n \mu_{\mathbf{F}^{s*}}(I_i) > \prod_{i=1}^n \mu_{\mathbf{F}^s}(I_i) = \Lambda_n(\mathbf{F}^s).$$

Either of them tell us that if $\hat{\mathbf{F}}_n^s$ is the NPMLE, then the weights assigned by $\hat{\mathbf{F}}_n^s$ have to be on the MI's only. ▫

§**2. Proof of the claim in Example 2.1:** Based on the model in Example 2.1, we can show that the possible observations are $(-\infty, 1] \times \{1, 2\}$, $(-\infty, 1] \times \{2\}$, $(1, 2] \times \{1\}$, $(1, 2] \times \{2\}$, and $(1, 2] \times \{1, 2\}$ with sizes N_1, N_2, N_3, N_4 and N_5 respectively where $N_1 + N_2 + N_3 + N_4 + N_5 = n$. Thus the MI's are $(-\infty, 1] \times \{2\}$,

$(1,2] \times \{1\}$ and $(1,2] \times \{2\}$ with weights s_1, s_2 and s_3 respectively. Then we can set up the log-likelihood function:

$$\mathcal{L}_n(s_1,s_2,s_3) = \frac{1}{n}(N_1+N_2)\log s_1 + N_3 \log s_2 + N_4 \log(1-s_1-s_2) + N_5 \log(1-s_1)$$

under the constraint $s_1+s_2+s_3=1$. By the differentiation on the weights and setting them equal 0, we can have

$$\frac{N_1+N_2}{s_1} - \frac{N_4}{1-s_1-s_2} - \frac{N_5}{1-s_1} = 0,$$
$$\frac{N_3}{s_2} - \frac{N_4}{1-s_1-s_2} = 0.$$

Solving them yields

$$\hat{s}_1 = \frac{N_1+N_2}{n},\ \ \hat{s}_2 = \frac{(n-N_1-N_2)N_3}{n(N_3+N_4)}.$$

Since $f_{T,C}(1,1)=0$ and $P(T=1)=P(T=2)=1/2$, we have $f_{T,C}(1,2)=\alpha_1=1/2$, $f_{T,C}(2,1)=\alpha_2/2>0$ and $f_{T,C}(2,2)=\alpha_3/2>0$ for some $\alpha_2>0, \alpha_3>0$ with $\alpha_2+\alpha_3=1$. By the SLLN, we have w.p.1,

$$\frac{N_1}{n} \to P(T=1, C=2, (L,R)=(-\infty,1), \Delta=2) = \frac{\alpha_1}{2} = 1/4,$$
$$\frac{N_2}{n} \to P(T=1, C=2, (L,R)=(-\infty,1), \Delta=1) = \frac{\alpha_1}{2} = 1/4,$$
$$\frac{N_3}{n} \to P(T=2, C=1, (L,R)=(1,2), \Delta=1) = \alpha_2/4,$$
$$\frac{N_4}{n} \to P(T=2, C=2, (L,R)=(1,2), \Delta=1) = \alpha_3/4,$$
$$\frac{N_5}{n} \to P(T=2, C=1 \text{ or } 2, (L,R)=(1,2), \Delta=2) = (\alpha_2+\alpha_3)/4 = 1/4.$$

Thus w.p.1, we have

$$\hat{s}_1 \to 1/2 = f_{T,C}(1,2),\ \ \hat{s}_2 \to \frac{\alpha_2/4(1-1/2)}{\alpha_2/4+\alpha_3/4} = \alpha_2/2 = f_{T,C}(2,1),$$

thus the estimators are consistent.

Denote $\mathbf{s}=(s_1,s_2)'$, then by the NPMLE property, we have $\frac{\partial \mathcal{L}_n(\hat{\mathbf{s}})}{\partial \mathbf{s}}=0$ where $\hat{\mathbf{s}}=(\hat{s}_1,\hat{s}_2)$. Then by the first Taylor expansion we have

$$\frac{\partial \mathcal{L}_n(\mathbf{s}^o)}{\partial \mathbf{s}} = \frac{\partial^2 \mathcal{L}_n(\mathbf{s}^o)}{\partial \mathbf{s}^2}(\mathbf{s}^o - \hat{\mathbf{s}}) + o_P(||\mathbf{s}^o - \hat{\mathbf{s}}||).$$

Due to the consistency, for n large enough, $||\hat{\mathbf{s}}(\omega) - \mathbf{s}^o|| < \frac{1}{n}$ for each $\omega \in \Omega$ where Ω denotes the sample space, then we have when $n \to \infty$, $o_P(\sqrt{n}||\mathbf{s}^o - \hat{\mathbf{s}}||) \to 0$.

