

On the Convergence of Observed Partial Likelihood under Incomplete Data with Two Class Possibilities

Tomoyuki Sugimoto

Department of Mathematical Sciences, Hirosaki University, Hirosaki, Japan
Email: tomoyuki@cc.hirosaki-u.ac.jp

Received October 21, 2013; revised November 21, 2013; accepted November 28, 2013

Copyright © 2014 Tomoyuki Sugimoto. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In accordance of the Creative Commons Attribution License all Copyrights © 2014 are reserved for SCIRP and the owner of the intellectual property Tomoyuki Sugimoto. All Copyright © 2014 are guarded by law and by SCIRP as a guardian.

ABSTRACT

In this paper, we discuss the theoretical validity of the observed partial likelihood (OPL) constructed in a Cox-type model under incomplete data with two class possibilities, such as missing binary covariates, a cure-mixture model or doubly censored data. A main result is establishing the asymptotic convergence of the OPL. To reach this result, as it is difficult to apply some standard tools in the survival analysis, we develop tools for weak convergence based on partial-sum processes. The result of the asymptotic convergence shown here indicates that a suitable order of the number of Monte Carlo trials is less than the square of the sample size. In addition, using numerical examples, we investigate how the asymptotic properties discussed here behave in a finite sample.

KEYWORDS

Cox's Regression Model; Logistic Regression Model; Incomplete Binary Data; Partial Likelihood; Partial-Sum Processes; Profile Likelihood

1. Introduction

Although the Cox model [1] is a standard tool for the analysis of time-to-event data, in practice analysts are often confronted with some problems in handling incomplete data beyond the right-censored form, such as interval-censored data [2], missing covariates [3-5] or (statistical) structural modelling. The inference for the Cox model under such cases of incomplete data can usually be performed based on the semiparametric profile likelihood (SPFL) [6] as a generalization of Cox's partial likelihood [7]. On the other hand, as a substitute for the SPFL method, one can analyse the same data using the imputation method, which yields a sum of partial likelihoods. By describing the sum of all possible partial likelihoods more exactly, we can formulate the marginal of partial likelihoods [8], that is, the observed partial likelihood (OPL). In this area, the theory for the SPFL has been studied by many authors (e.g., [6,9]). However, to the best of our knowledge, there has not been much development in mathematical theory for the OPL.

In this paper, we discuss the theoretical validity of the OPL which appears in a Cox-type model under incomplete data beyond the right-censored form. In this area, one advantage of the OPL is that the baseline hazard function as a nuisance included in a Cox-type model is eliminated completely in the inferential likelihood. This yields a more stable computational system for optimization than that of the SPFL. For example, in a Cox-cure model, a computational process based the EM algorithm to obtain the SPFL easily fails to converge if a suitable starting value is not provided (e.g., see [10]). The main disadvantages of the OPL are, for instance, that a great length of time is required for the exact computation and it is not clear how much the amount of computation can be reduced by the Monte Carlo (MC) method. However, even if the feasible number of MC trials is smaller than desirable to approximate the OPL, and hence the MC approximation is quite rough, it may be sufficient for a starting value in the computational process of the SPFL.

Generally, it is difficult to investigate computationally to what extent the MC approximations of the OPL are valid, since the exact computation requires a huge number of summands, as the sample size and incomplete information of data are increasing. For this reason, it is worth studying the OPL theoretically. However, it is not easy to complete such a study in one go, because standard tools to study asymptotic properties of Cox's partial likelihood or the SPFL cannot be applied directly to an objective of the OPL. Therefore in this paper, for the sake of simplicity, we focus on the OPL constructed in incomplete data composed of unobserved two class labels. Typical cases of this type occur in a Cox-type model with incomplete data, such as missing binary covariates, a cure-mixture model or doubly censored data. As a main result, we establish the asymptotic convergence of the OPL and derive a limit form of the OPL. This result is a foundation or precondition for applying an infinite-dimensional Laplace approximation for integral on the baseline hazard. Such a Laplace approximation method will yield the other limit form of the OPL [11], which is useful in discussing the consistency and asymptotic normality of the estimators. However, the method is not convenient for showing the convergence of the OPL. For these reasons, it is also valuable to discuss the convergence of the OPL using the arguments employed in this paper.

A matter of interest in practice concerns MC approximations of the OPL. One other significant point is that the result for the convergence of the exact OPL can be easily tailored to the context of the MC approximations. Based on such an argument, we show that a suitable order of the number of MC trials is less than $O(n^2)$. Further, in Section 4 we investigate how the asymptotic properties discussed here behave in a finite sample.

In Section 2 we formulate the OPL in incomplete data with two class possibilities, providing several examples of interest; in Section 3 we develop the tools to obtain the main result and show the convergence of the OPL, and in Section 4 we discuss the performances of MC approximations.

2. Observed Full and Partial Likelihoods

2.1. Notations and Motivated Examples

Let $T_i = \min(T_i^*, U_i)$ and $\Delta_i = 1(T_i^* \leq U_i)$ be the observed survival time and right-censoring indicator of the i -th individual, where $T_i^*, i = 1, \dots, n$ are continuous random variables independent of U_i and $1(\cdot)$ is the indicator function. Suppose that the individuals possess some difference between models or observations identified by the two classes. We define such a class variable by

$$q_i^* = \begin{cases} 1 & \text{if the } i\text{-th individual belongs to class 1} \\ 0 & \text{if the } i\text{-th individual belongs to class 0} \end{cases}$$

In the case that q_i^* expresses the difference between models, assume that the distribution of T_i^* follows the proportional hazards model formulated as

$$\lambda_i(t) = \lambda_0(t)r_i^{(j)}(\beta) \quad \text{and} \quad r_i^{(j)}(\beta) = r^{(j)}(\beta^T Z_i^{(j)}), \quad j = 0, 1,$$

where $\lambda_0(t)$ is the baseline hazard function, $r^{(j)}(x)$ is the function given by

$$r^{(1)}(x) = \exp(x), \quad \text{either } r^{(0)}(x) = \exp(x) \text{ or } r^{(0)}(x) = 0,$$

$Z_i^{(j)}$ is the covariate vector $(Z_{i1}^{(j)}, \dots, Z_{ib}^{(j)})^T$ from the population of the class j , and β is the regression coefficient vector $(\beta_1, \dots, \beta_b)^T$. As usual, the information on (T_i, Δ_i) can be re-expressed using the counting processes $\mathcal{N}_i(t) = 1(T_i \leq t, \Delta_i = 1)$ and at-risk processes $\mathcal{Y}_i(t) = 1(T_i \geq t)$.

In this paper, we consider incomplete data where some of the q_i^* 's are treated as missing. Let

$$C_i = 1(q_i^* \text{ is completely observed}) \quad \text{and}$$

$$\pi_i^{(j)} = \begin{cases} 1(q_i^* = j) & \text{for } j = 0, 1 \text{ if } C_i = 1 (q_i^* \text{ is complete}), \\ 1 & \text{for } j = 0, 1 \text{ if } C_i = 0 (q_i^* \text{ is missing}). \end{cases}$$

Each of these is used to construct the likelihoods. Further, if the event of $\{q_i^* = j\}$ can be expressed by a probability, we use the following notation and assumption

$$\Pr(q_i^* = j | X_i) = p_i^{(j)}(\alpha) \quad \text{and} \quad p_i^{(0)}(\alpha) = \exp(\alpha^T X_i) / \{1 + \exp(\alpha^T X_i)\},$$

where X_i is the covariate vector $(1, X_{i1}, \dots, X_{ia})^T$ related to q_i^* , and α is the regression coefficient vector $(\alpha_0, \alpha_1, \dots, \alpha_a)^T$. For simplicity, we will write $\theta = (\theta_0, \theta_1, \dots, \theta_{a+b})^T = (\alpha^T, \beta^T)^T$ hereafter.

Let $\mathbf{q}^* = (q_1^*, \dots, q_n^*)$ be the collection of true q_i^* 's. In many cases of incomplete data with two class possibilities, the observed full likelihood (OFL) can be generally written as

$$L_f^{(n)}(\theta, \Lambda_0) = \sum_{\mathbf{q} \in \{0,1\}^n} L_{\pi_f}^{(n)}(\theta, \Lambda_0; \mathbf{q}) \mathbb{I}_s(\mathbf{q}) \tag{2.1}$$

with the elements such that

$$L_{\pi_f}^{(n)}(\theta, \Lambda_0; \mathbf{q}) = \prod_{i=1}^n p_i^{(q_i)}(\alpha) \left\{ \lambda_0(T_i) r_i^{(q_i)}(\beta) \right\}^{\Delta_i} S_i^{(q_i)}(T_i; \beta, \Lambda_0) \pi_i^{(q_i)}$$

and $\mathbb{I}_s(\mathbf{q}) = \prod_{i=1}^n S^{q_i}$, where the space $\{0,1\}^n$ denotes the collection of all the vectors composed of 0 or 1 with length n , $\mathbf{q} = (q_1, \dots, q_n)$ expresses one element of $\{0,1\}^n$, in which there exists one true element \mathbf{q}^* ,

$$S_i^{(j)}(t; \beta, \Lambda_0) = \exp(-r_i^{(j)}(\beta) \Lambda_0(t)), \quad j = 0, 1$$

is the survival function of the i -th individual belonging to the class j , $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$ is the cumulative baseline hazard function, and S is given by either of $s=1$ or -1 in advance.

In the following three examples, we show how the form of the OFL is related to the representative cases. Hereafter, we will often omit θ when it is clear that a function depends on θ , e.g. $r_i^{(j)} = r_i^{(j)}(\beta)$, $p_i^{(j)} = p_i^{(j)}(\alpha)$, $S_i^{(j)}(t; \Lambda_0) = S_i^{(j)}(t; \beta, \Lambda_0)$ and so on.

