Technical Report: Consistency Of The Generalized MLE With Interval-Censored And Masked Competing Risks Data

Jiaping Wang[1] and Qiqing Yu[2]

[1]Department of Biostatistics, UNC-Chapel Hill, NC 27599, USA

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

Current Version: 02/21/2011

Short Title: GMLE with Interval-Censored and Masked Competing Risks Data

Key Words: Mixed case interval censoring, random partition masking, identifiability.

Abstract: We consider nonparametric estimation based on interval-censored competing risks data with masked failure causes. The generalized maximum likelihood estimator of the joint survival function of the failure time and the failure cause is studied under mixed case interval censorship and random partition masking. Strong consistency in the $L_1(\mu)$-topology is established for some finite measure μ derived from the underlying censoring and masking distributions. Under additional regularity assumptions we also establish the strong consistencies in the topologies of weak convergence, point-wise convergence and uniform convergence.

§**1. Introduction.** We consider the consistency of the generalized maximum likelihood estimator (GMLE) of the joint cumulative distribution function (cdf) $F_{T,C}$ of the failure time T and the failure cause C, based on interval-censored and masked competing risks data, called the ICMCR data hereafter.

An excellent review of the competing risks analysis in engineering and biomedical applications can be found in Crowder (2001). Most studies in the literature are about the statistical inferences for right-censored data with or without masked failure cause. Hudgens *et al.* (2001) study the nonparametric inference for interval-censored data with exact competing risks. Basu *et al.* (2003) discuss the Bayesian inference for ICMCR data under a parametric set-up on $F_{T,C}$. Groeneboom *et al.* (2008) provide the proof of the consistency of the GMLE of $F_{T,C}$ based on case 1 interval-censored data without masked failure cause. Wang *et al.* (2011) present an ICMCR cancer research data and some simulation results on the asymptotic properties of the GMLE with ICMCR data. This paper establish the consistency of the GMLE of $F_{T,C}$ with ICMCR data.

An ICMCR observation contains the information on the failure time and the associated failure cause of a J-component series system which stops functioning as soon as one of its J constituent components fails. Assume that the systems under study are non-repairable, the observation on the failure time T and the failure cause C of such systems can be described as follows. Let the random variable X_j denote the lifetime of the j^{th} component, $j = 1, 2, \cdots, J$. Let $T = \min\{X_1, X_2, \cdots, X_J\}$. 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 each system failure time T, say $X_C = T$. It is often in medical research or industrial experiments that T is interval censored and it is only known that $T \in (L, R]$, where $-\infty \le L < R \le \infty$. There are at least four different models for the interval censoring proposed by Ayer *et al.* (1995), Groeneboom and Wellner (1992), Wellner

(1995) and Schick and Yu (2000). We shall make use of the mixed case model proposed by Schick and Yu (2000), as it is more realistic.

To determine the cause C of the failure, one may not be able to pinpoint the exact one. In examining a failed system, one may first check parts one by one in detecting the failure cause and may stop at some point due to cost saving if it makes no sense economically to continue. Thus the failure cause C is masked by $\mathcal{M}$, a subset of integers which is called the minimum random set (MRS) (see Guess *et al.* (1991)). There are two possible models for the masking. One is called the conditional masking probability (CMP) model based on the symmetry assumption, which "is done purely for mathematical convenience without practical justification" (see Mukhopadhyay and Basu (2007, p.331[15])). The other one is introduced in Wang *et al.* (2011) and is called the random partition masking (RPM) model which does not make use of the symmetry assumption. The practical justification for the RPM model will be introduced in section 2. The RPM model has the following advantages over the CMP model (see Wang *et al.* (2011)):

(1) If $J = 2$, the CMP model is a special case of the RPM model;

(2) The CMP model relies on the symmetry assumptions but the RPM model doesn't.

One of the symmetry assumptions is stated as follows

S1 $f_{\mathcal{M}|T,C}(A|t,c)$ is constant in t (see Flehinger *et al.* (1996))

which is often not satisfied (see Example 6.1 in §6).

Thus we shall make use of the RPM model for masking. The mass assigned by the GMLE of a cdf with univariate interval-censored data is unique, but it is no longer always true for ICMCR data and it may not be consistent.

Example 1.1. Suppose that T is subject to case 1 interval censoring with censoring variable $Y = 1$ w.p.1., the possible observations on C are of the forms $W_j = \{1,2\}$, $\{3,4\}$, $\{1,3\}$ or $\{2,4\}$. Let n_j be the number of the observations of form W_j, $j = 1,2,3,4$. Denote

$p_c = P(T \leq 1, C = c)$, $c \in \{1, 2, 3, 4\}$. Then the GMLEs of p_c are not unique: $\hat{p}_2 = r_1 - \hat{p}_1$, $\hat{p}_3 = r_2 - \hat{p}_1$ and $\hat{p}_4 = 1 - \hat{p}_1 - \hat{p}_2 - \hat{p}_3$, where $\hat{p}_1$ is arbitrary in $[\max\{0, r_1 + r_2 - 1\}, \min\{r_1, r_2\}]$, $r_1 = n_1/(n_1 + n_2)$ and $r_2 = n_3/(n_3 + n_4)$ (see the derivation in Example 6.1). There are infinity many GMLE solutions for p_i's. Moreover, it is easy to show that the limit of the interval $[\max\{0, r_1 + r_2 - 1\}, \min\{r_1, r_2\}]$ is not a singleton, thus $\hat{p}_1$ is not consistent.

In the literature such non-uniqueness of the GMLE for bivariate interval-censored data is discovered out by Gentleman and Vandal (2002) and is called "mixture non-uniqueness". The bivariate cdf is not identifiable in the region where such non-uniqueness would happen. Thus identifiability conditions are needed for the GMLE being consistent in bivariate censoring.

Simulation results in Wang *et al.* (2011) suggest that under certain conditions, the GMLE of $F_{T,C}$ with ICMCR data is consistent, at least in a subset of the range of (T, C). Our main task in this paper is to study the identifiability conditions and to give a rigorous proof of the consistency of the GMLE. We use the framework in Schick and Yu (2000).

The rest of the paper is organized as follows. In section 2, we introduce the model for ICMCR data and derive the likelihood function in the nonparametric context. In section 3, we give the main result, the $L_1(\mu)$-consistency of the GMLE for some finite measure μ. In section 4, we provide the consistency results in the topologies of weak convergence, pointwise convergence and uniform convergence. In section 5, some proofs of the statements in the previous sections are provided. Section 6 is the Appendix.

§**2. The ICMCR Model and the GMLE.** We shall propose an ICMCR model which consists of two parts: the mixed case interval censorship model for the failure time T and the RPM model for the failure cause C.

First, we introduce the mixed case interval censorship model. Let K be a positive random integer and $\mathbf{Y} = \{Y_{k,j} : j = 1, 2, \cdots, k,\ k = 1, 2, \cdots\}$ be an array of random variables such that $Y_{k,1} < Y_{k,2} < \cdots < Y_{k,k}$. Given the event $\{K = k\}$, let (L, R) denote the vector of 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 contain T where $Y_{k,0} = -\infty$ and $Y_{k,k+1} = \infty$.

Second, we introduce the RPM model. Let $\mathcal{C}_r = \{1, 2, \cdots, J\}$ be the set of the labels associated with the failure causes (or competing risks) which is also the range of C and $\mathcal{J}$ be the collection of all the non-empty subsets of $\mathcal{C}_r$. One may not observe the exact C, but rather a random subset $\mathcal{M}$ of $\mathcal{C}_r$ which contains C. If the failure time T is right-censored, then the failure cause is completely unknown and we set $\mathcal{M} = \mathcal{C}_r$. If T is not right-censored, $\mathcal{M}$ is obtained as follows. Let $\mathcal{P}$ denote the set of all regular partitions. A regular partition is a collection $P_h = \{P_{h1}, \cdots, P_{hk_h}\}$ of pairwise disjoint non-empty subsets of $\mathcal{C}_r$ whose union is $\mathcal{C}_r$. By the definition, given a regular partition and given a $c \in \mathcal{C}_r$, there is a unique member of the partition which contains c. The set $\mathcal{P}$ has finitely many elements, say $n_{\mathcal{P}}$, and its elements can be ordered as $P_1, \cdots, P_{n_{\mathcal{P}}}$. Let Δ denote a random variable taking values in $\{1, \cdots, n_{\mathcal{P}}\}$. Given the events $\{\Delta = h\}$ and $\{T \text{ is not right-censored}\}$, $\mathcal{M}$ equals the (unique) random member of the partition P_h which contains C.

For example, $P_1 = \{\{1\}, \{2\}, \cdots, \{J\}\}$, $P_2 = \{\{1\}, \{2\}, \{3, 4, \cdots, J\})$ and $P_3 = \{\mathcal{C}_r\}$ are three such partitions. The partition P_2 can be interpreted as follows: in the process of determining the cause of failure in a system, exactly two steps will be taken. The first step can determine whether the failure is due to cause 1, and the second step can determine whether the failure is due to cause 2. If the failure is not due to these two causes, then the cause will be in $\{3, 4, \cdots, J\}$, no further investigation will be taken for cost saving. Then

$$\mathcal{M} = \begin{cases} \{1\} & \text{if } C = 1 \\ \{2\} & \text{if } C = 2 \\ \{3, 4, \cdots, J\} & \text{otherwise} \end{cases} \tag{2.1}$$

Thus, once the partition scheme $(P_{h1}, P_{h2}, \cdots, P_{hk_h})$ is chosen after the failure occurs, then $\mathcal{M}$ can be uniquely determined. However, it is worth mentioning that the aforementioned inspection scheme is only one of the six examination procedures corresponding to P_2. 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.

The second step can be either of the 2 remaining inspections. Thus P_2 corresponds to total of 6 (= 3!) examination schemes. All the 6 of them result in $\mathcal{M}$ in (2.1).

The random vector $(T, C, K, \Delta, \mathbf{Y})$ may not be observed, instead, (L, R) is the observable random vector on T and $\mathcal{M}$ is the observable random variable on C, where

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

The aforementioned scheme for masking is based on the partition on $\mathcal{C}_r$ and thus is called the random partition model for masking. Thus the ICMCR data can be modeled as follows. Given $K = k$ and $\Delta = h$, $(L, R, \mathcal{M})$ is given by

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

A1 The random vectors (T, C) and $(K, \Delta, \mathbf{Y})$ are independent.

Let $(L_i, R_i, \mathcal{M}_i)$, $i = 1$, ..., n be i.i.d. copies of $(L, R, \mathcal{M})$. The generalized likelihood function in the nonparametric context is

$$\Lambda_n(F) = \Pi_{i=1}^n \mu_F((L_i, R_i] \times \mathcal{M}_i), \quad \text{where } F \text{ is a bivariate cdf.}$$

The GMLE of F maximizes $\Lambda_n(F)$ over all bivariate cdf's F.