¿From the SLLN it follows w.p.1

$$\frac{\partial \mathcal{L}_n(\mathbf{s}^o)}{\partial \mathbf{s}} = \begin{pmatrix} \frac{N_1+N_2}{ns_1} - \frac{N_4}{n(1-s_1-s_2)} - \frac{N_5}{n(1-s_1)} \\ \frac{N_3}{ns_2} - \frac{N_4}{n(1-s_1-s_2)} \end{pmatrix}$$
$$\to \begin{pmatrix} 1 - \frac{\alpha_3/4}{(1-1/2-\alpha_2/2)} - \frac{1/4}{1-1/2} \\ \frac{\alpha_2/4}{\alpha_2/2} - \frac{\alpha_3/4}{1-1/2-\alpha_2/2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = E(\frac{\partial \mathcal{L}_n(\mathbf{s}^o)}{\partial \mathbf{s}}),$$

then by CLT, $\sqrt{n}\frac{\partial \mathcal{L}_n(\mathbf{s}^o)}{\partial \mathbf{s}} \to N(\mathbf{0}, \mathcal{I})$ in distribution where

$$\mathcal{I} = -\frac{\partial^2 \mathcal{L}_n(\mathbf{s}^o)}{\partial \mathbf{s}^2} = \begin{pmatrix} 1-\frac{1}{\alpha_3} & -\frac{1}{\alpha_3} \\ -\frac{1}{\alpha_3} & \frac{1}{\alpha_2}-\frac{1}{\alpha_3} \end{pmatrix}$$

is the Fisher Information matrix which can verified that it is positive definite. Thus we can obtain $\sqrt{n}(\mathbf{s}^o - \hat{\mathbf{s}}) \to N(\mathbf{0}, \mathcal{I}^{-1})$ in distribution. ▫

§**3. Derivation of the NPMLE in Example 2.3:** Based on the notations and the model given in Example 2.3, we can derive the log-likelihood function under the constraint $\sum_{i=1}^{8} s_i = 1$

$$\begin{aligned}\mathcal{L}_n(\mathbf{s}) = {} & N_1 \log s_1 + N_2 \log s_2 + N_3 \log s_3 + N_4 \log s_4 + N_5 \log(1 - s_1 - s_2 - s_3 - s_4) \\ & + N_6 \log(s_1 + s_2 + s_5 + s_6) + N_7 \log(1 - s_1 - s_2 - s_5 - s_6) \\ & + N_8 \log(s_1 + s_3 + s_5 + s_7) + N_9 \log(1 - s_3 - s_5 - s_7).\end{aligned}$$

Set the derivative on each s_j equal zero to obtain

$$\begin{aligned}\frac{\partial \mathcal{L}_n}{\partial s_1} = {} & \frac{N_1}{s_1} - \frac{N_5}{1 - s_1 - s_2 - s_3 - s_4} + \frac{N_6}{s_1 + s_2 + s_5 + s_6} \\ & - \frac{N_7}{1 - s_1 - s_2 - s_5 - s_6} + \frac{N_8}{s_1 + s_3 + s_5 + s_7} - \frac{N_9}{1 - s_1 - s_3 - s_5 - s_7} = 0,\end{aligned}$$

$$\frac{\partial \mathcal{L}_n}{\partial s_2} = \frac{N_2}{s_2} - \frac{N_5}{1 - s_1 - s_2 - s_3 - s_4} + \frac{N_6}{s_1 + s_2 + s_5 + s_6} - \frac{N_7}{1 - s_1 - s_2 - s_5 - s_6},$$

$$\frac{\partial \mathcal{L}_n}{\partial s_3} = \frac{N_1}{s_3} - \frac{N_5}{1 - s_1 - s_2 - s_3 - s_4} + \frac{N_8}{s_1 + s_3 + s_5 + s_7} - \frac{N_9}{1 - s_1 - s_3 - s_5 - s_7} = 0,$$

$$\frac{\partial \mathcal{L}_n}{\partial s_4} = \frac{N_1}{s_4} - \frac{N_5}{1 - s_1 - s_2 - s_3 - s_4} = 0,$$

$$\begin{aligned}\frac{\partial \mathcal{L}_n}{\partial s_5} = {} & \frac{N_6}{s_1 + s_2 + s_5 + s_6} - \frac{N_7}{1 - s_1 - s_2 - s_5 - s_6} \\ & + \frac{N_8}{s_1 + s_3 + s_5 + s_7} - \frac{N_9}{1 - s_1 - s_3 - s_5 - s_7} = 0,\end{aligned}$$

$$\frac{\partial \mathcal{L}_n}{\partial s_6} = \frac{N_6}{s_1 + s_2 + s_5 + s_6} - \frac{N_7}{1 - s_1 - s_2 - s_5 - s_6} = 0,$$

$$\frac{\partial \mathcal{L}_n}{\partial s_7} = \frac{N_8}{s_1 + s_3 + s_5 + s_7} - \frac{N_9}{1 - s_1 - s_3 - s_5 - s_7} = 0.$$

Solve them to get the NPMLE given in (2.5). ▫

§**4. Statement in the end of Step 2.2 of the Algorithm in Appendix I:** For $j = 1, 2, \cdots, J$, IMG_j is the set of all MI's induced from the observations in G_j.

Proof: If there is an element from IMG_j, say $(e_{l,i}, e_{r,i}] \times \mathcal{M}_i^*$ is not a MI, then either (1) there is an element from G_j, say $(l_k, r_k] \times \mathcal{M}_k$, such that $(e_{l,i}, e_{r,i}] \bigcap (l_k, r_k] \neq \phi$ and $(e_{l,i}, e_{r,i}] \not\subset (l_k, r_k]$, (2) $\exists (l_k, r_k] \times \mathcal{M}_k$ such that $(e_{l,i}, e_{r,i}] \subset (l_k, r_k]$, but $\mathcal{M}_i^* \not\subset \mathcal{M}_k$,

If case (1) is true then we either have $e_{l,i} < l_k < e_{r,i} < r_k$, or $l_k < e_{l,i} < l_r < e_{r,i}$, or $e_{l,i} < l_k < r_k < e_{r,i}$. But each of them is a contradiction to Step 2.1 of Step 2 in the algorithm where we order all the end points. Case (2) is impossible since Step 2.2 of Step 2 in the algorithm guarantees no such condition exists. ▫.