Example: Missing Binary Covariates. Let us assume that $r^{(0)}(x) = r^{(1)}(x)$ but $Z_i^{(0)} \neq Z_i^{(1)}$. For example, the first covariate is binary and may be missing, $Z_{i1}^{(q_i^*)} = q_i^*$ and $Z_{ij}^{(q_i^*)} = Z_{ij}$, $j \geq 2$. Then, we can write

$$r_i^{(q_i^*)}(\beta) = \exp(\beta^T Z_i^{(q_i^*)}) = q_i^* \exp(\beta^T Z_i^{(1)}) + (1 - q_i^*) \exp(\beta^T Z_i^{(0)}),$$

where $\beta^T Z_i^{(q_i^*)} = \beta_1 q_i^* + \sum_{j=2}^b \beta_j Z_{ij}$. In this case, the OFL is

$$L_f^{(n)}(\theta, \Lambda_0) = \prod_{i=1}^n \left[\sum_{j=0,1} p_i^{(j)}(\alpha) \left\{ \lambda_0(T_i) r_i^{(j)}(\beta) \right\}^{\Delta_i} S_i^{(j)}(T_i; \beta, \Lambda_0) \pi_i^{(j)} \right].$$

Using the binomial expansion, this can be rewritten as

$$L_f^{(n)}(\theta, \Lambda_0) = \sum_{\mathbf{q} \in \{0,1\}^n} \prod_{i=1}^n p_i^{(q_i)} \left\{ \lambda_0(T_i) r_i^{(q_i)}(\beta) \right\}^{\Delta_i} S_i^{(q_i)}(T_i; \Lambda_0) \pi_i^{(q_i)}. \tag{2.2}$$

Example: Cox Cure-Mixture Model. The Cox cure-mixture model [10,12-15] is presumed to hold the proportional hazards model for uncured individuals and to be zero-hazard for cured ones. That is, we assume $Z_i^{(0)} = Z_i^{(1)}$ but $r^{(0)}(x) = 0$, so that we can write

$$r_i^{(q_i^*)}(\beta) = q_i^* \exp(\beta^T Z_i^{(1)}).$$

We observe that $q_i^* = 1$ (i.e. $C_i = 1$) if $\Delta_i = 1$ and q_i^* is missing (i.e. $C_i = 0$) otherwise. The OFL is usually

$$L_f^{(n)}(\theta, \Lambda_0) = \prod_{i=1}^n \left\{ p_i^{(1)} \lambda_0(T_i) r_i^{(1)} S_i^{(1)}(T_i; \Lambda_0) \right\}^{\Delta_i} \left\{ p_i^{(0)} + p_i^{(1)} S_i^{(1)}(T_i; \Lambda_0) \right\}^{1-\Delta_i}; \tag{2.3}$$

note here that $S_i^{(0)}(t; \Lambda_0) = 1$ for all t . Then, (2.3) can be rewritten in the same form as (2.2).

Example: Doubly Censored Data. In doubly censored data [16,17], left-censored data may be included. Let \tilde{q}_i^* indicate whether the i -th observation is left-censored or not, the OFL is then

$$L_f^{(n)}(\theta, \Lambda_0) = \prod_{i=1}^n \left\{ \lambda_0(T_i) r_i^{(1)} \right\}^{\tilde{q}_i^* \Delta_i} S_i^{(1)}(T_i; \Lambda_0)^{\tilde{q}_i^*} \left\{ 1 - S_i^{(1)}(T_i; \Lambda_0) \right\}^{1-\tilde{q}_i^*}. \tag{2.4}$$

In the phenomenal meaning, the common model is assumed regardless of the type of observations, but we do not define q_i^* as \tilde{q}_i^* . Here we use the rule $r_i^{(0)} = 0$ and $Z_i^{(1)} = Z_i^{(0)}$ such that q_i^* designates the type of model rather than an observation. Under this rule, because we have $S_i^{(0)}(t; \Lambda_0) = 1$ for all t , (2.4) can be expressed as

$$L_f^{(n)}(\theta, \Lambda_0) = \sum_{\mathbf{q} \in \{0,1\}^n} \prod_{i=1}^n \left\{ \lambda_0(T_i) r_i^{(q_i)} \right\}^{\Delta_i} S_i^{(q_i)}(T_i; \Lambda_0) \pi_i^{(q_i)} (-1)^{q_i},$$

where note that q_i^* is defined as missing data (*i.e.* $C_i = 0$) if the i -th observation is left-censored, and as complete data of $q_i^* = 1$ (*i.e.* $C_i = 1$) otherwise.

2.2. Observed Partial Likelihood

Let $\mathbf{R}_n[\cdot]$ be the integral operator proposed by [18] to derive the partial likelihood in the Cox model without time-dependent covariates. Without loss of generality, let us suppose that there are no ties. By operating $\mathbf{R}_n[\cdot]$ to the OFL $L_f^{(n)}(\theta, \Lambda_0)$, we have the OPL

$$L_p^{(n)}(\theta) = \sum_{\mathbf{q} \in \{0,1\}^n} L_{\pi p}^{(n)}(\theta; \mathbf{q}) \mathbb{I}_s(\mathbf{q}),$$

where

$$L_{\pi p}^{(n)}(\theta; \mathbf{q}) = \mathbf{R}_n \left[L_{\pi f}^{(n)}(\theta, \Lambda_0; \mathbf{q}) \right] = \prod_{i=1}^n p_i^{(q_i)}(\alpha) \prod_{i=1}^n \left\{ r_i^{(q_i)}(\beta) / \sum_{j=1}^n \mathcal{Y}_j(T_i) r_j^{(q_j)}(\beta) \right\}^{\Delta_i} \pi_i^{(q_i)}.$$

Let $\ell_{p(n)}(\theta) = n^{-1} \log(L_p^{(n)}(\theta)) + n^{-1} \log(\prod_{i=1}^n n^{\Delta_i})$. To discuss an asymptotic form of $\ell_{p(n)}(\theta)$, we will prepare some convenient expressions. First, to pack the expression of $\pi_i^{(q_i)}$ into $r_i^{(q_i)}$ and $p_i^{(q_i)}$, we define

$$\begin{aligned} \tilde{Z}_i^{(q_i)} \tilde{r}_i^{(q_i)}(\beta)^h &= C_i \left[q_i^* \left\{ Z_i^{(1)} \right\}^l \left\{ r_i^{(1)}(\beta) \right\}^h + (1 - q_i^*) \left\{ Z_i^{(0)} \right\}^l \left\{ r_i^{(0)}(\beta) \right\}^h \right] \\ &\quad + (1 - C_i) \left[q_i \left\{ Z_i^{(1)} \right\}^l \left\{ r_i^{(1)}(\beta) \right\}^h + (1 - q_i) \left\{ Z_i^{(0)} \right\}^l \left\{ r_i^{(0)}(\beta) \right\}^h \right], \quad h, l = 0, 1, \end{aligned}$$

$$\text{and} \quad \tilde{X}_i^{(q_i)} = C_i (1 - q_i^*) X_i + (1 - C_i) (1 - q_i) X_i.$$

Using these expressions, let

$$\begin{aligned} \mathbb{S}_n^{(0)}(t, \mathbf{q}; \beta) &= \mathbb{E}_n \left[\mathcal{Y}_i(t) \tilde{r}_i^{(q_i)}(\beta) \right], \quad \mathbb{S}_n^{(1)}(t, \mathbf{q}; \beta) = \mathbb{E}_n \left[\mathcal{Y}_i(t) \tilde{Z}_i^{(q_i)} \tilde{r}_i^{(q_i)}(\beta) \right], \\ \text{and} \quad \mathbb{S}_n^{(0)}(\mathbf{q}; \alpha) &= \mathbb{E}_n \left[\alpha^T \tilde{X}_i^{(q_i)} - \log(1 + e^{\alpha^T X_i}) \right], \end{aligned}$$

where $\mathbb{E}_n[R_i] = n^{-1} \sum_{i=1}^n R_i$ is an empirical version of the theoretical expectation $\mathbb{E}[R_i]$. Further, as an important definition, let

$$\nu_n(\mathcal{A}) = 2^{-n} \sum_{x \in \{0,1\}^n} \mathbb{1}(x \in \mathcal{A})$$

be the n -dimensional version of Minkowski's measure ν_∞ . Then, using these notations, $\ell_{p(n)}(\theta)$ can be written as

$$\ell_{p(n)}(\theta) = n^{-1} \log \left\{ B_n \int_{\mathbf{q} \in \{0,1\}^n} \exp(n \ell_{\pi p(n)}(\theta; \mathbf{q})) \mathbb{I}_s(\mathbf{q}) d\nu_n(\mathbf{q}) \right\}, \quad (2.5)$$

where

$$\ell_{\pi p(n)}(\theta; \mathbf{q}) = \mathbb{S}_n^{(0)}(\mathbf{q}; \alpha) + \int_0^{\tau_e} \beta^T \mathbb{E}_n \left[\tilde{Z}_i^{(q_i)} d\mathcal{N}_i(t) \right] - \int_0^{\tau_e} \log \left\{ \mathbb{S}_n^{(0)}(t, \mathbf{q}; \beta) \right\} \mathbb{E}_n \left[d\mathcal{N}_i(t) \right],$$

τ_e is the greatest follow-up time and $B_n = 2^n / 2^{\sum_{i=1}^n C_i}$. The quantity of $2^{\sum_{i=1}^n C_i}$ is the total number that the same $\ell_{\pi p(n)}(\theta; \mathbf{q})$'s are repeated on $\mathbf{q} \in \{0,1\}^n$. In addition, we define

$$\bar{\ell}_{p(n)}(\theta) = \ell_{p(n)}(\theta) \Big|_{\ell_{\pi p(n)}(\theta; \mathbf{q}) = \bar{\ell}_{\pi p(n)}(\theta; \mathbf{q})}$$

$$\text{and} \quad \bar{\ell}_{\pi p(n)}(\theta; \mathbf{q}) = \mathbb{S}_n^{(0)}(\mathbf{q}; \alpha) + \int_0^{\tau_e} \left\{ \beta^T \bar{\mathbb{S}}_n^{(0)}(t, \mathbf{q}; \beta^*) - \log \left\{ \mathbb{S}_n^{(0)}(t, \mathbf{q}; \beta) \right\} \mathbb{S}_n^{(0)}(t, \mathbf{q}^*; \beta^*) \right\} d\Lambda_0^*(t),$$

which is $\ell_{\pi p(n)}(\theta; \mathbf{q})$ in which $\mathcal{N}_i(t)$'s are replaced by the true intensity

$$\Lambda_i^{**}(t) = q_i^* \int_0^t \mathcal{Y}_i(s) r_i^{(1)}(\beta^*) d\Lambda_0^*(s) + (1 - q_i^*) \int_0^t \mathcal{Y}_i(s) r_i^{(0)}(\beta^*) d\Lambda_0^*(s),$$

where $\bar{\mathbb{S}}_n^{(1)}(t, \mathbf{q}; \beta^*) = \mathbb{E}_n[\mathcal{Y}_i(t) \tilde{Z}_i^{(q_i)} r_i^{(q_i)}(\beta^*)]$, β^* and Λ_0^* are the true β and Λ_0 . In the case of $Z_i^{(1)} = Z_i^{(0)}$, such as Cox cure-mixture model or doubly censored data, we have $\bar{\mathbb{S}}_n^{(1)}(t, \mathbf{q}; \beta^*) = \mathbb{S}_n^{(1)}(t, \mathbf{q}^*; \beta^*)$.