Let $F_0 = F_{T,C}$ for convenience and $\mathbf{F}_0^s(t) = (F_{10}^s(t), \cdots, F_{J0}^s(t))'$, where $F_{j0}^s(t) = P(T \leq t, C = j)$ and A' is the transpose of the matrix A. Denote by $\mathcal{F}$ the collection of functions F on $[-\infty, \infty]^2$ satisfying $F(-\infty, -\infty) = 0$, $F(\infty, \infty) = 1$,

$$F(x, y) = \sum_{c \leq y, c \in \mathcal{C}_r} F_c^s(x) \quad \forall\ (x, y) \in [-\infty, \infty]^2 \tag{2.2}$$

and $F^s_c(x)$ is nondecreasing in x $\forall$ $c \in \mathcal{C}_r$. One can extend the domain of $F_{T,C}$ and its GMLE to $[-\infty, \infty]^2$ in an obvious way. Notice that the space of all cdf's is not complete but $\mathcal{F}$ is. For each $F \in \mathcal{F}$, let $\mathbf{F}^s(t) = (F^s_1(t), F^s_2(t), \cdots, F^s_J(t))'$ where F^s_1, ... , F^s_J are given in (2.2). Let $\mathcal{F}^s$ be the collection of all such $\mathbf{F}^s$. Obviously, $\mathbf{F}^s_0$ is a member of $\mathcal{F}^s$ and F_0 is a member of $\mathcal{F}$. For convenience, we define $G_{\mathbf{F}^s}(t, m) \overset{def}{=} \phi(m) \cdot \mathbf{F}^s(t) = \sum_{j \in m} F^s_j(t)$, where

$$\phi(A) \overset{def}{=} (\mathbf{1}_{(1 \in A)}, \mathbf{1}_{(2 \in A)}, \cdots, \mathbf{1}_{(J \in A)}) \quad A \in \mathcal{J}. \tag{2.3}$$

$\forall$ $w = (L, R] \times M$, define $\mu_F(w) = G_{\mathbf{F}^s}(R, M) - G_{\mathbf{F}^s}(L, M)$, which is measurable. Thus $0 \leq G_{\mathbf{F}^s} \leq 1$ and $0 \leq \mu_F \leq 1$. Then the normalized log likelihood function is

$$\mathcal{L}_n(F) = \frac{1}{n} \sum_{i=1}^{n} \log(G_{\mathbf{F}^s}(R_i, \mathcal{M}_i) - G_{\mathbf{F}^s}(L_i, \mathcal{M}_i)).$$

Note that $\mathcal{L}_n(F)$ depends on F only through the values of F at the points (L_i, j) or (R_i, j), $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, J$. Thus the maximizer of $\mathcal{L}_n(F)$ over the set $\mathcal{F}$ is not unique. However, there exists a unique maximizer $\hat{F}_n$ over the set $\mathcal{F}$ which is right continuous and piecewise constant with possible discontinuities only at the observed values of (R_i, j)'s or (L_i, j)'s. We call $\hat{F}_n$ the GMLE of F_0. Denote the corresponding GMLE of $\mathbf{F}^s_0$ by $\hat{\mathbf{F}}^s_n$, where $\hat{\mathbf{F}}^s_n(t) = (\hat{F}^s_{1n}(t), \cdots, \hat{F}^s_{Jn}(t))'$. In general, the GMLE does not have an explicit solution. Wang *et al.* (2011) propose to find a numerical solution of the GMLE through the self-consistency algorithm.

§**3. Main Results.** We assume $E(K) < \infty$ which is a very mild assumption, as explained in Schick and Yu (2000). Define a measure μ on the product σ-field $\mathcal{B} = \mathcal{A} \times \mathcal{B}_{\mathcal{J}}$ by

$$\mu(B) = \sum_{k=1}^{\infty} P(K = k) \sum_{i=1}^{k} P((Y_{k,i}, \mathcal{M}) \in B | K = k), \quad B \in \mathcal{B}, \tag{3.1}$$

where $\mathcal{A}$ is the Borel σ-field on $\Re$ and $\mathcal{B}_{\mathcal{J}}$ is the power set of $\mathcal{J}$. Define another measure ν on $\mathcal{A}$ by

$$\nu(A) = \sum_{k=1}^{\infty} P(K = k) \sum_{i=1}^{k} P(Y_{k,i} \in A | K = k), \quad A \in \mathcal{A}. \tag{3.2}$$

Notice that

$$\mu(B) = \sum_{k=1}^{\infty} P(K=k) \sum_{i=1}^{k} P(Y_{k,i} \in A | K=k) P(\mathcal{M} \in W | K=k, Y_{k,i} \in A), \tag{3.3}$$

for $B = A \times W$ with $W \in \mathcal{B}_{\mathcal{J}}$. By the assumption $E(K) < \infty$, we have

$$\mu(B) = \sum_{k=1}^{\infty} P(K=k) \sum_{i=1}^{k} P((Y_{k,i}, \mathcal{M}) \in B | K=k) \leq \textstyle\sum_{k=1}^{\infty} P(K=k)k = E(K) < \infty;$$
$$\nu(A) = \sum_{k=1}^{\infty} P(K=k) \sum_{i=1}^{k} P(Y_{k,i} \in A | K=k) \leq \sum_{k=1}^{\infty} k \cdot P(K=k) = E(K) < \infty.$$

The main $L_1(\mu)$-consistency result is stated as follows with its proof relegated to §5.

Theorem 3.1. *Suppose that A1 holds and $E(K) < \infty$. Then $\int |G_{\hat{\mathbf{F}}_n^s} - G_{\mathbf{F}_0^s}| d\mu \to 0$ a.s..*

In general, we do not have $\hat{F}_n(t,c) \to F_0(t,c) \ \forall \ t < \tau$ (see Example 3.1), where τ does not depend on c as in the univariate case.

Example 3.1. Let the range of (T,C) be $\{0,1,2\} \times \{1,2,3\}$ with partitions on $\mathcal{C}_r$: $P_1 = \{\{1\},\{2\},\{3\}\}$ and $P_2 = \{\{1,3\},\{2\}\}$. Assume the case 1 model with the censoring variable $U \sim Bin(1,p)$ for some $p > 0$ and define the conditional density function $f_{\Delta|U}(1|0) = 1$ and $f_{\Delta|U}(2|1) = 1$. It is obvious that $\hat{F}_{2n}^s(1)$ is consistent but $\hat{F}_{1n}^s(1)$ and $\hat{F}_{3n}^s(1)$ are not.

For each $t \in \Re$, we call t a support point of ν if $\nu(t-\epsilon, t+\epsilon) > 0$ for every $\epsilon > 0$. Denote by $\mathcal{S}$ the collection of all the support points of ν. Denote
$\mathcal{W} = \{t \in \mathcal{S} : \forall\, (k,i),\ \exists$ an open set $O_{k,i}$ $(\neq \emptyset)$ such that $t \in O_{k,i}$ and $P(Y_{k,i} \in O_{k,i}) = 0\}$.

Lemma 3.1. $\nu(\mathcal{W}) = 0$.

Proof: Since $\nu(\mathcal{W}) = \sum_{k=1}^{\infty} P(K=k) \sum_{i=1}^{k} P(Y_{k,i} \in \mathcal{W} | K=k) > 0$. The lemma follows from the definition of $\mathcal{W}$. □

Though Theorem 3.1 is applicable to Example 3.1, it does not tell whether $\hat{F}_n$ is consistent at a given point (t,c). Thus, if we want to obtain the consistency for the GMLE of each sub-distribution function, we need additional assumptions. First we give two definitions.

Definition 3.1. We say that f_C is identifiable at c if $\exists\ \{W_i\}_{i=1}^k$ such that $\phi(\{c\}) = \sum_{i=1}^k g_i\phi(W_i)$ for some constant g_i's, and $W_i \in P_{h_i}$ with $f_\Delta(h_i) > 0\ \forall\ i$.

Definition 3.2. We say that F is identifiable at $(r, c) \in \mathcal{S} \times \mathcal{C}_r$ if either $P(T > r) = 1$ or $\exists\ \{(l_i, r_i] \times W_i\}_{i=1}^w$ with $l_i,\ r_i \in \mathcal{S} \cup \{-\infty\}$ such that (1) $\max_i l_i < r = \min_i r_i$, (2) $P(\mathcal{M} = W_i|(L, R) = (l_i, r_i)) > 0$, and (3) $\phi(\{c\}) = \sum_{i=1}^w g_i\phi(W_i)$ for some constant g_i's.

A sufficient condition that f_C is identifiable at each c is as follows.

A2 The matrix $(\phi(M_1)', \cdots, \phi(M_w)')$ (see (2.3)) is of rank J, where $M_1, M_2, \cdots, M_w$ are all the distinct values of $\mathcal{M}$ such that $P(\mathcal{M} = M_\alpha) > 0$, $\alpha = 1, 2, \cdots, w$.

The following example may be helpful in understanding A2, Definitions 3.1, 3.2 and some regularity conditions.

Example 3.2 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 $f_{\Delta|U}(1|1) = 1$, $f_{\Delta|U}(2|2) = f_{\Delta|U}(3|2) = 1/2$. 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 GMLE of $(s_1, ..., s_8)$ is

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

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

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

In this example, A2 is satisfied and $_C$ is identifiable but F is identifiable only at those points (t, c) at which the W_j's related to t can satisfy A2. Thus the consistency results can be hold only for some special region.

Lemma 3.2. *Suppose that A2 holds. For each $c \in \mathcal{C}_r$, $\exists\, u_c \in \mathcal{S}$ such that F is identifiable at (u_c, c). Moreover, if F is identifiable at (u_c, c), then F is identifiable at (t, c) for almost all t (with respect to the measure ν) in the set $\{s: \ s \le u_c, s \in \mathcal{S}\}$.*

For proof, see §6. For $c \in \mathcal{C}_r$, define $\tau_c = \sup\{t \in S: \text{ F is identifiable at } (t, c)\}$.

Remark 3.1. It is shown in §6 that the collection $\{(l_i, r_i] \times W_i : i = 1, 2, \cdots, w_\tau\}$ satisfying the conditions in Definition 3.2 does not depend on r for $r < \tau_c$. By Lemma 3.2, F is identifiable at (t, c) for almost all $t \in \mathcal{S} \cap \Gamma_c$, where $\Gamma_c = \begin{cases} (-\infty, \tau_c] & \text{if } F \text{ is identifiable at } (\tau_c, c) \\ (-\infty, \tau_c) & \text{otherwise.} \end{cases}$

The proof of the following (strong) $L_1(\nu)$-consistency for the GMLE is given in §5.

Theorem 3.2. *Suppose that $E(K) < \infty$ and assumption A2 holds. Then $\forall\ c \in \mathcal{C}_r$, $\int_{\Gamma_c} |\hat{F}^s_{cn}(t) - F^s_{c0}(t)| d\nu(t) \to 0$ a.s..*

Based on this theorem and define $\Gamma = \bigcap_{c \in \mathcal{C}_r} \Gamma_c$, we can have the following result.

Corollary 3.1. *If $E(K) < \infty$, A2 holds and $c \in \mathcal{C}_r$, then $\int_\Gamma |\hat{F}_n(t, c) - F_0(t, c)| d\nu(t) \stackrel{a.s.}{\to} 0$.*

Proof: By definition, we have

$$\int_\Gamma |\hat{F}_n(t, c) - F_0(t, c)| d\nu(t) = \int_\Gamma |\sum_{j \le c} (\hat{F}_{jn}(t) - F_{j0}(t))| d\nu(t)$$

$$\le \int_\Gamma \sum_{j \le c} |\hat{F}_{jn}(t) - F_{j0}(t)| d\nu(t) \le \sum_{j \le c} \int_{\Gamma_j} |\hat{F}_{jn}(t) - F_{j0}(t)| d\nu(t) \to 0 \text{ a.s. } \square$$

Moreover, if $P((T_I(\{a\}), \mathcal{M} = M) > 0$ and denote $w = \{a\} \times M$, by Theorem 3.1,

$$\mu(w) |G_{\hat{\mathbf{F}}^s_n}(a, M) - G_{\mathbf{F}^s_0}(a, M)| \le \int |G_{\hat{\mathbf{F}}^s_n} - G_{\mathbf{F}^s_0}| d\mu \to 0.$$

So we have the following corollary.

Corollary 3.2. *If $E(K) < \infty$, $w = \{a\} \times M$ and $\mu(w) > 0$, then $G_{\hat{\mathbf{F}}^s_n}(a, M) \stackrel{a.s.}{\to} G_{\mathbf{F}^s_0}(a, M)$.*

Similarly, if $P(T_I(\{a\})) > 0$ for some $a \in \Gamma$, by Corollary 3.1, $\forall\ c \in \mathcal{C}_r$,

$$\nu(\{a\}) |\hat{F}_n(a, c) - F_0(a, c)| \le \int_\Gamma |\hat{F}_n(t, c) - F_0(t, c)| d\nu(t) \to 0.$$

So we have following corollary.

Corollary 3.3. *If $E(K) < \infty$, A2 holds, $a \in \Gamma$, $\nu(\{a\}) > 0$ and $c \in \mathcal{C}_r$, then $\hat{F}_n(a, c) \overset{a.s.}{\rightarrow} F_0(a, c)$.*

Fixing $c \in \mathcal{C}_r$, if $P(T_I(\{a\})) > 0$ for some $a \in \Gamma_c$, by Theorem 3.2,

$$\nu(\{a\})|\hat{F}_{cn}(a) - F_{c0}(a)| \leq \int_\Gamma |\hat{F}_{cn}(t) - F_{c0}(t)| d\nu(t) \rightarrow 0.$$

So we have following corollaries.