§**5. Statement in the end of Step 3 of Algorithm in Appendix I.** Let $MG = \bigcup_{j=1}^{J} IMG_j$. If two elements, say $(e_{l,i}, e_{r,i}] \times \mathcal{M}_i^*$ and $(e_{l,j}, e_{r,j}] \times \mathcal{M}_j^*$ for $i \neq j$ from MG are not disjoint, then $\exists\, (e_{l,k}, e_{r,k}] \times \mathcal{M}_k^* \in MG$ for some k such that $[(e_{l,k}, e_{r,k}] \times \mathcal{M}_k^*] \subset [(e_{l,i}, e_{r,i}] \times \mathcal{M}_i^*] \bigcap [(e_{l,j}, e_{r,j}] \times \mathcal{M}_j^*]$.

Proof: If $[(e_{l,i}, e_{r,i}] \times \mathcal{M}_i^*] \subset [(e_{l,j}, e_{r,j}] \times \mathcal{M}_j^*]$, then $[(e_{l,i}, e_{r,i}] \times \mathcal{M}_i^*] \bigcap [(e_{l,j}, e_{r,j}] \times \mathcal{M}_j^*] = (e_{l,i}, e_{r,i}] \times \mathcal{M}_i^* \in MG$ for $k = i$ and by the symmetry, same for the case that $[(e_{l,j}, e_{r,j}] \times \mathcal{M}_j^*] \subset [(e_{l,i}, e_{r,i}] \times \mathcal{M}_i^*]$.

Otherwise, we have $(e_{l,i}, e_{r,i}] \bigcap (e_{l,j}, e_{r,j}] = (e_l, e_r]$ with $e_{l,i} \leq e_l = e_{l,j} < e_r = e_{r,i} \leq e_{r,j}$ or $e_{l,j} \leq e_l = e_{l,i} < e_r = e_{r,j} \leq e_{r,i}$ and $\mathcal{M} = \mathcal{M}_i^* \bigcap \mathcal{M}_j^* \neq \phi$. By Step 2.1. of Step 2 in the algorithm, the ordered end points of $e_{l,i}, e_{r,i}, e_{l,j}$ and $e_{r,j}$ would be in the group G_s for some $s \in \mathcal{M}$. Thus if there are no any other end points appearing between e_l and e_r, then $(e_{l,k}, e_{r,k}] \times \mathcal{M}_k^* = (e_l, e_r] \times \mathcal{M}_k^*$ would be appearing in G_s as a MI for some $\mathcal{M}_k^* \subset \mathcal{M}$; otherwise, $e_l \leq e_{l,k} \leq e_{r,k} \leq e_r$ based on the definition of MI and $\mathcal{M}_k^* \subset \mathcal{M}$ if the end points $e_{l,k}, e_{r,k}$ appears between e_l and e_r, so $[(e_{l,k}, e_{r,k}] \times \mathcal{M}_k^*] \subset [(e_{l,i}, e_{r,i}] \times \mathcal{M}_i^*] \bigcap [(e_{l,j}, e_{r,j}] \times \mathcal{M}_j^*]$. □.

§**6. Proof of Remark 2.1.** Denote $\nu_c(A) = f_{\mathcal{M}|C}(A|c)$ and $\nu_{t,c}(A) = f_{\mathcal{M}|T,C}(A|t,c)$. Notice that if $T > Y_{K,K}$ then $\mathcal{M}^o \neq \mathcal{M}$ (see Mukhopadhyay (2006, p.806[13])), and

$$\text{if } T \leq Y_{K,K},\ C \in A \in P_h \text{ and } \Delta = h \text{ then } \mathcal{M}^o = \mathcal{M} = A. \tag{6.1}$$

We shall first state two lemmas.

Lemma 6.1. *For each* $A \in \mathcal{J}$ $\nu_c(A) = \begin{cases} 0 & \text{if } c \notin A \\ \sum_{h:\ A \in P_h} f_\Delta(h) & \text{if } c \in A. \end{cases}$ *Thus* $\nu_c(A)$ *is constant in* $c \in A$ *.*

Lemma 6.2. $\forall\ A \in \mathcal{J}$ $\nu_{t,c}(A) = \begin{cases} \sum_{h:\ A \in P_h} f_\Delta(h) & \text{if } c \in A \text{ and } f_{T,C}(t,c) > 0 \\ 0 & \text{if } c \notin A \text{ or } f_{T,C}(t,c) = 0. \end{cases}$

Now we shall prove that S1 holds iff S2 holds.