Remark 1. In (2.5), even if we consider a difference or quotient between $\ell_{p^{(n)}}(\theta)$ and $\bar{\ell}_{p^{(n)}}(\theta)$, we cannot remove the potential increasing factor n in the summands $\exp(n\ell_{\pi p^{(n)}}(\cdot))$ because of the existence of the operator $\sum_{q \in \{0,1\}^n}$. Thus, it may potentially be difficult to apply some of the standard tools in the survival analysis to the asymptotic discussion. For these reasons, our strategy to obtain a limit of $\ell_{p^{(n)}}(\theta)$ is to regard all the summands of $\sum_{q \in \{0,1\}^n}$ as a process on $\{0,1\}^n$. We will then derive the result of a weak convergence on $\{0,1\}^n$. \square

3. Convergence of the Observed Partial Likelihood

We will now discuss how the mean of the log OPL converges to a deterministic function and provide Theorem 1 of the main result. The following conditions are assumed for these discussions.

Conditions A. Let Θ be a compact set of θ which includes $\theta^* = (\alpha^{*\top}, \beta^{*\top})^\top$, where θ^* and α^* are the true θ and α . The true baseline function $\Lambda_0^*(t)$ is continuous and non-decreasing on $t \in [0, \tau_e]$ with $\Lambda_0^*(0) = 0$.

A1: $X_i, Z_i^{(j)}, i = 1, \dots, n$ are i.i.d. vectors from the population of the class $j = 0, 1$.

A2: $\Pr(\mathcal{Y}_i(\tau_e) = 1) > 0$. A3: $\Lambda_0^*(\tau_e) < \infty$.

A4: $E\left[\sup_{\beta \in \Theta} \left|Z_{il}^{(j)}\right| r_i^{(j)}(\beta)\right] < \infty, j = 0, 1$ for $l = 1, \dots, b$. A5: $E\left[|X_{il}|\right] < \infty$ for $l = 1, \dots, a$.

Condition A2 means $E[\{p_i^{*(0)} S_i^{*(0)}(t) + p_i^{*(1)} S_i^{*(1)}(t)\} S_i^U(t)] > 0$ for all $t \in [0, \tau_e]$ because of $\Pr(\mathcal{Y}_i(t) = 1) = E[\mathcal{Y}_i(t)]$, where $S_i^{*(j)}(t) = S_i^{(j)}(t; \beta^*, \Lambda_0^*)$ and $p_i^{*(j)} = p_i^{(j)}(\alpha^*)$ and $S_i^U(t)$ is the i -th survival function of right-censoring time under a given $(X_i, Z_i^{(q_i)})$. By the compact condition of Θ , we have $0 < E[p_i^{(0)}(\alpha)] < 1$ on $\alpha \in \Theta$ as a matter of course. However, in the case that there are no $p_i^{(j)}(\alpha)$ as in the example of doubly censored data, such conditions on α are omitted because $S_n^{(0)}(\mathbf{q}; \alpha) = 0$ always.

Theorem 1. Suppose that Conditions A are satisfied. Then, as $n \rightarrow \infty$, $\ell_{p^{(n)}}(\theta)$ converges almost surely to a deterministic function uniformly on $\theta \in \Theta$.

Theorem 1 is proved in Section 3.3. We prepare useful tools for such a proof in Sections 3.1 and 3.2 below. In Section 3.1, we discuss a relation needed to show that two OPL's converge to the same limit, determining a plan (Lemmas 1 and 2) to obtain Theorem 1. In Section 3.2, following the plan, we provide a tool (Lemma 3) to give a weak convergence of all possible partial-sum processes.

3.1. Relations between Two Observed Partial Likelihoods

Note that the OPL is constructed by an integral on $\{0,1\}^n$ with the measure ν_n . Thus, to give two OPL's with the same limit, it is predicted that the difference between the integrands of two OPL's should converge weakly to zeros on $\{0,1\}^n$, for example, by analogy of the dominated convergence theorem. Let

$$\ell_{p^{(\aleph)}}^{\iota}(\theta) = n^{-1} \log \left\{ B_n \int_{\mathbf{q} \in \{0,1\}^n} \varphi_n \left(\ell_{\pi p^{(\aleph)}}^{\iota}(\theta; \mathbf{q}) \right) \mathbb{I}_{\mathfrak{s}}(\mathbf{q}) d\nu_n(\mathbf{q}) \right\} \text{ for } \iota = A, B \text{ and } \aleph = n, \star$$

be functions that exist around $\ell_{p^{(n)}}(\theta)$, where $\varphi_n(x) = \exp(nx)$ for simplicity. Then, we have the following lemma about some $\ell_{\pi p^{(\aleph)}}^A(\theta; \mathbf{q})$ and $\ell_{\pi p^{(\aleph)}}^B(\theta; \mathbf{q})$.

Lemma 1. Suppose that

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} \nu_n \left(\mathbf{q} \in \{0,1\}^n : \left| \ell_{\pi p^{(\aleph)}}^A(\theta; \mathbf{q}) - \ell_{\pi p^{(\aleph)}}^B(\theta; \mathbf{q}) \right| > \varepsilon \right) = 0 \text{ for all } \varepsilon > 0, \tag{3.1}$$

with probability 1. Then, as $n \rightarrow \infty$,

$$\sup_{\theta \in \Theta} \left| \ell_{p^{(\aleph)}}^A(\theta) - \ell_{p^{(\aleph)}}^B(\theta) \right| \rightarrow_{\text{as}} 0,$$

where \rightarrow_{as} denotes almost sure convergence.

(Proof of Lemma 1). Using Taylor expansions of

$$\log f_1 - \log f_0 = \tilde{f}_0^{-1}(f_1 - f_0), \quad g_1^{1/n} - g_0^{1/n} = n^{-1} \{ \tilde{g}_0 \}^{1/n-1} (g_1 - g_0) \text{ and } e^{nh_1} - e^{nh_0} = ne^{nh_0} (h_1 - h_0),$$

the difference between $\ell_{p(\mathbb{S})}^A(\theta)$ and $\ell_{p(\mathbb{S})}^B(\theta)$ is derived as

$$\ell_{p(\mathbb{S})}^A(\theta) - \ell_{p(\mathbb{S})}^B(\theta) = \frac{\{\tilde{\ell}_2\}^{1/n}}{\tilde{\ell}_1 \tilde{\ell}_2} \int_{\mathbf{q} \in \{0,1\}^n} \tilde{\ell}_{3(\mathbf{q})} \left\{ \ell_{\pi p(\mathbb{S})}^A(\theta; \mathbf{q}) - \ell_{\pi p(\mathbb{S})}^B(\theta; \mathbf{q}) \right\} \mathbb{I}_{\mathbb{S}}(\mathbf{q}) d\nu_n(\mathbf{q})$$

where, with some $\xi_1, \xi_2, \xi_{3(\mathbf{q})} \in (0,1)$ on $\mathbf{q} \in \{0,1\}^n$ and $\xi'_l = 1 - \xi_l$ ($l=1,2,3(\mathbf{q})$),

$$\begin{aligned} \tilde{\ell}_1 &= \xi_1 \exp\left(\ell_{p(\mathbb{S})}^A(\theta)\right) + \xi'_1 \exp\left(\ell_{p(\mathbb{S})}^B(\theta)\right), \tilde{\ell}_2 = \xi_2 \varphi_n\left(\ell_{p(\mathbb{S})}^A(\theta)\right) + \xi'_2 \varphi_n\left(\ell_{p(\mathbb{S})}^B(\theta)\right) \\ &\text{and } \tilde{\ell}_{3(\mathbf{q})} = \varphi_n\left(\xi_{3(\mathbf{q})} \ell_{\pi p(\mathbb{S})}^A(\theta; \mathbf{q}) + \xi'_{3(\mathbf{q})} \ell_{\pi p(\mathbb{S})}^B(\theta; \mathbf{q})\right). \end{aligned}$$

Let us assume that $0 < \ell_{p(\mathbb{S})}^B(\theta) < \ell_{p(\mathbb{S})}^A(\theta)$ without loss of generality. Then, $\{\tilde{\ell}_2\}^{1/n}/\tilde{\ell}_1$ and $\int_{\tilde{\ell}_{3(\mathbf{q})}} d\nu_n(\mathbf{q})/\tilde{\ell}_2$ are bounded on $\theta \in \Theta$ by some finite values b_1 and b_2 . In fact,

$$\{\tilde{\ell}_2\}^{1/n}/\tilde{\ell}_1 < 1/\xi_1 \quad \text{and} \quad \int_{\tilde{\ell}_{3(\mathbf{q})}} d\nu_n(\mathbf{q})/\tilde{\ell}_2 < 1/\xi_2$$

are shown by $\exp\left(\ell_{p(\mathbb{S})}^B(\theta)\right)/\exp\left(\ell_{p(\mathbb{S})}^A(\theta)\right) < 1$. Therefore, we have

$$\begin{aligned} \sup_{\theta} \left| \ell_{p(\mathbb{S})}^A(\theta) - \ell_{p(\mathbb{S})}^B(\theta) \right| &\leq b_1 \sup_{\theta} \left| \int_{\tilde{\ell}_{3(\mathbf{q})}} \frac{\tilde{\ell}_{3(\mathbf{q})}}{\tilde{\ell}_2} \left\{ \ell_{\pi p(\mathbb{S})}^A(\theta; \mathbf{q}) - \ell_{\pi p(\mathbb{S})}^B(\theta; \mathbf{q}) \right\} \mathbb{I}_{\mathbb{S}}(\mathbf{q}) d\nu_n(\mathbf{q}) \right| \\ &\leq b_1 b_2 \sup_{\theta} \nu_n\left(\mathbf{q} \in \{0,1\}^n : \left| \ell_{\pi p(\mathbb{S})}^A(\theta; \mathbf{q}) - \ell_{\pi p(\mathbb{S})}^B(\theta; \mathbf{q}) \right| > 0\right). \end{aligned}$$

Applying (3.1) to the above inequality, this lemma is proved. \square

Using Lemma 1, for several patterns of $\ell_{\pi p(\mathbb{S})}^A$ and $\ell_{\pi p(\mathbb{S})}^B$ we can investigate whether they converge to the same limit. The important problem is how to show the condition (3.1). For this purpose, we make the use of meaning that a convergence in ν_n -probability implicated in (3.1), since ν_n is a probability measure on $\{0,1\}^n$. We have the following lemma to establish the condition (3.1).

Lemma 2. Suppose that

$$\sup_{\theta \in \Theta, \mathbf{q} \in \{0,1\}^n} \left| \ell_{\pi p(\mathbb{S})}^A(\theta; \mathbf{q}) - \ell_{\pi p(\mathbb{S})}^B(\theta; \mathbf{q}) \right| \rightarrow_p 0, \quad (3.2)$$

where \rightarrow_p denotes convergence in probability. Then, (3.1) is established.