Corollary 3.4. *If $E(K) < \infty$, A2 holds, $a \in \Gamma_c$ and $\nu(\{a\}) > 0$, then $\hat{F}_{cn}(a) \overset{a.s.}{\rightarrow} F_{c0}(a)$.*

Corollary 3.5. *Suppose that $\mathcal{S} = \{t_1, t_2, \cdots, t_d\}$ with $d < \infty$, $E(K) < \infty$ and A2 holds. Then the GMLE $\hat{F}$ has the following properties.*

1. *$\hat{F}_T(t_j)$ is consistent for $t_j \in \mathcal{S}$;*
2. *$\hat{f}_{C|T}(c|t_j)$ is consistent for $t_j \in \Gamma_c \cap \mathcal{S}$;*
3. *$\hat{f}(t_j, c) = \hat{f}_{C|T}(c|t_j)\hat{f}_T(t_j)$ is consistent for $t_j \in \Gamma_c \cap \mathcal{S}$.*

Proof: Since the inspection times are finite discrete, there are observations O_i with the form $(-\infty, t_j] \times M_i$ for $M_i \in P_h$. Since $\mu(O_i) > 0$, by Corollary 3.2, we can obtain

$$\begin{aligned} & G_{\hat{\mathbf{F}}_n^s}(t_j, M_i) \overset{a.s.}{\rightarrow} G_{\mathbf{F}_0^s}(t_j, M_i) \\ \Rightarrow & \sum_{M_i \in P_h} G_{\hat{\mathbf{F}}_n^s}(t_j, M_i) \overset{a.s.}{\rightarrow} \sum_{M_i \in P_h} G_{\mathbf{F}_0^s}(t_j, M_i) \\ \Rightarrow & \hat{F}_T(t_j) \overset{a.s.}{\rightarrow} F_T(t_j), \text{ which is the first statement.} \end{aligned}$$

For each $t_j \in \mathcal{S}$, $\nu(t_j) > 0$. Since $t_j \in \Gamma_c$, by Corollary 3.4, $\hat{F}^s_{cn}(t_j) \overset{a.s.}{\rightarrow} F^s_{c0}(t_j)$. Then we obtain $\hat{F}^s_{cn}(t_j) - \hat{F}^s_{cn}(t_{j-1}) \overset{a.s.}{\rightarrow} F^s_{c0}(t_j) - F^s_{c0}(t_{j-1})$, which yields $\hat{f}^s_{cn}(t_j) \overset{a.s.}{\rightarrow} f^s_{c0}(t_j)$. Thus $\hat{f}_{C|T}(c|t_j) = \frac{\hat{f}^s_{cn}(t_j)}{\hat{F}_T(t_j) - \hat{F}_T(t_{j-1})} \overset{a.s.}{\rightarrow} \frac{f^s_{c0}(t_j)}{F_T(t_j) - F_T(t_{j-1})} = f_{C|T}(c|t_j)$ by the continuous mapping, where $t_0 = -\infty$. Thus the second statement holds. Similarly, by the continuous mapping, the third statement follows from statements 1 and 2. □

If we define a measure V based on the distribution of the endpoints L and R with the MRS $\mathcal{M}$, then we can interpret the consistency easier than μ by the following definition:

$$V(B) = P(L \times \mathcal{M} \in B) + P(R \times \mathcal{M} \in B) \text{ for } B \in \mathcal{B}$$

Then in view of the set inclusion

$$\{L \times \mathcal{M} \in B\} \cup \{R \times \mathcal{M} \in B\} \subset \bigcup_{k=1}^{\infty} \bigcup_{i=1}^{k} \{K = k, Y_{k,i} \times \mathcal{M} \in B\}$$

we have $V(B) \leq 2\mu(B)$. Thus we have following corollary:

Corollary 3.6. *Suppose $E(K) < \infty$. Then*

(1) $\int |G_{\hat{\mathbf{F}}^s_n} - G_{\mathbf{F}^s_0}| dV \stackrel{a.s.}{\to} 0$;

(2) $G_{\hat{\mathbf{F}}^s_n}(a, M) \stackrel{a.s.}{\to} G_{\mathbf{F}^s_0}(a, M)$ $\forall\ w = \{a\} \times M$ *with* $V(w) > 0$.

§**4. Consistency Results in other topologies.** Suppose that $E(K) < \infty$ and $A2$ holds. Hereafter, we shall fix a $c \in \mathcal{C}_r$ and denote $\Omega_{\nu,c} = \{\lim_{n\to\infty} \int_{\Gamma_c} |\hat{F}^s_{cn}(t) - F^s_{c0}(t)| d\nu(t) = 0\}$ and $p_c = P(C = c)$. Then $P(\Omega_{\nu,c}) = 1$ by Theorem 3.2. Let x be a real number, we say that x is regular if $\nu((x-\epsilon, x]) > 0$ and $\nu([x, x+\epsilon)) > 0\ \forall\ \epsilon > 0$; x is strongly regular if $\nu((x-\epsilon, x)) > 0$ and $\nu((x, x+\epsilon)) > 0\ \forall\ \epsilon > 0$; and x is a point of increase of $F^s_{c0}(\cdot)$ if $F^s_{c0}(x+\epsilon) - F^s_{c0}(x-\epsilon) > 0$ $\forall\ \epsilon > 0$. Given $\omega \in \Omega_{\nu,c}$, define $D_c = \{t \in \Gamma_c : \lim_{n\to\infty} \hat{F}^s_{cn}(t;\omega) \neq F^s_{c0}(t)\}$. By Theorem 3.2,

$$\int_{\Gamma_c} |\hat{F}^s_{cn}(t;\omega) - F^s_{c0}(t)| d\nu = \int_{D_c} |\hat{F}^s_{cn}(t;\omega) - F^s_{c0}(t)| d\nu + \int_{\Gamma_c \setminus D_c} |\hat{F}^s_{cn}(t;\omega) - F^s_{c0}(t)| d\nu \to 0.$$

$$\Rightarrow \int_{D_c} |\hat{F}^s_{cn}(t;\omega) - F^s_{c0}(t)| d\nu \to 0. \quad \Rightarrow \nu(D_c) = 0.$$

The first proposition gives the strong consistency for the regular continuity points.

Proposition 4.1. *Suppose that $E(K) < \infty$, $A2$ holds and x is a regular continuity point of $F^s_{c0}(\cdot)$ in Γ_c. Then $\hat{F}^s_{cn}(x) \to F^s_{c0}(x)$ a.s..*

Proof: It suffices to show that $\forall\ \omega \in \Omega_{\nu,c}$, D_c does not contain any regular continuity points of $F^s_{c0}(\cdot)$. Otherwise, if x_0 is a regular continuity point of $F^s_{c0}(\cdot)$ and $x_0 \in D_c$, then $\lim_{n\to\infty} \hat{F}^s_{cn}(x_0;\omega) \neq F^s_{c0}(x_0)$. It follows that there exist $\delta > 0$ and $N_1 > 0$ such that $|\hat{F}^s_{cn}(x_0;\omega) - F^s_{c0}(x_0)| > \delta$ if $n > N_1$, that is, $\hat{F}^s_{cn}(x_0;\omega) - F^s_{c0}(x_0) > \delta$ or $\hat{F}^s_{cn}(x_0;\omega) - F^s_{c0}(x_0) < -\delta$. By the continuity of $F^s_{c0}(\cdot)$ at x_0, given $\epsilon > 0$, $\exists\ N_2 > 0$ such that $|F^s_{c0}(x) - F^s_{c0}(x_0)| < \frac{\delta}{n}\ \forall\ n > N_2$ and $|x - x_0| < \epsilon$. If $|\hat{F}^s_{cn}(x_0;\omega) - F^s_{c0}(x_0)| < \delta$ for $x \in [x_0, x_0 + \epsilon)$, then

$$
\begin{aligned}
&\hat{F}^s_{cn}(x_0;\omega) - \hat{F}^s_{cn}(x;\omega) + \hat{F}^s_{cn}(x;\omega) - F^s_{c0}(x_0) > \delta \\
\Rightarrow &\hat{F}^s_{cn}(x;\omega) - F^s_{c0}(x_0) > \delta + \hat{F}^s_{cn}(x;\omega) - \hat{F}^s_{cn}(x_0;\omega) \\
\Rightarrow &\hat{F}^s_{cn}(x;\omega) - F^s_{c0}(x_0) > \delta \qquad\qquad \text{(by nondecreasing monotonicity of } \hat{F}^s_{cn}(\cdot;\omega)) \\
\Rightarrow &\hat{F}^s_{cn}(x;\omega) - F^s_{c0}(x) + F^s_{c0}(x) - F^s_{c0}(x_0) > \delta \\
\Rightarrow &\hat{F}^s_{cn}(x;\omega) - F^s_{c0}(x) > \delta + F^s_{c0}(x_0) - F^s_{c0}(x) > \delta - \frac{\delta}{n} = \frac{(n-1)\delta}{n} \to \delta > 0 \\
\Rightarrow &\lim_{n\to\infty} \hat{F}^s_{cn}(x;\omega) - F^s_{c0}(x) > 0.
\end{aligned}
$$

Similarly we can show that

if $\hat{F}^s_{cn}(x_0;\omega) - F^s_{c0}(x_0) < -\delta$, for $x \in (x_0 - \epsilon, x_0]$, then $\lim_{n\to\infty} \hat{F}^s_{cn}(x;\omega) - F^s_{c0}(x) < 0$.

Either of them implies $\lim\limits_{n\to\infty} \hat{F}^s_{cn}(x;\omega) \neq F^s_{c0}(x)$. It follows that either $(x_0 - \epsilon, x_0] \subset D_c$ or $[x_0, x_0 + \epsilon) \subset D_c$. Consequently $\nu((x_0 - \epsilon, x_0]) = 0$ or $\nu([x_0, x_0 + \epsilon)) = 0$. In other words, x_0 is not regular, contradicting the assumption that x_0 is a regular point. The contradiction implies that $\hat{F}^s_{cn}(x_0;\omega) \to F^s_{c0}(x_0)$. □

The next two propositions give the consistency at the continuity point of F^s_{c0}.

Proposition 4.2. *Suppose that $E(K) < \infty$, A2 holds, $(a, b) \subset \Gamma_c$ and $(a, b) \subset \mathcal{S}$.*

(1) If x is a continuity point of $F^s_{c0}(\cdot)$ in (a, b), then $\hat{F}^s_{cn}(x) \stackrel{a.s.}{\to} F^s_{c0}(x)$.

(2) If $F^s_{c0}(a) = 0$, $F^s_{c0}(b-) = p_c$, and $F^s_{c0}(\cdot)$ is continuous at x, then $\hat{F}^s_{cn}(x) \stackrel{a.s.}{\to} F^s_{c0}(x)$.

Proof: Fix an $\omega \in \Omega_{\nu,c}$. Let x_0 be a continuity point of $F^s_{c0}(\cdot)$ in (a, b) which is also an interior point of $\mathcal{S}$. We shall show that $x_0 \notin D_c$. Otherwise, by the similar arguements as in the proof of Proposition 3.1, $\exists\ \epsilon > 0$ such that $[x_0, x_0 + \epsilon) \subset D_c$ and $(x_0 - \epsilon, x_0] \subset D_c$. Since $(a, b) \subset \mathcal{S}$, $\exists$ support points $x_1, x_2 \in (a, b)$ and $\eta > 0$ such that $(x_1 - \eta, x_1 + \eta) \subset (x_0 - \epsilon, x_0]$ and $(x_2 - \eta, x_2 + \eta) \subset [x_0, x_0 + \epsilon)$. Moreover, since $\nu(x_1 - \eta, x_1 + \eta) > 0$, $\nu(x_2 - \eta, x_2 + \eta) > 0$, we have $\nu(x_0 - \epsilon, x_0] > \nu(x_1 - \eta, x_1 + \eta) > 0$ and $\nu[x_0, x_0 + \epsilon) > \nu(x_2 - \eta, x_2 + \eta) > 0 \Rightarrow \nu(D_c) > 0$ which leads to a contradiction. This implies that the first statement of the proposition holds.