Suppose that S2 holds. Consider two types of $A \in \mathcal{J}$. (1) $\nu_c(A) = 0$ for some $c \in A$ and (2) otherwise. In case (1), by Lemma 6.1 $f_\Delta(h) = 0$ for each partition P_h with $A \in P_h$. It follows from Lemma 6.2 that $\nu_{t,c}(A) = 0$ and thus $\nu_{t,c}(A) = \nu_c(A)$. In case (2), that is, $A \in \mathcal{J}$ with $\nu_c(A) > 0$. If $f_T(t) > 0$, then $f_{T,C}(t,c) > 0$ by S2, and it follows from Lemmas 6.1 and 6.2 that $\nu_{t,c}(A) = \nu_c(A)$. In other words, S1 holds in both cases (1) and (2).

Now suppose that S1 holds. Then $\forall\ A \in \mathcal{J}$ with $\nu_c(A) > 0$, $\nu_c(A) = \nu_{t,c}(A)$ for all (t,c) with $f_T(t) > 0$. Then by Lemmas 6.1 and 6.2 $f_{T,C}(t,c) > 0$, that is, S2 holds.

The equation $\nu_c(A) = \sum_{h:\ A \in P_h} f_\Delta(h)$ if $c \in A$ and $A \in \mathcal{J}$ unique determine $\nu_c(\cdot)$ through $f_\Delta(\cdot)$, say $\nu_c = g(f_\Delta)$. Notice that $||\mathcal{P}|| \geq ||\mathcal{J}||$. By properly selecting a subset of $\mathcal{P}$, one can find an inverse g^{-1} such that $f_\Delta = g^{-1}(\nu_c)$. Under this restriction as well as under S1, the RPM and the CMP model are equivalent. □

§**7 Proof of Lemma 6.1.** For each $A \in \mathcal{J}$ and $c \in A$,

$$\begin{aligned}
\nu_c(A) =& P\{\mathcal{M} = A | C = c\} \\
=& \sum_h P\{\mathcal{M} = A | C = c, \Delta = h\} f_\Delta(h) && \text{(see (6.1))} \\
=& \sum_h P\{C \in A, A \in P_h | C = c, \Delta = h\} f_\Delta(h) \text{ (by the definitions of } \Delta \text{ and } \mathcal{M}) \\
=& \sum_{h:\ A \in P_h} P\{C \in A | C = c, \Delta = h\} f_\Delta(h) && \text{(as } A \text{ and } P_h \text{ are not random)} \\
=& \sum_{h:\ A \in P_h} P\{C \in A | C = c\} f_\Delta(h) && \text{(as } (T,C) \perp (\Delta, Y_{K,K})) \\
=& \sum_{h:\ A \in P_h} f_\Delta(h), \quad c \in A. \ \square
\end{aligned}$$

§**8 Proof of Lemma 6.2.** For each $A \in \mathcal{J}$ and $c \in A$,

$$\begin{aligned}
\nu_{t,c}(A) =& P\{\mathcal{M} = A | C = c, T = t\} \\
=& \sum_h P\{\mathcal{M} = A | C = c, T = t, \Delta = h\} f_\Delta(h) && \text{(see (6.1))} \\
=& \sum_h P\{C \in A, A \in P_h | C = c, T = t, \Delta = h\} f_\Delta(h) \text{ (by the definitions of } \mathcal{M} \text{ and } \Delta) \\
=& \sum_{h:\ c \in A \in P_h} P\{C = c | C = c, T = t, \Delta = h\} f_\Delta(h) && \text{(as } A \text{ and } P_h \text{ are not random)} \\
=& \sum_{h:\ c \in A \in P_h} P\{C = c | C = c, T = t\} f_\Delta(h) && \text{(as } (T,C) \perp (\Delta, Y_{K,K})) \\
=& \sum_{h:\ c \in A \in P_h} \mathbf{1}_{(f_{T,C}(t,c) > 0)} f_\Delta(h)
\end{aligned}$$

$$=\mathbf{1}_{(f_{T,C}(t,c)>0)} \sum_{h:\ c\in A\in P_h} f_\Delta(h) \qquad\qquad \text{(as } (T,C)\perp(\Delta, Y_{K,K})\text{)}$$
$$=\begin{cases} \sum_{h:\ A\in P_h} f_\Delta(h) & \text{if } c\in A \text{ and } f_{T,C}(t,c)>0, \\ 0 & \text{if } c\notin A \text{ or } f_{T,C}(t,c)=0. \end{cases} \square$$

§9 Proof of Theorem 3.1: Define $\mathbf{a}_k = (a_1,\cdots,a_k)$, $g(\mathbf{a}_k, M, k) = P(Y_{k,1}=a_1,\cdots,Y_{k,k}=a_k, \mathcal{M}=M, K=k)$ and $\mathcal{B}_s = \{(\mathbf{a}_k, M, k) : g(\mathbf{a}_k,M,k)>0, a_i\in\mathcal{S}, M\in\mathcal{J}, a_1<\cdots<a_k, k\le d\}$ with $a_0=-\infty$.