Remark 2. For simplicity, letting

$$G_n(\theta, \mathbf{q}) = \left| \ell_{\pi p(\mathbb{S})}^A(\theta; \mathbf{q}) - \ell_{\pi p(\mathbb{S})}^B(\theta; \mathbf{q}) \right| \quad \text{and} \quad f_n^\varepsilon(\mathbf{q}) = 1(\sup_{\theta \in \Theta} G_n(\theta, \mathbf{q}) > \varepsilon),$$

denote

$$Q_n^\varepsilon = \nu_n\left(\mathbf{q} \in \{0,1\}^n : f_n^\varepsilon(\mathbf{q}) = 1\right) = \int_{\mathbf{q} \in \{0,1\}^n} f_n^\varepsilon(\mathbf{q}) d\nu_n(\mathbf{q})$$

as the area of $f_n^\varepsilon(\mathbf{q}) = 1$. We can immediately show that (3.2) provides a version of convergence in probability of (3.1), i.e.

$$\lim_{n \rightarrow \infty} \Pr(Q_n^\varepsilon = 0) = 1 \quad \text{for all } \varepsilon > 0, \quad (3.3)$$

because it is always satisfied that

$$\sup_{\theta} \nu_n\left(\mathbf{q} \in \{0,1\}^n : G_n(\theta, \mathbf{q}) > \varepsilon\right) \leq \nu_n\left(\mathbf{q} \in \{0,1\}^n : \sup_{\theta} G_n(\theta, \mathbf{q}) > \varepsilon\right) \leq \sup_{\theta \in \Theta, \mathbf{q} \in \{0,1\}^n} G_n(\theta, \mathbf{q}).$$

Thus, we show that the operators of \lim and \Pr are mutually exchangeable in (3.3) in a proof of Lemma 2. \square (Proof of Lemma 2). Note that

$$Q_n^\varepsilon = \nu_n\left(\mathbf{q} \in \{0,1\}^n : f_n^\varepsilon(\mathbf{q}) = 1\right) = \int_{\mathbf{q} \in \{0,1\}^n} f_n^\varepsilon(\mathbf{q}) d\nu_n(\mathbf{q}) \leq 1$$

because $f_n^\varepsilon(\mathbf{q})$ is independent of $(q_{n+1}, q_{n+2}, \dots)$ due to the n -dimensional projection to $\{0,1\}^n$ from $\{0,1\}^\infty$. From condition (3.2), limits of $f_n^\varepsilon(\mathbf{q})$ are zeros almost everywhere on $\mathbf{q} \in \{0,1\}^n$, which can be eventually written as $\lim_n f_n^\varepsilon(\mathbf{q}) \leq o_p(1)$ independently of $\mathbf{q} \in \{0,1\}^\infty$. The dominated convergence theorem provides

$$\lim_{n \rightarrow \infty} Q_n^\epsilon = \int_{\mathbf{q} \in \{0,1\}^\infty} \lim_{n \rightarrow \infty} f_n^\epsilon(\mathbf{q}) d\nu_\infty(\mathbf{q}) \leq o_p(1).$$

This shows $\Pr(\lim_n Q_n^\epsilon = 0) = 1$ via Markov’s inequality such that

$$\Pr(\lim_n Q_n^\epsilon \geq \eta) \leq \int_{\mathbf{q}} \mathbb{E}[\lim_n f_n^\epsilon(\mathbf{q})] d\nu_\infty(\mathbf{q}) / \eta^2.$$

Therefore, condition (3.2) gives (3.1). \square

3.2. All Possible Partial-Sum Processes

Note that, by the portmanteau theorem, (3.2) is equivalent to, as $n \rightarrow \infty$,

$$\ell_{np(n)}^A(\theta; \mathbf{q}) - \ell_{np(n)}^B(\theta; \mathbf{q}) \rightarrow_D 0 \text{ on } (\theta, \mathbf{q}) \in \Theta \times \{0,1\}^\infty,$$

where \rightarrow_D denotes convergence in distribution. In this section, we develop a tool to show such a weak convergence on $\mathbf{q} \in \{0,1\}^\infty$.

For simplicity, let $C'_i = 1 - C_i$, $q_i^* = 1 - q_i$ and $q'_i = 1 - q_i$. An important key to obtaining (3.2) or a limit of $\ell_{p(n)}(\theta)$ is a convergence result of $\mathbb{S}_n^{(0)}(t, \mathbf{q}; \beta)$, $\overline{\mathbb{S}}_n^{(1)}(t, \mathbf{q}; \beta^*)$ and $\mathbb{S}_n^{(0)}(\mathbf{q}; \alpha)$ on $\mathbf{q} \in \{0,1\}^n$. As representations of more essential terms to consider in the convergence on $\mathbf{q} \in \{0,1\}^n$, we denote

$$\mathbb{T}_n(Y(t; \theta), \mathbf{q}) = \sum_{i=1}^n q_i Y_i(t; \theta) \text{ and } \mathbb{T}_n(Y(t; \theta), \mathbf{q}') = \sum_{i=1}^n q'_i Y_i(t; \theta),$$

where examples of $Y_i(t; \theta)$ are $C_i \mathcal{Y}_i(t) Z_i^{(1)l} r_i^{(1)}$, $C_i \mathcal{Y}_i(t) Z_i^{(0)h} r_i^{(0)}$, $C_i X_i$ and so on. Then, $n^{-1} \mathbb{T}_n(Y(t; \theta), \mathbf{q})$ can be regarded as $E_n[Y_i(t; \theta) q_i | \mathbf{q}^*]$, which is the empirical version of the conditional expectation on a unique \mathbf{q}^* but is calculated letting \mathbf{q} be fixed. Let

$$\mu_{q_i^*}(Y_i(t; \theta)) = \begin{cases} \mathbb{E}[Y_i(t; \theta) | q_i^* = 0] & \text{if } q_i^* = 0, \\ \mathbb{E}[Y_i(t; \theta) | q_i^* = 1] & \text{if } q_i^* = 1, \end{cases}$$

which is the conditional expectation of $Y_i(t; \theta)$ on given q_i^* . Although q_i^* is treated as unknown in many incomplete problems, this is not the case for terms included in the observed partial likelihood. Hence, for a fixed $\mathbf{q} \in \{0,1\}^n$, the conditional expectation of $\mathbb{T}_n(Y(t; \theta), \mathbf{q})$ on \mathbf{q}^* is $\mathbb{E}[\mathbb{T}_n(Y(t; \theta), \mathbf{q}) | \mathbf{q}^*] = \sum_{i=1}^n q_i \mu_{q_i^*}(Y_i(t; \theta))$. So, letting $s_n^{(l)}(t, \mathbf{q}; \beta)$, $\overline{s}_n^{(1)}(t, \mathbf{q}; \beta)$ and $s_n^{(0)}(\mathbf{q}; \alpha)$ be means to centralize $\mathbb{S}_n^{(l)}(t, \mathbf{q}; \beta)$, $\overline{\mathbb{S}}_n^{(1)}(t, \mathbf{q}; \beta)$ and $\mathbb{S}_n^{(0)}(\mathbf{q}; \alpha)$, then

$$\begin{aligned} s_n^{(l)}(t, \mathbf{q}; \beta) &= \mathbb{E} \left[C_i \mathcal{Y}_i(t) \{ q_i^* Z_i^{(1)l} r_i^{(1)}(\beta) + q_i^{*'} Z_i^{(0)l} r_i^{(0)}(\beta) \} \right] \\ &\quad + n^{-1} \sum_{i=1}^n q_i \mu_{q_i^*} \left(C_i \mathcal{Y}_i(t) Z_i^{(1)l} r_i^{(1)}(\beta) \right) + n^{-1} \sum_{i=1}^n q'_i \mu_{q_i^*} \left(C_i \mathcal{Y}_i(t) Z_i^{(0)l} r_i^{(0)}(\beta) \right), \\ \overline{s}_n^{(1)}(t, \mathbf{q}; \beta) &= \mathbb{E} \left[C_i \mathcal{Y}_i(t) \{ q_i^* Z_i^{(1)l} r_i^{(1)}(\beta) + q_i^{*'} Z_i^{(0)l} r_i^{(0)}(\beta) \} \right] \\ &\quad + n^{-1} \sum_{i=1}^n q_i \mu_{q_i^*} \left(C_i \mathcal{Y}_i(t) Z_i^{(1)l} r_i^{(1)}(\beta) \right) + n^{-1} \sum_{i=1}^n q'_i \mu_{q_i^*} \left(C_i \mathcal{Y}_i(t) Z_i^{(0)l} r_i^{(0)}(\beta) \right), \\ \text{and } s_n^{(0)}(\mathbf{q}; \alpha) &= \alpha^T \mathbb{E} \left[C_i q_i^{*'} X_i \right] - \mathbb{E} \left[\log \left(1 + e^{\alpha^T X_i} \right) \right] + \alpha^T n^{-1} \sum_{i=1}^n q'_i \mu_{q_i^*} \left(C_i X_i \right). \end{aligned}$$

Example: Missing Binary Covariates. For simplicity, we assume that $C_i = 1$ or 0 occurs independently of q_i^* . Letting $w_i^{(j)} = \Pr(C_i = j | X_i, Z_i^{(q_i^*)})$, then

$$\begin{aligned} \mathbb{E} \left[C_i q_i^{*'} X_i \right] &= \mathbb{E} \left[X_i \Pr \left(C_i = 1, q_i^* = 1 | X_i, Z_i^{(q_i^*)} \right) \right] = \mathbb{E} \left[X_i w_i^{(1)} p_i^{(0)} \right], \\ \mathbb{E} \left[C_i \mathcal{Y}_i(t) \{ q_i^* Z_i^{(1)l} r_i^{(1)} + q_i^{*'} Z_i^{(0)l} r_i^{(0)} \} \right] &= \sum_{j=0,1} \mathbb{E} \left[Z_i^{(j)l} r_i^{(j)} w_i^{(1)} p_i^{*(j)} S_i^{*(j)}(t) S_i^U(t) \right], \end{aligned}$$

while the expectations of terms which may form all possible partial-sums in $\mathbb{S}_n^{(0)}(\mathbf{q}; \alpha)$, $\mathbb{S}_n^{(l)}(t, \mathbf{q}; \beta)$ and $\overline{\mathbb{S}}_n^{(l)}(t, \mathbf{q}; \beta)$ are

$$\begin{aligned} \mu_{q_i^*}(C_i'X_i) &= \mathbb{E}\left[X_i \Pr(C_i = 0 | X_i, q_i^*) \middle| q_i^*\right] \\ &= \begin{cases} \mathbb{E}\left[X_i w_i^{(0)} \middle| q_i^* = 0\right] = \frac{\mathbb{E}\left[X_i w_i^{(0)} p_i^{*(0)}\right]}{\mathbb{E}\left[p_i^{*(0)}\right]} & \text{if } q_i^* = 0, \\ \mathbb{E}\left[X_i w_i^{(0)} \middle| q_i^* = 1\right] = \frac{\mathbb{E}\left[X_i w_i^{(0)} p_i^{*(1)}\right]}{\mathbb{E}\left[p_i^{*(1)}\right]} & \text{if } q_i^* = 1, \end{cases} \\ \mu_{q_i^*}(C_i' \mathcal{Y}_i(t) Z_i^{(j)l} r_i^{(j)h}) &= \mathbb{E}\left[Z_i^{(j)l} r_i^{(j)h} \Pr(C_i' \mathcal{Y}_i(t) = 1 | X_i, Z_i^{(q_i^*)}, q_i^*) \middle| q_i^*\right] \\ &= \begin{cases} \frac{\mathbb{E}\left[Z_i^{(j)l} r_i^{(j)h} w_i^{(0)} S_i^{*(0)}(t) S_i^U(t) p_i^{*(0)}\right]}{\mathbb{E}\left[p_i^{*(0)}\right]} & \text{if } q_i^* = 0, \\ \frac{\mathbb{E}\left[Z_i^{(j)l} r_i^{(j)h} w_i^{(0)} S_i^{*(1)}(t) S_i^U(t) p_i^{*(1)}\right]}{\mathbb{E}\left[p_i^{*(1)}\right]} & \text{if } q_i^* = 1, \end{cases} \quad j = 0, 1. \end{aligned}$$