Hereafter, assume that $F^s_{c0}(a) = 0$ and $F^s_{c0}(b-) = p_c$. Then by the monotonicity of the sub-distribution function, $\forall$ point $x \le a$, $0 \le F^s_{c0}(x) \le F^s_{c0}(a) = 0$, thus $\lim_{x\uparrow a} F^s_{c0}(x) = F^s_{c0}(a) = \lim_{x\downarrow a} F^s_{c0}(x)$ (due to the right continuity of a sub-distribution function). It follows that each $x \le a$ is a continuity point of F^s_{c0}. In a similar manner, we can show that each $x \ge b$ is a continuity point of F^s_{c0}.

First prove $a \notin D_c$, *i.e.*, $\hat{F}^s_{cn}(a;\omega) \to F^s_{c0}(a)$. Otherwise, since $F^s_{c0}(a) = 0$ for all $x \le a$, $F^s_{c0}(x) = 0$, a is a continuous point of $F^s_{c0}(x)$, by the similar arguments as in the first paragraph, $\exists\ \delta > 0$ such that $[a, a+\delta) \subset D_c$. By the assumption, $\exists$ some support point $x \in (a, a+\delta)$ such that $\nu(x-\delta_0, x+\delta_0) > 0$ for some $\delta_0 > 0$, thus we have $\nu[a, a+\delta) > 0$. It implies $\nu(D_c) > 0$, which leads to a contradiction. So $\hat{F}^s_{cn}(a;\omega) \to F^s_{c0}(a)$.

Furthermore, if $F^s_{c0}(a) = 0$, then by the monotonicity of the sub-distribution functions, $\forall$ point $x \le a$, $0 \le \hat{F}^s_{cn}(x;\omega) \le \hat{F}^s_{cn}(a;\omega) \to F^s_{c0}(a) = 0$, which implies that $\hat{F}^s_{cn}(x;\omega) \to F^s_{c0}(x)$. Similarly, we can get the same result for each point $x \ge b$. These two resuts together with statement (1) yield statement (2). $\square$

Proposition 4.3. *Suppose that $E(K) < \infty$ and A2 holds. If every point of increase of $F^s_{c0}(\cdot)$ in Γ_c is strongly regular, then $\hat{F}^s_{cn}(x) \overset{a.s.}{\to} F^s_{c0}(x)$ $\forall$ continuity point x of $F^s_{c0}(\cdot)$ in Γ_c.*

Proof: Fix an $\omega \in \Omega_{\nu,c}$. We show that every continuity point of $F^s_{c0}(\cdot)$ in Γ_c is not in D_c. Let x_0 be a continuity point of $F^s_{c0}(\cdot)$. If x_0 is a point of increase of $F^s_{c0}(\cdot)$, then by the assumption, it is strongly regular, hence it is regular and $x_0 \notin D_c$ by Proposition 4.1. If x_0 is not a point of increase of $F^s_{c0}(\cdot)$ and $x_0 \in D_c$, we shall show that it leads to a contradiction.

Now suppose $x_0 \in D_c$, then $\lim_{n\to\infty} \hat{F}^s_{cn}(x_0;\omega) \ne F^s_{c0}(x_0)$. There are two possibilities: (1) $\lim_{n\to\infty} \hat{F}^s_{cn}(x_0;\omega) > F^s_{c0}(x_0)$ and (2) $\lim_{n\to\infty} \hat{F}^s_{cn}(x_0;\omega) < F^s_{c0}(x_0)$.

In case (1), let $b := \sup\{x : F^s_{c0}(x) = F^s_{c0}(x_0)\}$. Then $\forall\ \delta > 0$, $F^s_{c0}(b+\delta) - F^s_{c0}(b-\delta) > 0$ by the monotonicity and the definition of b, thus b is a point of increase of $F^s_{c0}(\cdot)$ and $b > x_0$.

By the monotonicity of $\hat{F}^s_{cn}(x;\omega)$, $\forall\ x \in [x_0, b)$,

$$\lim_{n\to\infty} \hat{F}^s_{cn}(x;\omega) \geq \lim_{n\to\infty} \hat{F}^s_{cn}(x_0;\omega) > F^s_{c0}(x_0) = F^s_{c0}(x),$$

thus $[x_0, b) \subset D_c$. Since b is strongly regular by our assumption, $\nu(D_c) \geq \nu((x_0, b)) > 0$, which leads to a contradiction.

Similarly in case (2), let $a := \inf\{x : F^s_{c0}(x) = F^s_{c0}(x_0)\}$. Then a is a point of increase of $F^s_{c0}(\cdot)$ and $a < x_0$. For each $x \in [a, x_0)$,

$$\lim_{n\to\infty} \hat{F}^s_{cn}(x;\omega) \leq \lim_{n\to\infty} \hat{F}^s_{cn}(x_0;\omega) < F^s_{c0}(x_0) = F^s_{c0}(x),$$

thus $[a, x_0) \subset D$. Since a is strongly regular by our assumption, $\nu(D_c) \geq \nu((a, x_0)) > 0$, which also leads to a contradiction. $\square$

These propositions together with Corollary 3.4 yield the following results on the pointwise consistency in open intervals and in the entire Γ_c.

Corollary 4.1. *Suppose that $E(K) < \infty$, A2 holds, $(a, b) \subset \Gamma_c \cap \mathcal{S}$, and $\nu(\{x\}) > 0$ if x is a discontinuity point of $F^s_{c0}(\cdot)$ and $x \in (a, b)$. Then $\hat{F}^s_{cn}(x) \xrightarrow{a.s.} F^s_{c0}(x)\ \forall\ x \in (a, b)$. Moreover, if $F^s_{c0}(a) = 0$ and $F^s_{c0}(b-) = p_c$, then $\hat{F}^s_{cn}(x) \xrightarrow{a.s.} F^s_{c0}(x)\ \forall\ x \in \Gamma_c$.*

Corollary 4.2. *Suppose that $E(K) < \infty$, A2, holds, every point of increase of $F^s_{c0}(\cdot)$ is strongly regular and $\nu(\{x\}) > 0$ for each discontinuity point x of $F^s_{c0}(\cdot)$. Then $\hat{F}^s_{cn}(x) \xrightarrow{a.s.} F^s_{c0}(x)\ \forall\ x \in \Gamma_c$.*

The next two propositions give the uniform consistency in an interval.

Proposition 4.4. *Suppose that $E(K) < \infty$, A2 holds and $F^s_{c0}(\cdot)$ is continuous. If $x < y < \tau_c$ and $0 < F^s_{c0}(x) < F^s_{c0}(y) < p_c$ imply $\nu((x, y)) > 0$, then $\sup_{x\in\Gamma_c} |\hat{F}^s_{cn}(x) - F^s_{c0}(x)| \xrightarrow{a.s.} 0$.*

Proof: Fix an $\omega \in \Omega_{\nu,c}$. We shall first show that $D_c = \emptyset$. Otherwise, $\exists\ x_0 \in D_c$.

If $\lim_{n\to\infty} \hat{F}^s_{cn}(x_0;\omega) - F^s_{c0}(x_0) = d > 0$ for some d, then letting $b := \sup\{x \in \Gamma_c : F^s_{c0}(x) = F^s_{c0}(x_0)\}$, we have $\forall\ \delta > 0$, $F^s_{c0}(b+\delta) - F^s_{c0}(b-\delta) > 0$ by the monotonicity and the definition of b. So b is a point of increase of $F^s_{c0}(\cdot)$. Since $F^s_{c0}(\cdot)$ is continuous by the assumption, $0 < F^s_{c0}(b+\delta_0) - F^s_{c0}(b) < d/2 < d$ for some $\delta_0 > 0$, and $\forall\ x \in (b, b+\delta_0)$,

$$
\begin{aligned}
&\lim_{n\to\infty} \hat{F}^s_{cn}(x;\omega) - F^s_{c0}(x) \\
=&\lim_{n\to\infty} \hat{F}^s_{cn}(x;\omega) - \lim_{n\to\infty} \hat{F}^s_{cn}(x_0;\omega) + \lim_{n\to\infty} \hat{F}^s_{cn}(x_0;\omega) - F^s_{c0}(x_0) + F^s_{c0}(x_0) - F^s_{c0}(x) \\
\geq& d - d/2 = d/2 > 0
\end{aligned}
$$

by the monotonicity. It follows that $(b, b+\delta_0) \subset D_c$. $\Rightarrow \nu(D_c) > 0$, which leads to a contradiction. The proof for the case $\lim_{n\to\infty} \hat{F}^s_{cn}(x_0;\omega) - F^s_{c0}(x) = d < 0$ is similar. Thus $D_c = \emptyset$ and $\hat{F}^s_{cn}(\cdot;\omega)$ point-wisely converges to $F^s_{c0}(\cdot)$.

Given $\epsilon > 0$ and given $x_0 \in \Gamma_c$, by continuity and monotonicity of $F^s_{c0}(\cdot)$, we can choose finitely many points $\{a_0, a_1, \cdots, a_k\}$ such that $a_0 < a_1 < \cdots < a_k$, $a_0 = -\infty$, $0 \leq F^s_{c0}(a_i) - F^s_{c0}(a_{i-1}) < \epsilon$ for each $i = 1, 2, \cdots, k$. Then $\exists\ N > 0$ such that if $n > N$ then $|\hat{F}^s_{cn}(a_i;\omega) - F^s_{c0}(a_i)| < \epsilon$ for all $i = 1, 2, \cdots, k$ by the point-wise convergence given in the last paragraph. Since $x_0 \in (a_{j-1}, a_j)$ for some $j \leq k$, $|F^s_{c0}(y_0) - F^s_{c0}(x_0)| < \epsilon\ \forall\ y_0 \in [a_{j-1}, a_j]$. Moreover, $\forall\ n, m > N$,

$$
\begin{aligned}
&|\hat{F}^s_{cn}(a_i;\omega) - \hat{F}^s_{cm}(a_{i-1};\omega)| \\
\leq&|\hat{F}^s_{cn}(a_i;\omega) - F^s_{c0}(a_i)| + |F^s_{c0}(a_i) - F^s_{c0}(a_{i-1})| + |F^s_{c0}(a_{i-1}) - \hat{F}^s_{cm}(a_{i-1};\omega)| < 3\epsilon.
\end{aligned}
$$

Then by the monotonicity of $\hat{F}^s_{cn}(\cdot;\omega)$ and $\hat{F}^s_{cm}(\cdot;\omega)$, we obtain

$$
\begin{aligned}
|\hat{F}^s_{cn}(x_0;\omega) - \hat{F}^s_{cm}(x_0;\omega)| \leq& |\hat{F}^s_{cn}(x_0;\omega) - \hat{F}^s_{cn}(a_i;\omega)| + |\hat{F}^s_{cn}(a_i;\omega) - \hat{F}^s_{cn}(a_{i-1};\omega)| \\
&+ |\hat{F}^s_{cn}(a_{i-1};\omega) - \hat{F}^s_{cm}(a_i;\omega)| + |\hat{F}^s_{cm}(a_i;\omega) - \hat{F}^s_{cm}(a_{i-1};\omega)| \\
&+ |\hat{F}^s_{cm}(a_{i-1};\omega) - \hat{F}^s_{cm}(x_0;\omega)| < 11\epsilon.
\end{aligned}
$$

Since x_0 is arbitrary, thus we can say $\hat{F}^s_{cn}(\cdot;\omega)$ converges uniformly to $F^s_{c0}(\cdot)$. $\square$

Since a follow-up study often lasts for a certain period of time, say $[\tau_1, \tau_2]$, so it is likely that $F^s_{c0}(\tau_2) < p_c$ or $F^s_{c0}(\tau_1) > 0$. Then the condition in Proposition 4.4 is not satisfied. Thus we present the following proposition.

Proposition 4.5. *Suppose that for each $c \in \mathcal{C}_r$,*

(1) $F^s_{c0}(\cdot)$ is continuous on $(\tau_1, \tau_2]$;

(2) either $\nu(\{\tau_1\}) > 0$ or $F^s_{c0}(\tau_1) = 0$;

(3) either $\nu(\{\tau_2\}) > 0$ or $F^s_{c0}(\tau_2-) = p_c$;

(4) for all a and b in $(\tau_1, \tau_2) \cap \Gamma_c$, $0 < F^s_{c0}(a) < F^s_{c0}(b) < p_c$ implies $\nu((a, b)) > 0$.