Verify that

$$\text{Ł}(F) = E\{\log \mu_{\mathbf{F}^s}((L,R]\times\mathcal{M})\} = \sum_{(\mathbf{a}_k,M,k)\in\mathcal{B}_s} g(\mathbf{a}_k,M,k) h_k(\mathbf{s},\mathbf{a}_k,M)$$

with

$$h_k(\mathbf{s},\mathbf{a}_k,M) = \sum_{i=1}^{k}\sum_{l=1}^{m} s_l^o \mathbf{1}_{(A_l\subset(a_{i-1},a_i]\times M)} \log(\sum_{j=1}^{m} s_j \mathbf{1}_{(A_j\subset(a_{i-1},a_i]\times M)})$$
$$+\sum_{l=1}^{m} s_l^o \mathbf{1}_{(A_l\subset(a_k,\infty)\times\mathcal{C}_r)} \log(\sum_{j=1}^{m} s_j \mathbf{1}_{(A_j\subset(a_k,\infty)\times\mathcal{C}_r)}).$$

It is easy to verify that $h_k(\mathbf{s})$ is maximized by $\mathbf{s}\in D_s$ if and only if

$$\sum_{j=1}^{m} s_j \mathbf{1}_{(A_j\subset(a_{i-1},a_i]\times M)} = \sum_{l=1}^{m} s_l^o \mathbf{1}_{(A_l\subset(a_{i-1},a_i]\times M)}$$

for $i=1,\cdots,k$ and

$$\sum_{j=1}^{m} s_j \mathbf{1}_{(A_j\subset(a_k,\infty)\times\mathcal{C}_r)} = \sum_{l=1}^{m} s_l^o \mathbf{1}_{(A_l\subset(a_k,\infty)\times\mathcal{C}_r)}.$$

Notice that the total number of distinct equations are bigger than or equal to $m-1$, then by A2 we can have that $s_j = s_j^o$ for $j=1,\cdots,m$. Thus $\hat{\mathbf{s}}^o$ maximizes Ł(F) and any other values $\mathbf{s}$ in D_s that maximizes Ł(F) will coincide with $\mathbf{s}^o$ on $\mathcal{A}$.

Note that by the SLLN $\mathcal{L}_n(\mathbf{F}_0^s)\to$ Ł$(\mathbf{F}_0^s)$ almost surely. By the definition of the NPMLE, we have $\mathcal{L}_n(\hat{\mathbf{F}}_n^s)\ge\mathcal{L}_n(\mathbf{F}_0^s)$. Consequently,

$$\lim_{n\to\infty}\inf \mathcal{L}_n(\hat{\mathbf{F}}_n^s) \ge \lim_{n\to\infty}\inf \mathcal{L}_n(\mathbf{F}_0^s) = \text{Ł}(\mathbf{F}_0^s) \text{ almost surely.}$$

Define $\Omega' = \{\lim_{n\to\infty}\inf\mathcal{L}_n(\hat{\mathbf{F}}_n^s) \ge \text{Ł}(\mathbf{F}_0^s)\}$. Fix an $\omega\in\Omega'$, since $(\hat{s}_{1,n}(\omega),...,\hat{s}_{m,n}(\omega))$ is a finite and bounded vector, for each of its subsequence, there is a convergent subsequence. Let $(s_1^*,...,s_m^*)$ be a limiting vector of $(\hat{s}_{1,n}(\omega),...,\hat{s}_{m,n}(\omega))$, that is, for some subsequence $\{k_n\}$ such that $\hat{s}_{j,k_n}(\omega)\to s_j^*$ for $j=1,\cdots,m$. Thus if we can show Ł$(\mathbf{F}^{s*})\ge$ Ł$(\mathbf{F}_0^s)$, then we obtain the desired result where $\mathbf{F}^{s*}$ is a distribution function coinciding with $\mathbf{s}^*=(s_1^*,\cdots,s_{m-1}^*)$.

Let $t_{k_n}(\mathbf{a}_k,M,k)$ denote the value of the random variable

$$\frac{1}{n}\sum_{i=1}^{n}[\sum_{l=1}^{k}\mathbf{1}_{((T_i,C_i)\in(a_{l-1},a_l]\times M, \mathcal{M}_i=M, L_i=a_{l-1}, R_i=a_l)} \log(\sum_{u=1}^{m}\hat{s}_{u,k_n}\mathbf{1}_{(A_u\subset(a_{l-1},a_l]\times M)})$$
$$+\mathbf{1}_{(T_i>a_k, L_i=a_k, R_i=\infty)}\log(\sum_{u=1}^{m}\hat{s}_{u,k_n}\mathbf{1}_{(A_u\subset(a_k,\infty)\times\mathcal{C}_r)})]$$

at the point ω. By the definition of Ω', $\lim_{n\to\infty}\inf\sum_{(\mathbf{a}_k,M,k)\in\mathcal{B}_s} g(\mathbf{a}_k,M,k)t_{k_n}(\mathbf{a}_k,M,k)\ge$ Ł(F_0). It is easy to verify that $t_{k_n}(\mathbf{a}_k,M,k)\to g(\mathbf{a}_k,M,k)h_k(\mathbf{s}^*,\mathbf{a}_k,M)$ for each $(\mathbf{a}_k,M,k)\in\mathcal{B}_s$. Note that $t_{k_n}(\mathbf{a}_k,M,k)\le 0$, from Fatou's Lemma,

$$\lim_{n\to\infty}\sup\sum_{(\mathbf{a}_k,M,k)\in\mathcal{B}_s} t_{k_n}(\mathbf{a}_k,M,k) = -\lim_{n\to\infty}\inf\sum_{(\mathbf{a}_k,M,k)\in\mathcal{B}_s}(-t_{k_n}(\mathbf{a}_k,M,k))$$
$$\le -\sum_{(\mathbf{a}_k,M,k)\in\mathcal{B}_s}\lim_{n\to\infty}\inf(-t_{k_n}(\mathbf{a}_k,M,k)) \to \sum_{(\mathbf{a}_k,M,k)\in\mathcal{B}_s} g(\mathbf{a}_k,M,k)h_k(\mathbf{s}^*,\mathbf{a}_k,M) = \text{Ł}(\mathbf{F}^{s*}).$$