In these calculi, note that the Bayes rule is used, such as

$$\mathbb{E}\left[X_i w_i^{(0)} \middle| q_i^* = j\right] = \int X_i w_i^{(0)} d\Pr(X_i | q_i^* = j) = \int X_i w_i^{(0)} d\Pr(X_i) \frac{\Pr(q_i^* = j | X_i)}{\Pr(q_i^* = j)}.$$

Example: Cox Cure-Mixture Model. In this model, $C_i = \Delta_i$ is usually assumed. The expectations of terms which may form all possible partial-sums are $\mu_{q_i^*}(C_i' \mathcal{Y}_i(t) Z_i^{(0)l} r_i^{(0)}) = 0$ and

$$\begin{aligned} \mu_{q_i^*}(\Delta_i' \mathcal{Y}_i(t) Z_i^{(1)l} r_i^{(1)}) &= \mathbb{E}\left[Z_i^{(1)l} r_i^{(1)} \Pr\left((1 - \Delta_i) \mathcal{Y}_i(t) = 1 \middle| X_i, Z_i^{(q_i^*)}, q_i^*\right) \middle| q_i^*\right] \\ &= \begin{cases} \mathbb{E}\left[Z_i^{(1)l} r_i^{(1)} Q_i^0(t) \middle| q_i^* = 1\right] = \frac{\mathbb{E}\left[Z_i^l r_i(\beta) Q_i^0(t) p_i^{*(1)}\right]}{\mathbb{E}\left[p_i^{*(1)}\right]} & \text{if } q_i^* = 1 \\ \mathbb{E}\left[Z_i^{(1)l} r_i^{(1)} S_i^U(t) \middle| q_i^* = 0\right] = \frac{\mathbb{E}\left[Z_i^l r_i(\beta) S_i^U(t) p_i^{*(0)}\right]}{\mathbb{E}\left[p_i^{*(0)}\right]} & \text{if } q_i^* = 0 \end{cases}, \end{aligned}$$

where $Q_i^0(t) = \Pr(t \leq T_i, \Delta_i = 0 | X_i, Z_i^{(1)}, q_i^* = 1) = -\int_t^{\tau_c} S_i^{*(1)}(s) dS_i^U(s)$. Similarly, we have

$$\mu_{q_i^*}(\Delta_i' X_i) = \mathbb{E}\left[X_i \Pr\left(\Delta_i = 0 \middle| Z_i^{(q_i^*)}, q_i^*\right) \middle| q_i^*\right] = \begin{cases} \frac{\mathbb{E}\left[X_i Q_i^0(0) p_i^{*(1)}\right]}{\mathbb{E}\left[p_i^{*(1)}\right]} & \text{if } q_i^* = 1 \\ \frac{\mathbb{E}\left[X_i p_i^{*(0)}\right]}{\mathbb{E}\left[p_i^{*(0)}\right]} & \text{if } q_i^* = 0 \end{cases}.$$

On Weak Convergence. Let $\tilde{Y}_i(t; \theta) = Y_i(t; \theta) - \mu_{q_i^*}(Y_i(t; \theta))$. In our application, note that centred $\tilde{Y}_i(t; \theta)$, $i = 1, \dots, n$ are zero-means and mutually independent but are not sampled from an identical distribution. We will therefore discuss the partial-sum processes about $\tilde{Y}_i(t; \theta)$ sampled from two populations. Lemma 3 shows that $n^{-1} \mathbb{T}_n(\tilde{Y}(t; \theta), \mathbf{q})$ converges in probability to zero uniformly on $[0, \tau_g] \times \Theta \times \{0, 1\}^\infty$; recall that an n -dimensional element $\mathbf{q} \in \{0, 1\}^n$ means marginal collection of elements of $\{0, 1\}$. In advance, let

$$K_{t, \theta}^\delta = \left\{ (t_0, \theta_0) : \mathbb{E}\left[|\tilde{Y}_i(t_0; \theta_0) - \tilde{Y}_i(t; \theta)|\right] < \delta \right\}.$$

Remark 3. For example, if $\mathbb{E}\left[\tilde{Y}_i(t; \theta)^2\right] < \infty$, by Chebyshev's inequality, we immediately have

$$\lim_{n \rightarrow \infty} \sup_q \Pr\left(\left|n^{-1} \mathbb{T}_n(\tilde{Y}(t; \theta), \mathbf{q})\right| > \varepsilon\right) = 0.$$

However, a result of interest here is whether \Pr and \sup_q can be exchanged, that is, about

$$\lim_{n \rightarrow \infty} \Pr\left(\sup_q \left|n^{-1} \mathbb{T}_n(\tilde{Y}(t; \theta), \mathbf{q})\right|\right).$$

Incidentally, we cannot obtain the almost sure convergence in this problem, since

$$\sup_{\mathbf{q}} \left| n^{-1} \mathbb{T}_n(\tilde{Y}(t; \theta), \mathbf{q}) \right| = \mathbb{E}_n \left[1(\tilde{Y}_i(t; \theta) \geq 0) \tilde{Y}_i(t; \theta) \right]$$

is always apart from zero. \square

Lemma 3. Let $\tilde{Y}_i(\cdot; \theta)$, $i=1, \dots, n$ be random elements on $D[0, \tau_e]$ sampled from one of two distributions (populations) with $\mathbb{E}[\tilde{Y}_i(t; \theta)] = 0$ at every (t, θ) , where the population to which the i -th element $\tilde{Y}_i(\cdot; \theta)$ belongs is known and indexed by $q_i^* = 0$ or 1. Suppose that $\tilde{Y}_1(\cdot; \theta), \dots, \tilde{Y}_n(\cdot; \theta)$ are mutually independent and have $\mathbb{E}[\tilde{Y}_i(t; \theta)] = 0$ at every (t, θ) . If the following three conditions are satisfied,

- (i) The class of functions $\{\tilde{Y}_i(t; \theta) : (t, \theta) \in [0, \tau_e] \times \Theta\}$ is Glivenko-Cantelli,
- (ii) $\mathbb{E} \left[\sup_{t \in [0, \tau_e], \theta \in \Theta} |\tilde{Y}_i(t; \theta)| \right] < \infty$ and
- (iii) $\mathbb{E} \left[\sup_{(t_0, \theta_0) \in K_{t, \theta}^\delta} |\tilde{Y}_i(t_0; \theta_0) - \tilde{Y}_i(t; \theta)| \right] \leq O(1)\delta$ for every $K_{t, \theta}^\delta \subset [0, \tau_e] \times \Theta$,

then, as $n \rightarrow \infty$

$$n^{-1} \mathbb{T}_n(\tilde{Y}(t; \theta), \mathbf{q}) \rightarrow_D 0 \text{ on } (t, \theta, \mathbf{q}) \in [0, \tau_e] \times \Theta \times \{0, 1\}^\infty.$$

Lemma 3 is proved in Appendix A.1. The following examples show that the conditions needed in Lemma 3 are satisfied for $Y_i(t; \alpha) = C_i' X_i$, $Y_i(t; \beta) = C_i' \mathcal{Y}_i(t) Z_i^{(1)l} r_i^{(1)h}$ and $Y_i(t; \beta) = C_i' \mathcal{Y}_i(t) Z_i^{(0)l} r_i^{(0)h}$ ($l, h = 0, 1$).

Example 1. Let $\tilde{Y}_i(t; \theta) = C_i' X_i - \mu_{q_i^*}(C_i' X_i)$. From Condition A5, we have Condition (ii). Since $\tilde{Y}_i(t; \theta)$ is independent of (t, θ) , Conditions (i) and (iii) are clearly satisfied.

Example 2. Let $\tilde{Y}_i(t; \beta) = C_i' \mathcal{Y}_i(t) Z_i^{(j)l} r_i^{(j)h} - \mu_{q_i^*}(C_i' \mathcal{Y}_i(t) Z_i^{(j)l} r_i^{(j)h})$, $j=0, 1$. Conditions (i) and (ii) are shown by Conditions A1 and A4. Arbitrary $(t_0, \beta_0) \in K_{t, \beta}^\delta$ satisfies $|\beta_0 - \beta| \leq O(1)\delta$ by Condition A4 and $\mathbb{E}[|Y_i(t_0; \beta_0) - Y_i(t; \beta_0)|] \leq O(1)\delta$ by Condition A3, where $|\cdot|$ means the Euclidean norm for vectors. Hence, Condition (iii) is satisfied.