Then the GMLE satisfies $\sup_{x\in[\tau_1,\tau_2]\cap\Gamma_c} |\hat{F}^s_{cn}(x) - F^s_{c0}(x)| \to 0$ a.s.

Proof: We only give the proof for the case $\nu(\{\tau_1\}) > 0$ and $F^s_{c0}(\tau_2-) = p_c$. We shall show in the next paragraph that $\nu(D_c \cap [\tau_1, \tau_2]) = 0$, which implies that $\hat{F}^s_{cn}(x;\omega) \to F^s_{c0}(x)$ point-wisely for all $x \in [\tau_1, \tau_2] \cap \Gamma_c$ and by the continuity assumption on $F^s_{c0}(\cdot)$, we can have the convergence is uniform on $[\tau_1, \tau_2] \cap \Gamma_c$.

By Corollary 3.4, we have $\hat{F}^s_{cn}(\tau_1;\omega) \to F^s_{c0}(\tau_1)$. So it is done if $F^s_{c0}(\tau_1) = p_c$. Now assume that $F^s_{c0}(\tau_1) < p_c$. Then $D_1 = D_c \cap [\tau_1, \tau_2]$ is empty. Otherwise, by the continuity of $F^s_{c0}(\cdot)$, and the monotonicity of $F^s_{c0}(\cdot)$ and $\hat{F}^s_{cn}(\cdot;\omega)$, we have

$$\lim_{n\to\infty} \hat{F}^s_{cn}(\tau_1;\omega) = F^s_{c0}(\tau_1) < F^s_{c0}(\tau_2-) = p_c.$$

Then by an arguement similar to the proof of Proposition 4.4, we can show that D_1 contains an open interval (a, b) such that $0 < F^s_{c0}(a) < F^s_{c0}(b) < p_c$ and $\tau_1 < a < b < \tau_2$. However, it leads to a contradiction, as $\nu(D_c) \geq \nu(D_1) \geq \nu((a, b)) > 0$. ▫

§5. Proofs of the theorems in §3.

Proof of Theorem 3.1. We shall prove this theorem in three steps.

Step 1 (preliminary). By the strong law of large numbers (SLLN) for each $F \in \mathcal{F}$,

$\mathcal{L}_n(F) = \frac{1}{n}\sum_{i=1}^{n} \log \mu_F((L_i, R_i] \times \mathcal{M}_i) \overset{a.s.}{\rightarrow} E\{\log \mu_F((L, R] \times \mathcal{M})\} = \mathcal{L}(F)$, where

$$\begin{aligned}
&\mathcal{L}(F) = E(E(\log \mu_F((L, R] \times \mathcal{M})|K = k, \Delta = h)) \\
&= \sum_{k=1}^{\infty}\sum_{h=1}^{n_{\mathcal{P}}} P(K = k, \Delta = h) E([\sum_{i=0}^{k-1}\sum_{j=1}^{|P_h|} \mu_{F_0}((Y_{k,i}, Y_{k,i+1}] \times P_{hj}) \log \mu_F((Y_{k,i}, Y_{k,i+1}] \times P_{hj}) \\
&\quad + \mu_{F_0}((Y_{k,k}, Y_{k,k+1}) \times \mathcal{C}_r) \log \mu_F((Y_{k,k}, Y_{k,k+1}) \times \mathcal{C}_r)]|K = k, \Delta = h) \\
&= \sum_{k=1}^{\infty}\sum_{h=1}^{n_{\mathcal{P}}} P(K = k, \Delta = h) E(H_{F,k,h}(\mathbf{Y}, P_h)|K = k, \Delta = h) \\
&H_{F,k,h}(\mathbf{y}, P_h) = \sum_{i=0}^{k-1}\sum_{j=1}^{|P_h|} \mu_{F_0}((y_{k,i}, y_{k,i+1}] \times P_{hj}) \log \mu_F((y_{k,i}, y_{k,i+1}] \times P_{hj}) \\
&\qquad\qquad\qquad + \mu_{F_0}((y_{k,k}, y_{k,k+1}) \times \mathcal{C}_r) \log \mu_F((y_{k,k}, y_{k,k+1}) \times \mathcal{C}_r)
\end{aligned}$$

for $-\infty = y_{k,0} < y_{k,1} < \cdots < y_{k,k} < y_{k,k+1} = \infty$, and $0 \log 0 \overset{def}{=} 0$ and $\log 0 \overset{def}{=} -\infty$.

For each positive integer k and real numbers $y_{k,1} < y_{k,2} < \cdots < y_{k,k}$, verify that (1) $\sum_{i=0}^{k-1}\sum_{j=1}^{|P_h|} \mu_F((y_{k,i}, y_{k,i+1}] \times P_{hj}) + \mu_F((y_{k,k}, y_{k,k+1}) \times \mathcal{C}_r) = 1$ for each $F \in \mathcal{F}$; (2) $\sup\{|x \log x| : x \in [0, 1]\} < 1$; (3) $|H_{F_0,k,h}| \le (k+1) \cdot |P_h|$ (by (2)). It follows from Shannon-Kolmogorov inequality (see Ferguson, 1996) and statements (1) and (3) that $H_{F,k,h}(\mathbf{y}, P_h)$ is maximized by a function $F \in \mathcal{F}$ iff $\mu_F((y_{k,i}, y_{k,i+1}] \times P_{hj}) = \mu_{F_0}((y_{k,i}, y_{k,i+1}] \times P_{hj})$ and $\mu_F((y_{k,k}, y_{k,k+1}) \times \mathcal{C}_r) = \mu_{F_0}((y_{k,k}, y_{k,k+1}) \times \mathcal{C}_r)$, which are equivalent to

$$G_{\mathbf{F}^s}(y_{k,i+1}, P_{hj}) - G_{\mathbf{F}^s}(y_{k,i}, P_{hj}) = G_{\mathbf{F}_0^s}(y_{k,i+1}, P_{hj}) - G_{\mathbf{F}_0^s}(y_{k,i}, P_{hj}),$$

$$\text{for } 0 \le i < k \text{ and } 1 \le j \le |P_h|, \text{ and } 1 - G_{\mathbf{F}^s}(y_{k,k}, \mathcal{C}_r) = 1 - G_{\mathbf{F}_0^s}(y_{k,k}, \mathcal{C}_r).$$

Thus we can say $H_{F,k,h}(\mathbf{y}, P_h)$ is maximized by a function $F \in \mathcal{F}$ iff

$$G_{\mathbf{F}^s}(y_{k,i}, P_{hj}) = G_{\mathbf{F}_0^s}(y_{k,i}, P_{hj}) \text{ for } i = 1, 2, \cdots, k, \; j = 1, 2, \cdots, |P_h|. \tag{5.1}$$

Moreover, by statements (2) and (3),

$$|\mathcal{L}(F_0)| \le \sum_{k=1}^{\infty}\sum_{h=1}^{n_{\mathcal{P}}} P(K = k, \Delta = h)(k + 1) \cdot |P_h| \le J \cdot (E(K) + 1) < \infty.$$

Thus from Shannon-Kolmogorov inequality (see Ferguson, 1996) we have the statement:

AS1 (1) The cdf F_0 maximizes $\mathcal{L}(\cdot)$ over the set $\mathcal{F}$. (2) Any other F in $\mathcal{F}$ maximizes $\mathcal{L}(\cdot)$ iff $G_{\mathbf{F}^s} = G_{\mathbf{F}_0^s}$ a.e. μ (that is, for each possible (k,h), (5.1) holds a.e. (w.r.t. measure μ) in the set $\{(y_{k,i}, P_{hj}) : i = 1, \cdots, k$ and $j = 1, \cdots, |P_h|\}$).

Step 2 (existence of the limit of an arbitrary subsequence of the GMLE). Let $\mathcal{R}_2 = \{(l,r) : -\infty \le l < r \le +\infty\}$ and let α be an arbitrary positive integer, then there are finitely many extended real numbers,

$-\infty = q_0 < q_1 < \cdots < q_\beta < \infty$, such that $\mu((q_{i-1}, q_i) \times \mathcal{J}) < 2^{-\alpha}$.

Now form the sets $U_0, U_1, \cdots, U_{2\beta}$ by setting $U_{2i-1} = (q_{i-1}, q_i),\ i = 1, 2, \cdots, \beta$, $U_{2i} = [q_i, q_i],\ i = 0, 2, \cdots, \beta$. Let $\mathcal{U}_\alpha$ denote the collection of all nonempty sets of the form $U_{ij} = \mathcal{R}_2 \cap (U_i \times U_j),\ 0 \le i \le j \le 2\beta$. Then we take $\mathcal{U} = \bigcup_\alpha \mathcal{U}_\alpha$.

Let Ω be the sample space. For each $\omega \in \Omega$, let $\{\hat{\mathbf{F}}_n^s(t,\omega)\} = \{(\hat{F}_{1n}^s(t,\omega), \cdots, \hat{F}_{Jn}^s(t,\omega)\}$ be a sequence of GMLEs of $\mathbf{F}_0^s(\cdot)$, since $\hat{F}_{jn}^s(\cdot;\omega)$ is monotone and bounded for each $j = 1, \cdots, J$, by Helly's selection theorem, $\exists$ a subsequence $\{n'\}$ such that $\hat{\mathbf{F}}_n^s(t,\omega) \to \mathbf{F}^s(t,\omega)$ where $\mathbf{F}^s(t,\omega)$ is the corresponding limit function and denoted by $\mathbf{F}_\omega^s$. So $G_{\hat{\mathbf{F}}_{n'}^s(\cdot,\omega)}(t,m) = \sum_{j\in m} \hat{F}_{jn'}^s(t;\omega) \to \sum_{j\in m} F_j^s(t;\omega) = G_{\mathbf{F}_\omega^s}(t,m)$ point-wise for each $t \in \Re$, $m \in \mathcal{J}$ and for some $F \in \mathcal{F}$ with the corresponding function F in $\mathcal{F}$ by $F_\omega(t,c) = \sum_{j\le c}(\mathbf{F}_\omega^s(t))_j$, where $c \in \mathcal{C}_r$. Now let Ω_0 be the set of all sample points ω such that each point-wise limit $G_{\mathbf{F}_\omega^s}$ of the sequence $\{G_{\hat{\mathbf{F}}_n^s(\cdot,\omega)}\}$ satisfies $\mathcal{L}(F_\omega) \ge \mathcal{L}(F_0)$, thus for each $\omega \in \Omega_0$, all the limit points of $\{G_{\hat{\mathbf{F}}_n^s(\cdot,\omega)}\}$ equal $G_{\mathbf{F}_0^s}$ a.e. μ in view of (AS1) in Step 1 and this implies

$$\lim_{n'\to\infty} \int_{\Re\times\mathcal{J}} |G_{\hat{\mathbf{F}}_{n'}^s(\cdot,\omega)}(t,m) - G_{\mathbf{F}_0^s}(t,m)| d\mu(t,m) = 0.$$

Thus the desired result follows if we can show Ω_0 has probability 1.

Let $\hat{Q}_n$ be the empirical estimator of Q, the distribution of $(L, R, \mathcal{M})$. By SLLN, $\Omega_{U,W} = \{\omega : \hat{Q}_n(U, W; \omega) \to Q(U, W)\}$ has probability 1 for each Borel subset U of $\mathcal{R}_2$ and

$W \in \mathcal{J}$, so does $\Omega' = \{\omega : \mathcal{L}_n(F_0;\omega) \to \mathcal{L}(F_0)\}$ and $\Omega^* = \Omega' \cap (\bigcap_{U\in\mathcal{U}} \bigcap_{W\in\mathcal{J}} \Omega_{U,W})$, thus $P(\Omega^*) = 1$. We are done if we can show that $\Omega^* \subset \Omega_0$.