Thus we obtain $Ł(\mathbf{F}_0^s) \leq Ł(\mathbf{F}^{s*})$. As $\mathbf{F}_0^s$ maximizes $Ł(\mathbf{F}^s)$, we conclude that $Ł(\mathbf{F}^{s*}) = Ł(\mathbf{F}_0^s)$ and therefore $s_j^* = s_j^o$ for each $j = 1, \cdots, m$. Since ω is arbitrary and $P(\Omega') = 1$, the consistency is thus established. ▫

§**10 Proof of Theorem 3.2:** Denote $a_0 = -\infty$, $\mathbf{a}_k = (a_1, \cdots, a_k)$,

$g(\mathbf{a}_k, M, k) = P(Y_{k,1} = a_1, \cdots, Y_{k,k} = a_k, \mathcal{M} = M, K = k)$ and

$\mathcal{B}_s = \{(\mathbf{a}_k, M, k) : g(\mathbf{a}_k, M, k) > 0, a_i \in \mathcal{S}, M \in \mathcal{J}, a_1 < \cdots < a_k, k \leq d\}$. Then

$$
\begin{aligned}
\mathcal{L}_n(\mathbf{F}^s) =& \frac{1}{n} \sum_{i=1}^n \log \mu_{\mathbf{F}^s}((L_i, R_i] \times \mathcal{M}_i) \\
=& \frac{1}{n} \sum_{i=1}^n \sum_{(\mathbf{a}_k, M, k) \in \mathcal{B}_s} (\sum_{j=1}^k \mathbf{1}(L_i = a_{j-1}, R_i = a_j, \mathcal{M}_i = M) \log(\mu_{\mathbf{F}^s}((a_{j-1}, a_j] \times M)) \\
& + \mathbf{1}(L_i = a_k, R_i = \infty, \mathcal{M}_i = \mathcal{C}_r) \log(\mu_{\mathbf{F}^s}((a_k, \infty) \times \mathcal{C}_r))).
\end{aligned}
$$

Also we can verify that

$$
\begin{aligned}
& Ł(\mathbf{F}^s) = E\{\log \mu_{\mathbf{F}^s}((L, R] \times \mathcal{M})\} \\
=& \sum_{(\mathbf{a}_k, M, k) \in \mathcal{B}_s} g(\mathbf{a}_k, M, k) [\sum_{i=1}^k \sum_{l=1}^m s_l^o \mathbf{1}_{(A_l \subset (a_{i-1}, a_i] \times M)} \log(\sum_{j=1}^m s_j \mathbf{1}_{(A_j \subset (a_{i-1}, a_i] \times M)}) \\
& + \sum_{l=1}^m s_l^o \mathbf{1}_{(A_l \subset (a_k, \infty) \times \mathcal{C}_r)} \log(\sum_{j=1}^m s_j \mathbf{1}_{(A_j \subset (a_k, \infty) \times \mathcal{C}_r)})] \\
=& \sum_{(\mathbf{a}_k, M, k) \in \mathcal{B}_s} [\sum_{i=1}^k g(\mathbf{a}_k, M, k) \sum_{l=1}^m s_l^o \mathbf{1}_{(A_l \subset (a_{i-1}, a_i] \times M)} \log(\sum_{j=1}^m s_j \mathbf{1}_{(A_j \subset (a_{i-1}, a_i] \times M)})] \\
& + \sum_{i=1}^d [(\sum_{(\mathbf{a}_k, M, k) \in \mathcal{B}_s} \mathbf{1}_{(a_k = t_i)} g(\mathbf{a}_k, M, k)) \sum_{l=1}^m s_l^o \mathbf{1}_{(A_l \subset (t_i, \infty) \times \mathcal{C}_r)} \log(\sum_{j=1}^m s_j \mathbf{1}_{(A_j \subset (t_i, \infty) \times \mathcal{C}_r)})] \\
=& \sum_{i=0}^d \sum_{v=i+1}^d [(\sum_{(\mathbf{a}_k, M, k) \in \mathcal{B}_s, (t_i, t_v) = (a_{u-1}, a_u), 1 \leq u \leq k} g(\mathbf{a}_k, M, k) \sum_{l=1}^m s_l^o \mathbf{1}_{(A_l \subset (t_i, t_v] \times M)}) \\
& \cdot \log(\sum_{j=1}^m s_j \mathbf{1}_{(A_j \subset (t_i, t_v] \times M)})] \\
& + \sum_{i=1}^d [(\sum_{(\mathbf{a}_k, M, k) \in \mathcal{B}_s} \mathbf{1}_{(a_k = t_i)} g(\mathbf{a}_k, M, k)) \sum_{l=1}^m s_l^o \mathbf{1}_{(A_l \subset (t_i, \infty) \times \mathcal{C}_r)} \log(\sum_{j=1}^m s_j \mathbf{1}_{(A_j \subset (t_i, \infty) \times \mathcal{C}_r)})]
\end{aligned}
$$

where $t_0 = -\infty$.