3.3. Proof of Theorem 1

Consider $\bar{\ell}_{p(n)}(\theta)$, $\tilde{\ell}_{p(n)}(\theta)$ and $\ell_{p(n)}^\dagger(\theta)$ in which the random quantities are reduced less than that of $\ell_{p(n)}(\theta)$, where $\bar{\ell}_{p(n)}(\theta)$ and $\tilde{\ell}_{p(n)}(\theta)$ are $\ell_{p(n)}(\theta)$ in which $\ell_{\pi p(n)}(\theta)$ is replaced by $\tilde{\ell}_{\pi p(n)}(\theta; \mathbf{q})$ and $\ell_{\pi p(n)}^\dagger(\theta; \mathbf{q})$, i.e. $\tilde{\ell}_{p(n)}(\theta) = \ell_{p(n)}(\theta) \Big|_{\ell_{\pi p(n)}(\theta; \mathbf{q}) = \tilde{\ell}_{\pi p(n)}(\theta; \mathbf{q})}$, $\ell_{p(n)}^\dagger(\theta) = \ell_{p(n)}(\theta) \Big|_{\ell_{\pi p(n)}(\theta; \mathbf{q}) = \ell_{\pi p(n)}^\dagger(\theta; \mathbf{q})}$ and

$$\begin{aligned} \tilde{\ell}_{\pi p(n)}(\theta; \mathbf{q}) &= s_n^{(0)}(\mathbf{q}; \alpha) + \int_0^{\tau_e} \left[\beta^T \bar{s}_n^{(1)}(t, \mathbf{q}; \beta^*) - \log \{ \mathbb{S}_n^{(0)}(t, \mathbf{q}; \beta) \} s_n^{(0)}(t, \mathbf{q}^*; \beta^*) \right] d\Lambda_0^*(t), \\ \ell_{\pi p(n)}^\dagger(\theta; \mathbf{q}) &= s_n^{(0)}(\mathbf{q}; \alpha) + \int_0^{\tau_e} \left[\beta^T \bar{s}_n^{(1)}(t, \mathbf{q}; \beta^*) - \log \{ s_n^{(0)}(t, \mathbf{q}; \beta) \} s_n^{(0)}(t, \mathbf{q}^*; \beta^*) \right] d\Lambda_0^*(t). \end{aligned}$$

$\ell_{\pi p(n)}$ and $\bar{\ell}_{\pi p(n)}$: It is satisfied that

$$\sup_s \left| \int_0^s \mathbb{E}_n [d\mathcal{N}_i(t)] - \int_0^s \mathbb{S}_n^{(0)}(t, \mathbf{q}^*; \beta^*) d\Lambda_0^*(t) \right| \rightarrow_{as} 0$$

$$\text{and } \sup_{\beta, s, \mathbf{q} \in \{0, 1\}^\infty} \left| \beta^T \int_0^s \mathbb{E}_n [\tilde{Z}_i^{(q_i)} d\mathcal{N}_i(t)] - \beta^T \int_0^s \bar{\mathbb{S}}_n^{(1)}(t, \mathbf{q}; \beta^*) d\Lambda_0^*(t) \right| \rightarrow_{as} 0$$

by Conditions A1 and A4, similar to the standard Cox model (see [19]). Thus,

$$\sup_{\theta \in \Theta, \mathbf{q} \in \{0, 1\}^\infty} \left| \ell_{\pi p(n)}(\theta; \mathbf{q}) - \bar{\ell}_{\pi p(n)}(\theta; \mathbf{q}) \right| \rightarrow_{as} 0$$

is obtained as $n \rightarrow \infty$.

$\bar{\ell}_{\pi p(n)}$ and $\tilde{\ell}_{\pi p(n)}$: We have

$$\sup_{\theta \in \Theta, \mathbf{q} \in \{0, 1\}^\infty} \left| \mathbb{S}_n^{(0)}(\mathbf{q}; \alpha) - s_n^{(0)}(\mathbf{q}; \alpha) \right| \rightarrow_p 0$$

using Lemma 3 in Example 1 and applying the strong law of large numbers (SLLN) to $\log p_i^{(1)}(\alpha)$ and $\alpha^T C_i q_i^* X_i$ by Conditions A1 and A5. For the latter application, note that $\mathbb{E}[|\log p_i^{(1)}|] \leq \log 2 + \mathbb{E}[|\alpha^T X_i|]$ because of $\log(1+x) \leq \log 2 + 1(x \geq 1) \log x$ ($x \geq 0$). Also, it is shown that

$$\sup_t \left| \mathbb{S}_n^{(0)}(t, \mathbf{q}^*; \beta^*) - s_n^{(0)}(t, \mathbf{q}^*; \beta^*) \right| \rightarrow_{\text{as}} 0$$

by applying the SLLN on $D[0, \tau_e]$ (see [1]) from Conditions A1 and A4. In addition, we have

$$\sup_{\mathbf{q} \in \{0,1\}^\infty} \left| \overline{\mathbb{S}}_n^{(1)}(t, \mathbf{q}; \beta^*) - \overline{s}_n^{(1)}(t, \mathbf{q}; \beta^*) \right| \rightarrow_p 0$$

using Lemma 3 in Example 2. Hence, $\sup_{\theta \in \Theta, \mathbf{q} \in \{0,1\}^\infty} \left| \overline{\ell}_{p(n)}(\theta; \mathbf{q}) - \tilde{\ell}_{p(n)}(\theta; \mathbf{q}) \right|$ converges in probability to zero as $n \rightarrow \infty$.

$\tilde{\ell}_{p(n)}$ and $\ell_{p(n)}^\dagger$: It satisfies that

$$\sup_{\theta, t, \mathbf{q}} \left| \log \mathbb{S}_n^{(0)}(t, \mathbf{q}; \beta) - \log s_n^{(0)}(t, \mathbf{q}; \beta) \right| \rightarrow_p 0$$

using Lemma 3 in Example 2 and the continuous mapping theorem about log-function. For the latter application, note that $s_n^{(0)}(t, \mathbf{q}; \beta)$ is bounded away from zero on $[0, \tau_e] \times \{0,1\}^\infty \times \Theta$ by Condition A2. Hence,

$$\sup_{\theta \in \Theta, \mathbf{q} \in \{0,1\}^\infty} \left| \tilde{\ell}_{p(n)}(\theta; \mathbf{q}) - \ell_{p(n)}^\dagger(\theta; \mathbf{q}) \right|$$

converges in probability to zero as $n \rightarrow \infty$.

Applying the above three results to Lemmas 1 and 2, therefore, we obtain

$$\sup_\theta \left| \ell_{p(n)}(\theta) - \overline{\ell}_{p(n)}(\theta) \right| \rightarrow_{\text{as}} 0, \quad \sup_\theta \left| \overline{\ell}_{p(n)}(\theta) - \tilde{\ell}_{p(n)}(\theta) \right| \rightarrow_{\text{as}} 0 \quad \text{and} \quad \sup_\theta \left| \tilde{\ell}_{p(n)}(\theta) - \ell_{p(n)}^\dagger(\theta) \right| \rightarrow_{\text{as}} 0,$$

respectively, so that we conclude

$$\sup_{\theta \in \Theta} \left| \ell_{p(n)}(\theta) - \ell_{p(n)}^\dagger(\theta) \right| \rightarrow_{\text{as}} 0 \quad \text{as } n \rightarrow \infty. \tag{3.4}$$

Although (3.4) shows that the limit of $\ell_{p(n)}(\theta)$ is equivalent to that of $\ell_{p(n)}^\dagger(\theta)$, $\ell_{p(n)}^\dagger(\theta)$ still depends on n and q_i^* 's. We will therefore investigate the limit form of $\ell_{p(n)}^\dagger(\theta)$ further.

In discussing a convergence about the form $n^{-1} \sum_{i=1}^n q_i \mu_{q_i^*}(\cdot)$ and $n^{-1} \sum_{i=1}^n q_i' \mu_{q_i^*}(\cdot)$ included in $s_n^{(0)}$ and $\overline{s}_n^{(1)}$ of the partial sums, note that they can be written as

$$\begin{aligned} n^{-1} \sum_{i=1}^n q_i \mu_{q_i^*}(\cdot) &= \mathbb{E}_n \left[1(q_i = 1, q_i^* = 1) \right] \mu_1(\cdot) + \mathbb{E}_n \left[1(q_i = 1, q_i^* = 0) \right] \mu_0(\cdot), \\ n^{-1} \sum_{i=1}^n q_i' \mu_{q_i^*}(\cdot) &= \mathbb{E}_n \left[1(q_i = 0, q_i^* = 1) \right] \mu_1(\cdot) + \mathbb{E}_n \left[1(q_i = 0, q_i^* = 0) \right] \mu_0(\cdot). \end{aligned}$$

Let $v_h^{(j)}(\mathbf{q}) = \mathbb{E} \left[1(q_i = h, q_i^* = j) \right]$, $h, j = 0, 1$. Similarly to Lemma 3 (proof of **s2**), we show that

$$\mathbb{E}_n \left[1(q_i = h, q_i^* = l) \right] \rightarrow_p v_h^{(j)}(\mathbf{q}) \quad \text{as } n \rightarrow \infty$$

at arbitrary point $\mathbf{q} \in \{0,1\}^\infty$. In particular, because of

$$\Pr(q_i = h, q_i^* = j | X_i) = \Pr(q_i^* = j | X_i) \Pr(q_i = h | X_i, q_i^* = j),$$

note that

$$v_h^{(j)}(\mathbf{q}) = \mathbb{E} \left[\mathbb{E} \left[1(q_i = h, q_i^* = j) | X_i \right] \right] = \mathbb{E} \left[p_i^{*(j)} \Pr(q_i = h | X_i, q_i^* = j) \right]$$

and then $v_1^{(j)}(\mathbf{q}) + v_0^{(j)}(\mathbf{q}) = \mathbb{E} \left[p_i^{*(j)} \right]$. We have the following lemma.

Lemma 4. $n^{-1} \sum_{i=1}^n q_i \mu_{q_i^*}(\cdot)$ and $n^{-1} \sum_{i=1}^n q_i' \mu_{q_i^*}(\cdot)$ converge in probability to

$$v_1^{(1)}(\mathbf{q}) \mu_1(\cdot) + v_1^{(0)}(\mathbf{q}) \mu_0(\cdot) \quad \text{and} \quad v_0^{(0)}(\mathbf{q}) \mu_1(\cdot) + v_0^{(1)}(\mathbf{q}) \mu_0(\cdot)$$

uniformly on $\mathbf{q} \in \{0,1\}^\infty$.