For $\omega^* \in \Omega^*$, to simplify the notations, let $G_n^*(t,m) = G_{\hat{\mathbf{F}}_n^s(\cdot,\omega^*)}(t,m)$ which is the GMLE of the distribution function, for $t \in \Re$, $m \in \mathcal{J}$ and

$$Q_n(U,W) = \hat{Q}_n(U,W;\omega^*),\ U \in \mathcal{B}(\mathcal{R}_2),\ W \in \mathcal{J}.$$

Without loss of the generality, assume $\{n\} = \{n'\}$. Let G^* be a point-wise limiting function of $\{G_n^*\}$ where G^* denotes $G_{\mathbf{F}^s_{\omega^*}}$. If denoting F_{ω^*} by F_*, obviously $\mathcal{L}(F_*) \le \mathcal{L}(F_0)$. Also, $\mathcal{L}(F_0) \le \lim\inf_{n\to\infty} \mathcal{L}_n(\hat{F}_n;\omega^*)$, because $\mathcal{L}_n(F_0;\omega^*) \le \mathcal{L}_n(\hat{F}_n;\omega^*)$ by the definition of GMLE and the fact that $\mathcal{L}_n(F_0;\omega^*) \to \mathcal{L}(F_0)$ by the choice of ω^*. If we can show that

$$\limsup_{n\to\infty} \mathcal{L}_n(\hat{F}_n;\omega^*) \le \mathcal{L}(F_*) \tag{5.2}$$

then $\mathcal{L}(F_0) \le \mathcal{L}(F_*)$. Notice that $\mathcal{L}(F_*)$ depends on $\omega^* \in \Omega^*$. It will further conclude that Ω_0 contains Ω^* by the arbitrary choice of ω^*, and thus has probability 1. In addition,

$$\limsup_{n\to\infty} \int_{\Re\times\mathcal{J}} |G_{\hat{\mathbf{F}}_n^s}(t,m) - G_{\mathbf{F}_0^s}(t,m)| d\mu(t,m) = 0 \tag{5.3}$$

for each $\omega \in \Omega_0$. Thus this theorem follows if we can prove (5.2).

Step 3 (to prove that statement (5.2) holds for each $\omega^* \in \Omega^*$). Notice that

$$\mathcal{L}_n(F_n^*) = \int_{\mathcal{R}_2\times\mathcal{J}} \log(\mu_{F_n^*}((l,r]\times m))dQ_n(l,r,m)$$

Inequality (5.2) can be written as

$$\limsup_{n\to\infty} \int_{\mathcal{R}_2\times\mathcal{J}} \log(\mu_{F_n^*}((l,r]\times m))dQ_n(l,r,m) \le \int_{\mathcal{R}_2\times\mathcal{J}} \log(\mu_{F_*}((l,r]\times m))dQ(l,r,m) \tag{5.4}$$

Fix a positive integer α and a negative integer γ. Then

$$\begin{aligned}\int_{\mathcal{R}_2\times\mathcal{J}} \log(\mu_{F_n^*}((l,r]\times m))dQ_n(l,r,m) &\le \int_{\mathcal{R}_2\times\mathcal{J}} \gamma \vee \log((\mu_{F_n^*}((l,r]\times m))dQ_n(l,r,m) \\ &\le \sum_{U\in\mathcal{U}_\alpha, W\in\mathcal{J}} M_n(U,W)Q_n(U,W)\end{aligned}$$

where $M_n(U,W) = \sup\{\gamma \vee \log(\mu_{F_n^*}((l,r] \times W)) : (l,r) \in \overline{U}\}$, $\overline{U}$ is the closure of U. Let $r_U = \sup\{r : (l,r) \in \overline{U}\}$ and $l_U = \inf\{l : (l,r) \in \overline{U}\}$. Then for each $U \in \mathcal{U}_\alpha$ and $W \in \mathcal{J}$,

(1) $r_U = \sup\{r : (l,r) \in U\}$ and $l_U = \inf\{l : (l,r) \in U\}$, as $(l_U, r_U) \in \overline{U}$;

(2) $\gamma \vee \log(x)$ is a bounded continuous function of x on $(0,1]$ for each $\gamma \in (-\infty, 0]$;

(3) $\mu_{F_n^*}(B) \to \mu_{F_*}(B)$ for each B with the form $(a,b] \times W$.

$$\begin{aligned}
M_n(U,W) =& \sup\{\gamma \vee \log(\mu_{F_n^*}((l,r] \times W)) : (l,r) \in \overline{U}\} \\
=& \gamma \vee \log(\mu_{F_n^*}((l_U, r_U] \times W)) \qquad \text{(by definitions of } l_U,\ r_U \text{ and } M_n(U,W)) \\
\to& \gamma \vee \log(\mu_{F_*}((l_U, r_U] \times W)) \text{ (by the aforementioned statements (2) and (3))} \\
=& \sup\{\gamma \vee \log(\mu_{F_*}((l,r] \times W)) : (l,r) \in \overline{U}\} \\
\overset{def}{=}& M(U,W),
\end{aligned}$$

By the choice of ω^*, $Q_n(U,W) \to Q(U,W)$ for all $U \in \mathcal{U}_\alpha$ and $W \in \mathcal{J}$. It follows that

$$\sum_{U \in \mathcal{U}_\alpha, W \in \mathcal{J}} M_n(U,W) Q_n(U,W) \to \sum_{U \in \mathcal{U}_\alpha, W \in \mathcal{J}} M(U,W) Q(U,W).$$

Let $m(U,W) = \inf\{\gamma \vee \log(\mu_{F_*}((l,r] \times W)) : (l,r) \in \overline{U}\}$. It follows from the bound that $|\gamma \vee \log(x) - \gamma \vee \log(y)| \leq e^{-\gamma}|y - x|$ for $0 \leq x,\ y \leq 1$, we have the inequality

$$\begin{aligned}
& M(U,W) - m(U,W) \leq e^{-\gamma} \sup\{\mu_{F_*}((l_U, r_U] \times W) - \mu_{F_*}((l,r] \times W) : (l,r) \in \overline{U}\} \\
=& e^{-\gamma} \sup\{\sum_{j \in W} (F^s_{*j}(r_U) - F^s_{*j}(l_U)) - \sum_{j \in W} (F^s_{*j}(r) - F^s_{*j}(l)) : (l,r) \in \overline{U}\} \\
=& e^{-\gamma} \sup\{\sum_{j \in W} (F^s_{*j}(r_U) - F^s_{*j}(r) + F^s_{*j}(l) - F^s_{*j}(l_U)) : (l,r) \in \overline{U}\} \\
=& e^{-\gamma} \sup\{\sum_{j \in W} (F^s_{*j}(r_U) - F^s_{*j}(r)) + \sum_{j \in W} (F^s_{*j}(l) - F^s_{*j}(l_U)) : (l,r) \in \overline{U}\} \\
=& e^{-\gamma} \sup\{\mu_{F_*}((r, r_U] \times W) + \mu_{F_*}((l_U, l] \times W) : (l,r) \in \overline{U}\} \text{ for } U \in \mathcal{U}_\alpha,\ W \in \mathcal{J}.
\end{aligned}$$

This implies that

(1) if $U = \mathcal{R}_2 \cap [(q_{i-1}, q_i) \times (q_{j-1}, q_j)]$, then $M(U,W) - m(U,W) > 2/\alpha$ implies either $\mu_{F_*}((q_{i-1}, q_i] \times W) > e^\gamma/\alpha$ or $\mu_{F_*}((q_{j-1}, q_j] \times W) > e^\gamma/\alpha$;

(2) if $U = \mathcal{R}_2 \cap [[q_i, q_i] \times (q_{j-1}, q_j)]$, then $M(U,W) - m(U,W) > 2/\alpha$ implies $\mu_{F_*}((q_{j-1}, q_j] \times W) > e^\gamma/\alpha$;

(3) if $U = \mathcal{R}_2 \cap [(q_{i-1}, q_i) \times [q_j, q_j]]$, then $M(U,W) - m(U,W) > 2/\alpha$ implies $\mu_{F_*}((q_{i-1}, q_i] \times W) > e^\gamma/\alpha$;

(4) if $U = \mathcal{R}_2 \cap [\{a\} \times \{b\}]$, then $M(U,W) - m(U,W) = 0$.

The aforementioned statements yield

$$\begin{aligned}
&\sum_{U \in \mathcal{U}_\alpha, W \in \mathcal{J}} (M(U,W) - m(U,W))Q(U,W) \\
\leq & 2/\alpha + |\gamma| \sum_{U \in \mathcal{U}_\alpha, W \in \mathcal{J}} Q(U,W)\mathbf{1}((M(U,W) - m(U,W)) \geq 2/\alpha) \\
\leq & 2/\alpha + |\gamma| \sum_{i=1}^{\beta} \sum_{W \in \mathcal{J}} P(q_{i-1} < L < q_i, \mathcal{M} = W)\mathbf{1}(\mu_{F_*}((q_{i-1}, q_i] \times W) > e^\gamma/\alpha) \\
& + |\gamma| \sum_{i=1}^{\beta} \sum_{W \in \mathcal{J}} P(q_{i-1} < R < q_i, \mathcal{M} = W)\mathbf{1}(\mu_{F_*}((q_{i-1}, q_i] \times W) > e^\gamma/\alpha) \\
\leq & 2/\alpha + |\gamma| \sum_{i=1}^{\beta} P(q_{i-1} < L < q_i, \mathcal{M} \in \mathcal{J})\mathbf{1}(\mu_{F_*}((q_{i-1}, q_i] \times \mathcal{C}_r) > e^\gamma/\alpha) \\
& + |\gamma| \sum_{i=1}^{\beta} P(q_{i-1} < R < q_i, \mathcal{M} \in \mathcal{J})\mathbf{1}(\mu_{F_*}((q_{i-1}, q_i] \times \mathcal{C}_r) > e^\gamma/\alpha) \\
\leq & 2/\alpha + |\gamma|(1 + \alpha e^{-\gamma})2^{1-\alpha}.
\end{aligned}$$

The last step make use of the inequality

$$P(q_{i-1} < L < q_i, \mathcal{M} \in \mathcal{J}) + P(q_{i-1} < R < q_i, \mathcal{M} \in \mathcal{J}) \leq 2\mu((q_{i-1}, q_i) \times \mathcal{J}) \leq 2^{1-\alpha}$$

and the fact that at most $1 + \alpha e^{-\gamma}$ of the terms

$$\mu_{F_*}((q_0, q_1] \times \mathcal{C}_r),\ \mu_{F_*}((q_1, q_2] \times \mathcal{C}_r),\ \cdots,\ \mu_{F_*}((q_{\beta-1}, q_\beta] \times \mathcal{C}_r) \text{ exceed } e^\gamma/\alpha.$$

Thus we have

$$
\begin{aligned}
&\limsup_{n\to\infty} \int_{\mathcal{R}_2\times\mathcal{J}} \log(\mu_{F_n^*}((l,r]\times m))dQ_n(l,r,m) \\
&\le \int_{\mathcal{R}_2\times\mathcal{J}} \gamma \vee \log(\mu_{F_*}((l,r]\times m))dQ(l,r,m) + 2/\alpha + |\gamma|(1+\alpha e^{-\gamma})2^{1-\alpha} \\
&\to \int_{\mathcal{R}_2\times\mathcal{J}} \gamma \vee \log(\mu_{F_*}((l,r]\times m))dQ(l,r,m) \text{ as } \alpha\to\infty \\
&\to \int_{\mathcal{R}_2\times\mathcal{J}} \log(\mu_{F_*}((l,r]\times m))dQ(l,r,m) \text{ as } \gamma\to-\infty \text{ (which is (5.4) or (5.2)).} \ \square
\end{aligned}
$$

Proof of Theorem 3.2. Given an $A \in \mathcal{A}$, define an event $T_I(A) = \{L \in A \text{ or } R \in A\}$, then $P(T_I(A) \cap \{\mathcal{M} \in W\}) = P(T_I(A))P(\mathcal{M} \in W|T_I(A))$, which is equivalent to

$$
\begin{aligned}
&\mu(B) = \sum_{k=1}^{\infty} P(K=k) \sum_{i=1}^{k} P(Y_{k,i} \in A|K=k)P(\mathcal{M} \in W|L \in A \text{ or } R \in A) \\
&= \nu(A)P(\mathcal{M} \in W|T_I(A)) \text{ with } \nu(A) > 0.
\end{aligned}
$$