If we order all possible distinct observations from 1 to β, say $\{I_1, I_2, \cdots, I_\beta\}$ with the form $I_h = (t_i, t_v] \times M$ for $t_i, t_v \in \mathcal{S} \bigcup \{-\infty\}$ or $I_h = (t_i, \infty) \times \mathcal{C}_r$ for $t_i \in \mathcal{S}$, and define

$$
p_h = \begin{cases} \sum_{(\mathbf{a}_k, M, k) \in \mathcal{B}_s, (a_{u-1}, a_u) = (t_i, t_v), 1 \leq u \leq k} g(\mathbf{a}_k, M, k) \sum_{l=1}^m s_l^o \mathbf{1}_{(A_l \subset I_h)} & \text{if } I_h = (t_i, t_v] \times M \\ \sum_{(\mathbf{a}_k, M, k) \in \mathcal{B}_s} \mathbf{1}_{(a_k = t_i)} g(\mathbf{a}_k, M, k) \sum_{l=1}^m s_l^o \mathbf{1}_{(A_l \subset I_h)} & \text{if } I_h = (t_i, \infty) \times \mathcal{C}_r, \end{cases}
$$

then we can rewrite

$$
Ł(\mathbf{F}^s) = \sum_{h=1}^\beta p_h \log(\sum_{l=1}^m s_l \delta_{hl})
$$

where $\delta_{hl} = \mathbf{1}_{(A_l \subset I_h)}$. Thus by the assumptions, $p_h > 0$ for $h = 1, 2, \cdots, \beta$.

Letting $\mathcal{I} = -E(\frac{\partial^2 \mathcal{L}_n(\mathbf{F}_0^s)}{\partial \mathbf{s} \partial \mathbf{s}'})$, we can verify

$$
\frac{\partial^2 \mathcal{L}_n(\mathbf{F}_0^s)}{\partial \mathbf{s} \partial \mathbf{s}'} \to E(\frac{\partial^2 \mathcal{L}_n(\mathbf{F}_0^s)}{\partial \mathbf{s} \partial \mathbf{s}'}) = -\mathcal{I} \text{ a.s..}
$$

It follows that

$$\frac{\partial^2 \mathcal{L}_n(\mathbf{F}_0^s)}{\partial \mathbf{s} \partial \mathbf{s}'} = -\mathcal{I} + o_P(1). \tag{T1}$$

Since $\frac{\partial \mathcal{L}_n}{\partial \mathbf{s}}$ is an $(m-1)\times 1$ vector, $\frac{\partial^2 \mathcal{L}_n}{\partial \mathbf{s}\partial \mathbf{s}'}$ is $(m-1)\times(m-1)$ matrix. Verify that

$$\begin{aligned}\mathcal{I} = & nE(\frac{\partial \mathcal{L}_n(\mathbf{F}_0^s)}{\partial \mathbf{s}}\frac{\partial \mathcal{L}_n(\mathbf{F}_0^s)}{\partial \mathbf{s}'}) = -\frac{\partial^2 \text{Ł}(\mathbf{F}_0^s)}{\partial \mathbf{s}\partial \mathbf{s}'} \\ =&(\sum_{h=1}^{\beta} p_h \frac{(\delta_{hu}-\delta_{hm})\cdot(\delta_{hv}-\delta_{hm})}{(\sum_{l=1}^m \delta_{hl} s_l^o)^2})_{(m-1)\times(m-1)} \\ =&\mathbf{U}\mathbf{U}'\end{aligned}$$

where

$$\mathbf{U} = \begin{pmatrix} \frac{(\delta_{11}-\delta_{1m})\sqrt{p_1}}{\sum_{l=1}^m \delta_{1l}s_l^o} & \cdots & \frac{(\delta_{\beta 1}-\delta_{\beta m})\sqrt{p_\beta}}{\sum_{l=1}^m \delta_{\beta l}s_l^o} \\ \vdots & \vdots & \vdots \\ \frac{(\delta_{1(m-1)}-\delta_{1m})\sqrt{p_1}}{\sum_{l=1}^m \delta_{1l}s_l^o} & \cdots & \frac{(\delta_{\beta(m-1)}-\delta_{\beta m})\sqrt{p_\beta}}{\sum_{l=1}^m \delta_{\beta l}s_l^o} \end{pmatrix}$$

Now we show that $\mathcal{I}$ is nonsingular.