A proof of Lemma 4 is provided briefly in Appendix A.2 since it is similar to Lemma 3. Now, applying Lemma 4 to $s_n^{(0)}(t, \mathbf{q}; \beta)$, $\overline{s}_n^{(1)}(t, \mathbf{q}; \beta)$ and $s_n^{(0)}(\mathbf{q}; \alpha)$, we obtain their limits as

$$\begin{aligned} s_*^{(0)}(t, \mathbf{q}; \beta) &= \mathbb{E} \left[C_i \mathcal{Y}_i(t) \left\{ q_i^* Z_i^{(1)l} r_i^{(1)}(\beta) + q_i^{*'} Z_i^{(0)l} r_i^{(0)}(\beta) \right\} \right] \\ &\quad + v_1^{(1)}(\mathbf{q}) \mu_1 \left(C_i \mathcal{Y}_i(t) Z_i^{(1)l} r_i^{(1)}(\beta) \right) + v_1^{(0)}(\mathbf{q}) \mu_0 \left(C_i \mathcal{Y}_i(t) Z_i^{(1)l} r_i^{(1)}(\beta) \right) \\ &\quad + v_0^{(1)}(\mathbf{q}) \mu_1 \left(C_i \mathcal{Y}_i(t) Z_i^{(0)l} r_i^{(0)}(\beta) \right) + v_0^{(0)}(\mathbf{q}) \mu_0 \left(C_i \mathcal{Y}_i(t) Z_i^{(0)l} r_i^{(0)}(\beta) \right), \end{aligned}$$

$$s_*^{(1)}(t, \mathbf{q}; \beta) = \mathbb{E} \left[C_i \mathcal{Y}_i(t) \left\{ q_i^* Z_i^{(1)} r_i^{(1)}(\beta) + q_i^{**} Z_i^{(0)} r_i^{(0)}(\beta) \right\} \right] \\ + v_1^{(1)}(\mathbf{q}) \mu_1 \left(C_i' \mathcal{Y}_i(t) Z_i^{(1)'} r_i^{(1)}(\beta) \right) + v_1^{(0)}(\mathbf{q}) \mu_0 \left(C_i' \mathcal{Y}_i(t) Z_i^{(1)'} r_i^{(0)}(\beta) \right) \\ + v_0^{(1)}(\mathbf{q}) \mu_1 \left(C_i' \mathcal{Y}_i(t) Z_i^{(0)'} r_i^{(1)}(\beta) \right) + v_0^{(0)}(\mathbf{q}) \mu_0 \left(C_i' \mathcal{Y}_i(t) Z_i^{(0)'} r_i^{(0)}(\beta) \right),$$

$$\text{and } s_*^{(0)}(\mathbf{q}; \alpha) = \alpha^T \mathbb{E} \left[C_i q_i^* X_i \right] - \mathbb{E} \left[\log \left(1 + e^{\alpha^T X_i} \right) \right] + \alpha^T v_0^{(1)}(\mathbf{q}) \mu_1 \left(C_i' X_i \right) + \alpha^T v_0^{(0)}(\mathbf{q}) \mu_0 \left(C_i' X_i \right).$$

Let $\ell_{p(*)}^\dagger(\theta; \mathbf{q})$ be $\ell_{p(n)}^\dagger(\theta; \mathbf{q})$, in which $\ell_{\pi p(n)}^\dagger(\theta; \mathbf{q})$ is replaced by

$$\ell_{\pi p(*)}^\dagger(\theta; \mathbf{q}) = s_*^{(0)}(\mathbf{q}; \alpha) + \int_0^{\tau_e} \left[\beta^T \bar{s}_*^{(1)}(t, \mathbf{q}^*; \beta^*) - \log \left\{ s_*^{(0)}(t, \mathbf{q}; \beta) \right\} s_*^{(0)}(t, \mathbf{q}^*; \beta^*) \right] d\Lambda_*^*(t).$$

$\ell_{\pi p(n)}^\dagger$ and $\ell_{\pi p(*)}^\dagger$: We obtain $\sup_{\theta, t, \mathbf{q}} \left| \log s_n^{(0)}(t, \mathbf{q}; \beta) - \log s_*^{(0)}(t, \mathbf{q}; \beta) \right| \rightarrow_p 0$, $\sup_{t, \mathbf{q}} \left| \bar{s}_n^{(1)}(t, \mathbf{q}; \beta^*) - \bar{s}_*^{(1)}(t, \mathbf{q}; \beta^*) \right| \rightarrow_p 0$ and $\sup_{\theta, \mathbf{q}} \left| \log s_n^{(0)}(\mathbf{q}; \alpha) - \log s_*^{(0)}(\mathbf{q}; \alpha) \right| \rightarrow_p 0$ by Lemma 4 and the continuous mapping theorem about log-function. Therefore, $\sup_{\theta \in \Theta, \mathbf{q} \in \{0,1\}^\infty} \left| \ell_{\pi p(n)}^\dagger(\theta; \mathbf{q}) - \ell_{\pi p(*)}^\dagger(\theta; \mathbf{q}) \right|$ converges in probability to zero as $n \rightarrow \infty$. Hence, using Lemmas 1 and 2, we can show $\sup_{\theta \in \Theta} \left| \ell_{p(n)}^\dagger(\theta) - \ell_{p(*)}^\dagger(\theta) \right| \rightarrow_{\text{as}} 0$, so that a triangle combination of this result and (3.4) yields

$$\sup_{\theta \in \Theta} \left| \ell_{p(n)}(\theta) - \ell_{p(*)}^\dagger(\theta) \right| \rightarrow_{\text{as}} 0 \text{ as } n \rightarrow \infty. \tag{3.5}$$

On a Limit Form. The result of (3.5) shows only that the limit of $\ell_{p(n)}(\theta)$ is equivalent to that of $\ell_{p(*)}^\dagger(\theta)$. Here we discuss a limit form of $\ell_{p(*)}^\dagger(\theta)$. To consider the case of $s = -1$, let $\{0,1\}_+^n$ and $\{0,1\}_-^n$ be the subsets of $\{0,1\}^n$ such that $\mathbb{I}_s(\mathbf{q}) = 1$ if $\mathbf{q} \in \{0,1\}_+^n$ and $\mathbb{I}_s(\mathbf{q}) = -1$ if $\mathbf{q} \in \{0,1\}_-^n$. For simplicity, let

$$f^\dagger(\mathbf{q}) = \exp \left(\ell_{\pi p(*)}^\dagger(\theta; \mathbf{q}) \right).$$

Then,

$$\exp \left(\ell_{p(*)}^\dagger(\theta) \right) = B_n^{1/n} \left\{ \int_{\mathbf{q} \in \{0,1\}_+^n} f^\dagger(\mathbf{q})^n d\nu_n(\mathbf{q}) - \int_{\mathbf{q} \in \{0,1\}_-^n} f^\dagger(\mathbf{q})^n d\nu_n(\mathbf{q}) \right\}^{1/n} \\ = B_n^{1/n} \left\{ \int_{\mathbf{q} \in \{0,1\}_+^n} f^\dagger(\mathbf{q})^n d\nu_n(\mathbf{q}) \right\}^{1/n} \times \left\{ 1 - \frac{\int_{\mathbf{q} \in \{0,1\}_-^n} f^\dagger(\mathbf{q})^n d\nu_n(\mathbf{q})}{\int_{\mathbf{q} \in \{0,1\}_+^n} f^\dagger(\mathbf{q})^n d\nu_n(\mathbf{q})} \right\}^{1/n}.$$

Because of $\exp \left(\ell_{p(*)}^\dagger(\theta) \right) > 0$, we have

$$1 > \int_{\mathbf{q} \in \{0,1\}_-^n} f^\dagger(\mathbf{q})^n d\nu_n(\mathbf{q}) / \int_{\mathbf{q} \in \{0,1\}_+^n} f^\dagger(\mathbf{q})^n d\nu_n(\mathbf{q}) \geq 0,$$

so that, via the general binomial theorem, we can show that

$$\lim_{n \rightarrow \infty} \left\{ 1 - \int_{\mathbf{q} \in \{0,1\}_-^n} f^\dagger(\mathbf{q})^n d\nu_n(\mathbf{q}) / \int_{\mathbf{q} \in \{0,1\}_+^n} f^\dagger(\mathbf{q})^n d\nu_n(\mathbf{q}) \right\}^{1/n} = 1.$$

Also, $(B_n)^{1/n} \rightarrow_{\text{as}} B_* = 2^{1-E[C_i]}$. Therefore, because of $\nu_n \left(\{0,1\}_+^n \right) \leq 1$, and similar to the derivation of the ℓ^∞ -norm on the Banach space which results in the essential supremum, we conclude

$$\lim_{n \rightarrow \infty} \exp \left(\ell_{p(*)}^\dagger(\theta) \right) = B_* \lim_{n \rightarrow \infty} \left\{ \int_{\mathbf{q} \in \{0,1\}_+^n} f^\dagger(\mathbf{q})^n d\nu_n(\mathbf{q}) \right\}^{1/n} \\ = B_* \text{ess.sup}_{\mathbf{q} \in \{0,1\}_+^\infty} \exp \left(\ell_{\pi p(*)}^\dagger(\theta; \mathbf{q}) \right). \tag{3.6}$$

In addition, (3.6) is derived in the case of $s = 1$. Results (3.5) and (3.6) show that Theorem 1 is complete. \square

A limit function of $\ell_{p(n)}(\theta)$ is concretely provided by (3.6), which is summarized as follows.

Corollary 1. If Theorem 1 holds, then a limit expression to which $\ell_{p(n)}(\theta)$ converges almost surely as $n \rightarrow \infty$ is $\log B_* + \text{ess.sup}_{\mathbf{q} \in \{0,1\}_+^\infty} \ell_{\pi p(*)}^\dagger(\theta; \mathbf{q})$.

4. Additional Considerations

4.1. Monte Carlo (MC) Approximations

It usually takes a long time for the exact computation of the OPL. So, another subject of interest is the performance of its MC approximations. Let $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_{2^n}$ be all the elements of $\{0, 1\}^n$ labelled in order such that

$$\ell_{\pi p(n)}(\theta; \mathbf{q}_{2^n}) \mathbb{I}_s(\mathbf{q}_{2^n}) \leq \dots \leq \ell_{\pi p(n)}(\theta; \mathbf{q}_2) \mathbb{I}_s(\mathbf{q}_2) \leq \ell_{\pi p(n)}(\theta; \mathbf{q}_1) \mathbb{I}_s(\mathbf{q}_1).$$

We assign a point $j/2^n \in (0, 1]$ to \mathbf{q}_j using $\zeta(j/2^n) = \mathbf{q}_j$ and let $\nu_n(\zeta(t))$ denote the distribution $\nu_n(\mathbf{q} \in \{0, 1\}^n : \zeta^{-1}(\mathbf{q}) \leq t)$. Using these notations, we redefine $\exp(n\ell_{\pi p(n)}(\theta))$ as

$$\mathbb{M}_n(\theta) = \exp(n\ell_{\pi p(n)}(\theta)) = B_n \int_{t \in (0, 1]} \exp(n\ell_{\pi p(n)}(\theta; \zeta(t))) \mathbb{I}_s(\zeta(t)) d\nu_n(\zeta(t)).$$