By Lemma 3.2 and the definition of Γ_c, for almost all $t \in \Gamma_c \cap \mathcal{S}$, F is identifiable at (t,c). Then by Remark 3.1, $\exists\, W_1, \cdots, W_w$, $w < \infty$ with $f_{\mathcal{M}|T_I(\{t\})}(W_i) = P(\mathcal{M} = W_i|T_I(\{t\})) > 0$ and some constants g_i's such that

$$
\phi(\{c\}) = \textstyle\sum_{i=1}^{w} g_i \phi(W_i) \text{ for almost all } t \in \Gamma_c \cap \mathcal{S}. \tag{5.5}
$$

By Theorem 3.1,

$$
\begin{aligned}
&\int |G_{\hat{\mathbf{F}}_n^s} - G_{\mathbf{F}_0^s}| d\mu \overset{a.s.}{\to} 0 \Rightarrow \sum_{M\in\mathcal{J}} \int_{\Re} |G_{\hat{\mathbf{F}}_n^s}(t,M) - G_{\mathbf{F}_0^s}(t,M)| d\mu(t,M) \overset{a.s.}{\to} 0 \\
&\Rightarrow \sum_{M\in\mathcal{J}} \int_{\mathcal{S}} |G_{\hat{\mathbf{F}}_n^s}(t,M) - G_{\mathbf{F}_0^s}(t,M)| d\mu(t,M) \overset{a.s.}{\to} 0 \\
&\Rightarrow \sum_{M\in\mathcal{J}} \int_{\mathcal{S}\cap\Gamma_c} |G_{\hat{\mathbf{F}}_n^s}(t,M) - G_{\mathbf{F}_0^s}(t,M)| d\mu(t,M) \overset{a.s.}{\to} 0.
\end{aligned}
$$

Then we have for each $i = 1, 2, \cdots, w$, w.p.1,

$$
\begin{aligned}
&\int_{\mathcal{S}\cap\Gamma_c} |G_{\hat{\mathbf{F}}_n^s}(t,W_i) - G_{\mathbf{F}_0^s}(t,W_i)| d\mu(t,W_i) \to 0. \\
&\Rightarrow \int_{\mathcal{S}\cap\Gamma_c} |G_{\hat{\mathbf{F}}_n^s}(t,W_i) - G_{\mathbf{F}_0^s}(t,W_i)| f_{\mathcal{M}|T_I(\{t\})}(W_i) d\nu(t) \to 0.
\end{aligned}
$$

By Lemma 3.2, $f_{\mathcal{M}|T_I(\{t\})}(W_i) > 0$ for almost all $t \in \mathcal{S} \cap \Gamma_c$. Since $\mathcal{M}$ is finite and discrete, we can define $h_i(t) = f_{\mathcal{M}|T_I(\{t\})}(W_i)$, notice that $\mathcal{S}^*$ denotes the set of all points in $\mathcal{S}$ except the set of ν-measure zero, then $h_i(t) > 0\ \forall\ t \in \mathcal{S}^* \cap \Gamma_c$, *i.e.*, $\mathcal{S}^* \cap \Gamma_c \subseteq \{h_i(t) > 0\} = \{h_i(t) \geq 1/m\} \cup \{0 < h_i(t) < 1/m\}$ for each arbitrary positive integer m. If we define $A_m = \mathcal{S}^* \cap \Gamma_c \cap \{h_i(t) \geq 1/m\}$ and $B_m = \mathcal{S}^* \cap \Gamma_c \cap \{0 < h_i(t) < 1/m\}$, then $A_m \cup B_m = \mathcal{S}^* \cap \Gamma_c$ and $A_m \cap B_m = \emptyset$. Thus we have, w.p.1,

$$
\begin{aligned}
&\sum_{M\in\mathcal{J}} \int_{\mathcal{S}\cap\Gamma_c} |G_{\hat{\mathbf{F}}_n^s}(t,M) - G_{\mathbf{F}_0^s}(t,M)| d\mu(t,M) \\
=&\int_{\mathcal{S}\cap\Gamma_c} |G_{\hat{\mathbf{F}}_n^s}(t,W_i) - G_{\mathbf{F}_0^s}(t,W_i)| f_{\mathcal{M}|T_I(\{t\})}(W_i) d\nu(t) \\
=&\int_{\mathcal{S}^*\cap\Gamma_c} |G_{\hat{\mathbf{F}}_n^s}(t,W_i) - G_{\mathbf{F}_0^s}(t,W_i)| f_{\mathcal{M}|T_I(\{t\})}(W_i) d\nu(t) \\
=&\int_{A_m} |G_{\hat{\mathbf{F}}_n^s}(t,W_i) - G_{\mathbf{F}_0^s}(t,W_i)| h_i(t) d\nu(t) + \int_{B_m} |G_{\hat{\mathbf{F}}_n^s}(t,W_i) - G_{\mathbf{F}_0^s}(t,W_i)| h_i(t) d\nu(t) \\
\geq&\frac{1}{m}\int_{A_m} |G_{\hat{\mathbf{F}}_n^s}(t,W_i) - G_{\mathbf{F}_0^s}(t,W_i)| d\nu(t) + \int_{B_m} |G_{\hat{\mathbf{F}}_n^s}(t,W_i) - G_{\mathbf{F}_0^s}(t,W_i)| h_i(t) d\nu(t) \\
\geq&\frac{1}{m}\int_{A_m} |G_{\hat{\mathbf{F}}_n^s}(t,W_i) - G_{\mathbf{F}_0^s}(t,W_i)| d\nu(t),
\end{aligned}
$$

which implies $\lim_{n\to\infty} \int_{A_m} |\phi(W_i)(\hat{\mathbf{F}}_n^s(t) - \mathbf{F}_0^s(t))| d\nu(t) = 0$ w.p.1.

Since m is arbitrary, let $m \to \infty$, then $A_1 \subseteq A_2 \subseteq \cdots \subseteq \mathcal{S}^* \cap \Gamma_c$ and $\mathcal{S}^* \cap \Gamma_c \supseteq B_1 \supseteq B_2 \supseteq \cdots \supseteq \emptyset$ and $\nu(\mathcal{S} \cap \Gamma_c) < \infty$. Thus w.p.1,

$$
\begin{aligned}
&\lim_{n\to\infty} \int_{\mathcal{S}^*\cap\Gamma_c} |\phi(W_i)(\hat{\mathbf{F}}_n^s(t) - \mathbf{F}_0^s(t))| d\nu(t) \\
=&\lim_{n\to\infty} \Big(\int_{A_m} |\phi(W_i)(\hat{\mathbf{F}}_n^s(t) - \mathbf{F}_0^s(t))| d\nu(t) + \int_{B_m} |\phi(W_i)(\hat{\mathbf{F}}_n^s(t) - \mathbf{F}_0^s(t))| d\nu(t)\Big) \\
\leq&\lim_{n\to\infty} \Big(\int_{A_m} |\phi(W_i)(\hat{\mathbf{F}}_n^s(t) - \mathbf{F}_0^s(t))| d\nu(t) + \int_{B_m} |\phi(W_i)(\hat{\mathbf{F}}_n^s(t) + \mathbf{F}_0^s(t))| d\nu(t)\Big) \\
\leq&\lim_{n\to\infty} \Big(\int_{A_m} |\phi(W_i)(\hat{\mathbf{F}}_n^s(t) - \mathbf{F}_0^s(t))| d\nu(t) + 2\int_{B_m} d\nu(t)\Big) \\
&\text{(as } 0 \leq \phi(W_i)\hat{\mathbf{F}}_n^s(t), \phi(W_i)\mathbf{F}_0^s(t) \leq 1\text{)} \\
=&\lim_{n\to\infty} \Big(\int_{A_m} |\phi(W_i)(\hat{\mathbf{F}}_n^s(t) - \mathbf{F}_0^s(t))| d\nu(t) + 2\nu(B_m)\Big)
\end{aligned}
$$

$$= \lim_{n\to\infty} \int_{A_m} |\phi(W_i)(\hat{\mathbf{F}}_n^s(t) - \mathbf{F}_0^s(t))| d\nu(t) + 2\nu(B_m) \text{ (as } \nu \text{ is a finite positive measure).}$$

$$\Rightarrow \lim_{n\to\infty} \int_{\mathcal{S}^*\cap\Gamma_c} |\phi(W_i)(\hat{\mathbf{F}}_n^s(t) - \mathbf{F}_0^s(t))| d\nu(t) \le \lim_{m\to\infty} 2\nu(B_m) \overset{a.s.}{=} 0.$$

$$\Rightarrow \int_{\mathcal{S}^*\cap\Gamma_c} |g_i| \cdot |\phi(W_i)(\hat{\mathbf{F}}_n^s(t) - \mathbf{F}_0^s(t))| d\nu(t) \overset{a.s.}{\to} 0 \text{ (see (5.6)).} \qquad (5.6)$$

So we can have

$$\begin{aligned}
&\int_{\mathcal{S}^*\cap\Gamma_c} |\hat{F}_{cn}^s(t) - F_{c0}^s(t)| d\nu(t) \\
&= \int_{\mathcal{S}^*\cap\Gamma_c} |\phi(\{c\})(\hat{\mathbf{F}}_n^s(t) - \mathbf{F}_0^s(t))| d\nu(t) \\
&= \int_{\mathcal{S}^*\cap\Gamma_c} |(\sum_{i=1}^{w} g_i \phi(W_i))(\hat{\mathbf{F}}_n^s(t) - \mathbf{F}_0^s(t))| d\nu(t) \\
&\le \sum_{i=1}^{w} \int_{\mathcal{S}^*\cap\Gamma_c} |g_i||\phi(W_i)(\hat{\mathbf{F}}_n^s(t) - \mathbf{F}_0^s(t))| d\nu(t) \to 0 \text{ a.s. by (5.6).}
\end{aligned}$$

Let $\mathcal{S}^{*c}$ be the complement of $\mathcal{S}^*$. Since $\int_{\mathcal{S}^{*c}\cap\Gamma_c} |\hat{F}_{cn}^s(t) - F_{c0}^s(t)| d\nu(t) = 0$ by Theorem 3.1,

$$\begin{aligned}
&\int_{\Gamma_c} |\hat{F}_{cn}^s(t) - F_{c0}^s(t)| d\nu(t) \\
&= \int_{\mathcal{S}^*\cap\Gamma_c} |\hat{F}_{cn}^s(t) - F_{c0}^s(t)| d\nu(t) + \int_{\mathcal{S}^{*c}\cap\Gamma_c} |\hat{F}_{cn}^s(t) - F_{c0}^s(t)| d\nu(t) \\
&= \int_{\mathcal{S}^*\cap\Gamma_c} |\hat{F}_{cn}^s(t) - F_{c0}^s(t)| d\nu(t) \to 0 \text{ a.s. } \square
\end{aligned}$$

REFERENCES CITED

* Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T. and Silverman, E. (1995). An empirical distribution function for sampling incomplete information. *Annal of Mathematical Statistics* 26 641-647.

* Basu, S., Sen, A., and Banerjee, M. (2003). Bayesian analysis of competing risks with partially masked cause of failure. *Applied Statistics* 52 77-93.

* Crowder, M. (2001). Classical Competing Risks. Chapman & Hall/CRC.

* Flehinger, B. J., Reiser, B., and Yashchin, E. (1996). Inference about defects in the presence of masking. *Technometrics* 38 247-255.

* Ferguson, T. S. (1996). A course in large sample theory. Chapman & Hall.

* Gentleman, R. and Vandal, A. C. (2002). Nonparametric estimation of the bivariate CDF for arbitrarily censored data. *Canadian Journal of Statistics.* 30 557-571.

* Groeneboom, P., Maathuis, M. H. and Wellner, J. A. (2008). Current status data with competing risks: consistency and rates of convergence of the MLE. *Annals of Statistics* 36 1031-1063.

* Groeneboom, P. and Wellner, J. A. (1992). Information bounds and nonparametric maximum likelihood estimation. *Birkhäuser Verlag, Basel.*

* 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.

* Mukhopadhyay C., and Basu S. (2007). Bayesian Analysis of Masked Series System Lifetime Data. *Communications in Statistics - Theory And Methods* 36 329-348.