Assume $M_1, M_2, \cdots, M_w$ be the distinct MI's w.r.t. MCR part. For each $u \in \{1,\ 2,\ \cdots,\ w\}$, let $A_j = (t_{j-1}, t_j] \times M_u$ for $j = 1, 2, \cdots, d_u$ with $d_u \leq d$ be the MI's, since t_j is the right endpoint of A_j w.r.t. the interval part, by reordering the observations $I_i = (l_i, r_i] \times W_i$ with $M_u \subset W_i$, $i = 1, 2, \cdots, n$, without loss of generalization, we can assume that the right endpoint of these I_i's w.r.t. the interval part is equal to t_j for $j = 1, \cdots, d_u$. Thus $I_i \bigcap A_j = \emptyset$ for $j > i$, $i = 1, 2, \cdots, d_u$, which is an upper triangle matrix. By combining all these upper triangle matrices for $u = 1, 2, \cdots, w$ together, putting the I_i's that repeated appearing in these upper matrices into the same positions in the new matrix, then rearranging the observations I_i, $i = 1, 2, \cdots, n$ such that for $j > i$, $i = 1, 2, \cdots, m-1$, $I_i \bigcap A_j = \emptyset$, the matrix of $\mathbf{U}$ has the upper triangle matrix form

$$\mathbf{U} = \begin{pmatrix} \frac{\sqrt{p_1}}{s_1^0} & \cdot & \cdots & \cdot & \cdots & \frac{(\delta_{\beta 1}-\delta_{\beta m})\sqrt{p_\beta}}{\sum_{l=1}^m \delta_{\beta l}s_l^o} \\ 0 & \frac{\sqrt{p_2}}{s_2^0+\delta_{21}s_1^0} & \cdots & \cdot & \cdots & \frac{(\delta_{\beta 2}-\delta_{\beta m})\sqrt{p_\beta}}{\sum_{l=1}^m \delta_{\beta l}s_l^o} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{\sqrt{p_{m-1}}}{s_{m-1}^0+\sum_{l=1}^{m-1}\delta_{ml}s_l^o} & \cdots & \frac{(\delta_{\beta(m-1)}-\delta_{\beta m})\sqrt{p_\beta}}{\sum_{l=1}^m \delta_{\beta l}s_l^o} \end{pmatrix}$$

Since $s_i^o > 0$ and $p_i > 0$ for $i = 1, 2, \cdots, m-1$, then it follows that matrix $\mathbf{U}$ is of full rank and $\mathbf{U}\mathbf{U}'$ is nonsingular and has the upper triangle matrix form. Then $\mathcal{I} = \mathbf{U}\mathbf{U}'$ is also nonsingular.

By first order Taylor expansion for $\hat{\mathbf{F}}_n^s$ around $\mathbf{F}_0^s$, we have

$$\begin{aligned}\frac{\partial \mathcal{L}(\hat{\mathbf{F}}_n^s)}{\partial \mathbf{s}} = & \frac{\partial \mathcal{L}_n(\mathbf{F}_0^s)}{\partial \mathbf{s}} + \frac{\partial^2 \mathcal{L}_n(\mathbf{F}_0^s)}{\partial \mathbf{s}\partial \mathbf{s}'}\Delta_n + o_p(||\Delta_n||) \\ =&\frac{\partial \mathcal{L}_n(\mathbf{F}_0^s)}{\partial \mathbf{s}} - \mathcal{I}\Delta_n + o_p(||\Delta_n||) \text{ by Eq. (T1)}\end{aligned}$$

where $\Delta_n = (\hat{s}_i - s_i^o)_{(m-1)\times 1} = (\mu_{\hat{F}_n}(A_i) - \mu_{\mathbf{F}_0^s}(A_i))_{(m-1)\times 1}$.

Let $\Omega_n = \{\inf_{i\leq m} \hat{s}_i = 0\}$, then $\frac{\partial \mathcal{L}_n(\hat{\mathbf{F}}_n^s)}{\partial \mathbf{s}} = 0$ by the definition of the NPMLE and $\frac{\partial \mathcal{L}_n(\hat{\mathbf{F}}_n^s)}{\partial \mathbf{s}} = 0$ except on the event Ω_n.

Since each $A_i \in \mathcal{A}$ is of the form either $(t_{i-1}, t_i] \times M_j$ or $(t_i, \infty) \times \mathcal{C}_r$, by the last consistency theorem, $\mu_{\hat{\mathbf{F}}_n^s}((t_{i-1}, t_i] \times M_j) \to \mu_{\mathbf{F}_0^s}((t_{i-1}, t_i] \times M_j)$ a.s. which is equivalent to $\hat{s}_i \to s_i^o$ a.s. And with the given assumptions, we can have

$$P(\Omega_n) \to 0 \text{ a.s. when } n \to \infty.$$

Then by CLT we have $\sqrt{n}\frac{\partial \mathcal{L}_n(\mathbf{F}_0^s)}{\partial \mathbf{s}}$ is asymptotically normal with mean $\mathbf{0}$ and dispersion matrix $\mathcal{I}$. That is

$$\sqrt{n}\frac{\partial \mathcal{L}_n(\mathbf{F}_0^s)}{\partial \mathbf{s}} \to N(\mathbf{0}, \mathcal{I})$$

which implies

$$\sqrt{n}\frac{\partial \mathcal{L}_n(\mathbf{F}_0^s)}{\partial \mathbf{s}} \sim N(\mathbf{0}, \mathcal{I}) + o_P(1).$$

This shows that $\sqrt{n}\Delta_n = \mathcal{I}^{-1}\sqrt{n}\frac{\partial \mathcal{L}_n(\mathbf{F}_0^s)}{\partial \mathbf{s}} + o_P(1) \to N(\mathbf{0}, \mathcal{I}^{-1})$ when $n \to \infty$. Thus we can obtain the desired result. $\square$