* Schick, A. and Yu, Q. (2000) Consistency of the GMLE with mixed case interval censored data. *Scandinavian Journal of Statistics* 27 45-55.

* Wang, J., Yu, Q. and Wong, G. Y. C. (2011). The random partition masking model for interval censored and masked competing risks data. submitted.

* Wellner, J. A. (1995). Interval censoring case 2: alternative hypotheses. In *Analysis of censored data*, Proceedings of the workshop on analysis of censored data, Dec. 28,

1994-Jan. 1, 1995, University of Pune, India. IMS Lecture Notes, Monograph Series 27 271-291. H. L. Koul and J. V. Deshpande, editors.

* 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.

§6. Appendix II.

Example 6.1. Consider a discrete case. Suppose that $J = 2$, $T \in \{1, 2\}$ and $C \in \{1, 2\}$. There are two partitions denoted by $P_1 = (\{1\}, \{2\})$ 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 the case 2 model with $P(Y_1 = 1, Y_2 = 2) = 1$. 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 the GMLE of F_0 is consistent for $t \in \{1\} \cup ([2, \infty)$ and is asymptotically normally distributed.

Proof of Example 6.1. Based on the model in Example 1.1, 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 GMLE 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, $||\hat{\mathbf{s}}(\omega) - \mathbf{s}^o|| < \frac{1}{n}\ \forall\ \omega \in \Omega$ 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. $\square$

Derivation of Example 1.1. Under given assumptions and notations in the example, the log likelihood function is

$$\mathcal{L}_1 \text{ln}(p_1 + p_2) + n_2 \text{ln}(1 - p_1 - p_2) + n_3 \text{ln}(p_1 + p_3) + n_4 \text{ln}(1 - p_1 - p_3),$$

denote $n_1 + n_2 + n_3 + n_4$ as the total number of the observations. Then the normal equations are $\frac{n_1}{p_1+p_2} - \frac{n_2}{1-p_1-P_2} + \frac{n_3}{p_1+p_3} - \frac{n_4}{1-p_1-p_3} = 0$, $\frac{n_1}{p_1+p_2} - \frac{n_2}{1-p_1-p_2} = 0$, $\frac{n_3}{p_1+p_3} - \frac{n_4}{1-p_1-p_3} = 0$ which reduce to $\frac{n_1}{p_1+p_2} - \frac{n_2}{1-p_1-p_2} = 0$, $\frac{n_3}{p_1+p_3} - \frac{n_4}{1-p_1-p_3} = 0$. Solving them leads to the GMLEs in Example 1.1.□

Derivation of Example 3.2. Based on the notations and the model, 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 the example. □

Proof of Lemma 3.2. Fix $c \in \mathcal{C}_r$. By assumption A2, there are some constant g_α's such that

$$\phi(\{c\}) = \sum_{\alpha=1}^{w} g_\alpha \phi(M_\alpha). \tag{6.1}$$

Given N values of K with the corresponding inspection times, say, $(k_i, y_{k_i,1}, \cdots, y_{k_i,k_i})$, $i = 1, \cdots, N$, the possible values of $(L, R, \mathcal{M})$ from these N sets of inspection times are $(-\infty, y_{k_i,1}, V_i)$, $(y_{k_i,1}, y_{k_i,2}, V_i)$, $\cdots$, $(y_{k_i,k_i-1}, y_{k_i,k_i}, V_i)$, $(y_{k_i,k_i}, \infty, \mathcal{C}_r)$, where $V_i \in P_{h_i}$ for some h_i. By A2 and the definitions of $\mathcal{S}$ and ν (defined before Definition 3.1), there is at least a finite N such that

(1) $P(\Delta = h_i | (Y_{k_i,1}, \cdots, Y_{k_i,k_i}) = (y_{k_i,1}, \cdots, y_{k_i,k_i}), K = k_i) > 0$, $i = 1$, ..., N;

(2) for each $\alpha \in \{1, ..., w\}$, $M_\alpha = V_{i_\alpha}$ for some $i_\alpha \in \{1, ..., N\}$;

(3) $\mathcal{S}_N = \{y_{k_i,j} : j = 1, \cdots, k_i, i = 1, 2, \cdots, N\} \subset \mathcal{S}$.

Let $t_1 < t_2 < \cdots < t_m$ be all the distinct values of the elements in $\mathcal{S}_N$.

Now we verify that F is identifiable at (r, c) with $r = \min_\alpha y_{k_{i_\alpha},1}$ by taking $(l_\alpha, r_\alpha] \times W_\alpha = (-\infty, y_{k_{i_\alpha},1}] \times V_{i_\alpha}$, $\alpha = 1, \cdots, w$ that satisfy the three conditions in Definition 3.2 as follows.

(1) Notice that $\max_\alpha l_\alpha = -\infty < r = \min_\alpha r_\alpha$ by the choices of l_α and r.

(2) Since $\{(-\infty, y_{k_i,1}] \times V_i : V_i \in P_{h_i}, i = 1, \cdots, N\} \supseteq \{(-\infty, y_{k_{i_\alpha},1}] \times M_\alpha : \alpha = 1, \cdots, w\}$, it implies that for each $M_\alpha = V_{i_\alpha}$, $\exists\ h_i$ with $P(\Delta = h_i | (L, R) = (-\infty, y_{k_i,1})) > 0$ such that $M_\alpha \in P_{h_i}$. Thus it implies $P(\mathcal{M} = M_\alpha | (L, R) = (-\infty, y_{k_{i_\alpha},1})) > 0$.

(3) We have $\phi(\{c\}) = \sum_{i=1}^{w} g_i \phi(M_i)$ by (6.1).

Thus F is identifiable at (r, c), where $r \in \mathcal{S}$. Let $\mathcal{S}_c$ be the set of all t in $\mathcal{S}$ such that F is identifiable at (t, c). Then we have just shown that $\mathcal{S}_c$ is not empty.

Now let $u_c \in \mathcal{S}_c$. Then $\exists \{(l_i, r_i] \times W_i\}_{i=1}^{w_\tau}$ where $(l_i, r_i) = (y_{k_i,j_i-1}, y_{k_i,j_i})$ for some $j_i \leq k_i$ such that the three conditions in Definition 3.2 hold for these (l_i, r_i)'s by letting $r = u_c$. Denote $\mathcal{S}^* = \mathcal{S} \setminus \mathcal{W}$, then Lemma 3.1 yields $\nu(\mathcal{S} \setminus \mathcal{S}^*) = 0$. For any $t < r$ with $t \in \mathcal{S}^*$, due to the aforementioned $\{(l_i, r_i] \times W_i : i = 1, 2, \cdots, w_\tau\}$, there exists $(l_i^0, r_i^0] \times W_i$ such that $l_i^0 < t \leq r_i^0$ for each $i = 1, \cdots, w_\tau$. In fact, since $t < r \leq r_i$ for $i = 1, \cdots, w_\tau$, based on the model, we have

$$(l_i^0, r_i^0) = \begin{cases} (l_i, r_i) = (y_{k_i,j_i-1}, y_{k_i,j_i}) \text{ (which is given above)} & \text{if } t \in (l_i, r_i] \\ (y_{k_i,j_i-s-1}, y_{k_i,j_i-s}),\ 1 \leq s < j_i \text{ with } t \in (y_{k_i,j_i-s-1}, y_{k_i,j_i-s}] & \text{otherwise.} \end{cases}$$

Since $t \in \mathcal{S}^*$, by the definition of $\mathcal{S}^*$, $Y_{k,i}$ is defined at t for some (k, i). Then for this pair of (k, i), $P(\Delta = h | Y_{k,i} = t)$ is well defined for each h. Thus $P(\Delta = h | R = t)$ is well defined for each h and $P(\Delta = h | R = t) > 0$ for some h. Then there exists $(l^*, t^*] \times W_{s,h}$ for $t^* = t$ and some $l^* < t^*$ such that $P(\Delta = h | (L, R) = (l^*, t^*)) > 0$ and $W_{s,h} \in P_h$, where l^* could be $-\infty$.

If we set $w_t = w_\tau + 1$, $l_{w_t}^0 = l^*$, $r_{w_t}^0 = t^*$ and $W_{w_t} = W_{s,h}$, then we can verify $(l_i^0, r_i^0] \times W_i$, $i = 1, 2, \cdots, w_t$, satisfy the three conditions in Definition 3.2 as follows.

(1) Obviously, $l_i^0, r_i^0 \in \mathcal{S} \cup \{-\infty\}$ and $l_i^0 < t \leq r_i^0$ for $i = 1, 2, \cdots, w_t$, then we have $\max_i l_i^0 < t = \min_i r_i^0$.

(2) For each $i = 1, 2, \cdots, w_t$, $\exists\ h_i$ such that $P(\Delta = h_i | (L, R) = (l_i^0, r_i^0)) > 0$ by the given assumptions, then it implies that $P(\mathcal{M} = W_i | (L, R) = (l_i^0, r_i^0)) > 0$.

(3) Moreover, $\phi(\{c\}) = \sum_{i=1}^{w_t} g_i^0 \phi(W_i) = \sum_{i=1}^{w_\tau} g_i^0 \phi(W_i)$ by setting $g_{w_t}^0 = 0$.

Thus F is identifiable at (t, c) for almost all $t \in \mathcal{S}$ and $t < u_c$. $\square$

Proof of Remark 3.1. An obvious result from Lemma 3.2 is that if F is also identifiable at (τ_c, c), then $\{(l_i, r_i] \times W_i : i = 1, 2, \cdots, w_\tau\}$ satisfy the three conditions in Definition 3.2

and thus $\phi(\{c\}) = \sum_{i=1}^{w_\tau} g_i \phi(W_i)$ for some constant g_i's. Thus the latter linear combination can be applied to almost all $t \in (-\infty, \tau_c] \cap \mathcal{S}$.

Otherwise, consider a sequence in $\mathcal{S}^*$, say $y_1 < y_2 < \cdots < \tau_c$ such that $\lim_{m\to\infty} y_m = \tau_c$. By Lemma 3.2 and the definition of τ_c, F is identifiable at (y_m, c), so $\exists$ $\{(l_{m,i}, r_{m,i}] \times W_{m,i} : i = 1, 2, \cdots, w_m\}$ satisfying the three conditions in Definition 3.2 for each y_m, define $S_{y_m} = \{W_{m,1}, W_{m,2}, \cdots, W_{m,w_{y_m}}\}$ be the set of distinct values of $W_{m,i}$, $i = 1, 2, \cdots, w_m$, thus $S_{y_m} \subset \mathcal{J}$ and $\phi(\{c\}) = \sum_{i=1}^{w_{y_m}} g_{m,i} \phi(W_{m,i})$ for some constant $g_{m,i}$'s with $w_{y_m} \le J$ based on the random partition masking model and assumption $A2$ for each m. Since $\mathcal{J}$ is finite, the possibilities of S_{y_m} are also finite, say there are N_S distinct S_{y_m}'s, denoted by $S_1, S_2, \cdots, S_{N_S}$. If we define a map $H : \{y_m : m = 1, 2, \cdots\} \to \{S_i : i = 1, 2, \cdots, N_S\}$, then $H^{-1}(S_i) \subset \{y_m : m = 1, 2, \cdots\}$ for each $i = 1, 2, \cdots, N_S$, where H^{-1} is defined in an obvious way and $\bigcup_{i=1}^{N_S} H^{-1}(S_i) = \{y_m : m = 1, 2, \cdots\}$. Since $y_m \to \tau_c$, at least $\exists$ a subsequence $H^{-1}(S_j) = \{y_{m_1} : m_1 = 1, 2, \cdots\} \subset \{y_m : m = 1, 2, \cdots\}$ for some $1 \le j \le N_S$ such that $y_{m_1} \to \tau_c$ as $m_1 \to \infty$. Since $S_{m_1} = S_j = \{W_1, W_2, \cdots, W_w\}, w \le J$ for each m_1 and $\phi(\{c\}) = \sum_{i=1}^{w} g_i \phi(W_i)$ for some constant g_i's, then by Lemma 3.2 and the definition of τ_c, this linear combination can be applied to almost all $t \in (-\infty, \tau_c) \cap \mathcal{S}$. $\square$