Given fixed data $(T_i, \Delta_i, C_i, X_i, Z_i^{(0)}, Z_i^{(1)})$, $i = 1, \dots, n$, let $\tilde{\mathbf{q}}_1, \dots, \tilde{\mathbf{q}}_k$ be k random elements from $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_{2^n}\}$, where $\tilde{\mathbf{q}}_j = (\tilde{q}_{j1}, \dots, \tilde{q}_{jn})$ and \tilde{q}_{ji} is either 1 or 0 with an equal probability 0.5 if $C_i = 0$ and $\tilde{q}_{ji} = q_i^*$ if $C_i = 1$. An MC approximation of $\tilde{\mathbb{M}}_n(\theta)$ is

$$\tilde{\mathbb{M}}_n^{(k)}(\theta) = B_n \int_{t \in (0, 1]} \exp(n\ell_{\pi p(n)}(\theta; \zeta(t))) \mathbb{I}_s(\zeta(t)) d\tilde{\nu}_n^{(k)}(\zeta(t))$$

using $\{\tilde{\mathbf{q}}_1, \dots, \tilde{\mathbf{q}}_k\}$ and the corresponding empirical measure

$$\tilde{\nu}_n^{(k)}(\zeta(t)) = k^{-1} \sum_{j=1}^k 1(\zeta^{-1}(\tilde{\mathbf{q}}_j) \leq t).$$

By the standard asymptotic theory, as $k \rightarrow \infty$, it follows that

$$\tilde{\mathbb{M}}_n^{(k)}(\theta) \xrightarrow{\text{as}} \mathbb{M}_n(\theta) \text{ and } \sqrt{k} \{ \tilde{\mathbb{M}}_n^{(k)}(\theta) - \mathbb{M}_n(\theta) \} \rightarrow_D N(0, \mathbb{V}_n(\theta)), \tag{4.1}$$

provided $\mathbb{M}_n(\theta)$ exists and $\mathbb{V}_n(\theta) < \infty$, where

$$\mathbb{V}_n(\theta) = B_n^2 \int_{t \in (0, 1]} \exp(2n\ell_{\pi p(n)}(\theta; \zeta(t))) d\nu_n(\zeta(t)) - \mathbb{M}_n(\theta)^2.$$

To evaluate the quantity of $\mathbb{V}_n(\theta)$ in the case of $n \rightarrow \infty$, consider $\mathbb{V}_n(\theta)^{1/n}$, then, as $n \rightarrow \infty$

$$\mathbb{V}_n(\theta)^{1/n} \xrightarrow{\text{as}} B_* \text{ess.sup}_{t \in (0, 1]} \exp(2\ell_{\pi p(*)}^\dagger(\theta; \zeta(t))),$$

similar to the discussion for (3.6). As this result means that $\mathbb{V}_n(\theta)$ may increase exponentially according to n , direct use of (4.1) is not particularly productive. Therefore, although (4.1) is the rationale in this context, it will be modified, as $k \rightarrow \infty$, to

$$\begin{aligned} n^{-p} \log(\tilde{\mathbb{M}}_n^{(k)}(\theta)) &\xrightarrow{\text{as}} n^{-p} \log(\mathbb{M}_n(\theta)) \\ \text{and } \sqrt{k} n^{-p} \{ \log(\tilde{\mathbb{M}}_n^{(k)}(\theta)) - \log(\mathbb{M}_n(\theta)) \} &\rightarrow_D N(0, \mathbb{M}_n(\theta) / n^{2p} \mathbb{M}_n(\theta)^2) \end{aligned} \tag{4.2}$$

using the delta method. Now consider the other aspect of (4.2) under $p = 1$. Applying Theorem 1 and Corollary 1 to such a problem, we obtain the following results

$$\begin{aligned} n^{-1} \log(\tilde{\mathbb{M}}_n^{(k)}(\theta)) &\xrightarrow{\text{as}} \log B_* + \text{ess.sup}_{t \in \{t_1, \dots, t_k\}} \ell_{\pi p(*)}^\dagger(\theta; \zeta(t)) \text{ as } n \rightarrow \infty, \\ &\xrightarrow{\text{as}} \log B_* + \text{ess.sup}_{t \in (0, 1)} \ell_{\pi p(*)}^\dagger(\theta; \zeta(t)) \text{ as } k \rightarrow \infty, \end{aligned}$$

where $t_j = \zeta^{-1}(\tilde{\mathbf{Q}}_j^n)$. This means $\lim_{n \rightarrow \infty} \mathbb{V}_n(\theta) / n^2 \mathbb{M}_n(\theta)^2 \leq O(1)$ from the point of view of the k -asymptotic variance in the second line of (4.2). Hence, we can show that the order of $\mathbb{V}_n(\theta) / \mathbb{M}_n(\theta)^2$ is less than $O(n^2)$ in (4.2) under $p = 0$. That is, using the MC method, a computational load of $O(2^n)$ needed in the exact computation can be reduced to one of at most $O(n^2)$.

4.2. Numerical Examples

We will investigate two circumstances in the finite samples using the Cox cure-mixture model. One is how a relation such as (3.6) obtained as $n \rightarrow \infty$ is located in the finite samples. The other is to observe numerically the practical size of the error in MC approximations, $\mathbb{V}_n(\theta) / k \mathbb{M}_n(\theta)^2$, which was shown to be less than

$O(n^2)/k$ in the previous section.

Ovarian Cancer Data: For the first purpose, we use survival data of ovarian cancer patients [20]. We set the covariates as $X_i = (Treat, Age, Rdisease)$ and $Z_i^{(1)} = X_i$, $i = 1, \dots, 26$, where *Treat* is the type of chemotherapy (0 = single, 1 = combined), *Age* is the age of the patient (in years) and *Rdisease* is the extent of residual disease (0 = complete, 1 = incomplete). The maximum of the OPL $\ell_{p(n)}(\theta)$ is achieved approximately at

$$\hat{\theta} = (\hat{\alpha}^T, \hat{\beta}^T) = (1.702, 1.020, -0.054, -9.276, -1.643, 0.158, -2.277)^T.$$

Here, let $\theta^{(j)} = (j/5) \times \hat{\theta}$, $j = 1, \dots, 5$. **Figure 1** shows plots of

$$y(t; \theta) = (B_n)^{1/n} \exp(\ell_{p(n)}(\theta; \zeta(t))) \text{ on } t \in (0, 1]$$

and $\mathbb{M}_n(\theta)^{1/n}$ at $\theta = \theta^{(j)}$, where $B_n = 2^{14}$ and the y-axis is drawn in exponential scale. Although the total number of $\ell_{p(n)}(\theta; \mathbf{q})$ is 2^{26} , in fact $\ell_{p(n)}(\theta; \mathbf{q})$'s of 2^{14} are sorted on $t = \zeta^{-1}(\mathbf{q}) \in (0, 1)$. This data are small in size ($n = 26$). However, circumstances close to the relation in (3.6) are observed at least at $\theta = \theta^{(5)}, \theta^{(4)}$ and $\theta^{(3)}$.

Simulated Data: For the second purpose, we prepare simulated data with $\alpha_0^* = \log(0.6/0.4)$, $\alpha_1^* = -1$, $\beta_0^* = 1$ and $X_i = Z_i^{(1)}$, where X_i follows the standard uniform distribution. The latent distribution of T_i^* is standard exponential and the censoring follows a uniform distribution [0, 3.65]. Under these settings, the simulated means of cure and censored rate are about 48% and 58%, respectively. We generate 100 pairs of simulated data of size n . We perform m MC approximations for each simulated data set. Let $\tilde{\mathbb{M}}_n^{j(k)}(\theta)$ be the j -th element of m $\tilde{\mathbb{M}}_n^{(k)}(\theta)$'s. For each simulated data set, we estimate $\log \mathbb{M}_n(\theta)$ and $\nabla_n(\theta)/k \mathbb{M}_n(\theta)^2$ by $\log(\bar{\mathbb{M}}_{n,m}^{(k)}(\theta))$ and $\bar{V}_{n,m}^{(k)}(\theta)/\bar{\mathbb{M}}_{n,m}^{(k)}(\theta)^2$, where

$$\bar{\mathbb{M}}_{n,m}^{(k)}(\theta) = m^{-1} \sum_{j=1}^m \tilde{\mathbb{M}}_n^{j(k)}(\theta) \text{ and } \bar{V}_{n,m}^{(k)}(\theta) = m^{-1} \sum_{j=1}^m (\tilde{\mathbb{M}}_n^{j(k)}(\theta) - \bar{\mathbb{M}}_{n,m}^{(k)}(\theta))^2.$$

We use these to observe a better estimation performance than

$$\hat{l}_{n,m}^{(k)}(\theta) = m^{-1} \sum_{j=1}^m \log \tilde{\mathbb{M}}_n^{j(k)}(\theta) \text{ and } m^{-1} \sum_{j=1}^m (\log \tilde{\mathbb{M}}_n^{j(k)}(\theta) - \log \bar{\mathbb{M}}_{n,m}^{(k)}(\theta))^2.$$

Figure 2 shows simulated averages and standard errors (SEs) of 100 pairs of

$$\log \mathbb{M}_n(\theta), \log(\bar{\mathbb{M}}_{n,m}^{(k)}(\theta)) \text{ and } \hat{l}_{n,m}^{(k)}(\theta)$$

computed at $\theta = \theta^{(j)}$ in simulated data of $n = 30$ under $k = 5n$ and $m = 1000$, where $\theta^{(j)} = (2j/10) \times \theta^*$ ($j = 1, \dots, 10$). Although k is considerably smaller than the $2^{n - \sum C_i}$ needed in the exact method, $\log(\bar{\mathbb{M}}_{n,m}^{(k)}(\theta))$ approximates $\log \mathbb{M}_n(\theta)$ well enough.

Further, even if the approximations were reduced to $m = 30$, the simulated average of $\bar{V}_{30,m}^{(k)}(\theta)/\bar{\mathbb{M}}_{30,m}^{(k)}(\theta)^2$ would still yield sufficiently good approximations of $\nabla_{30}(\theta)/k \mathbb{M}_{30}(\theta)^2$. Based on these empirical findings,

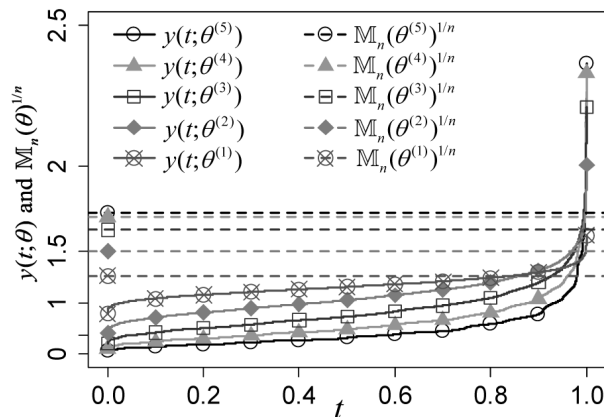


Figure 1. Plots of $(B_n)^{1/n} \exp(\ell_{p(n)}(\theta; \zeta(t)))$ on $t \in (0, 1]$ (solid curves) and $\mathbb{M}_n(\theta)^{1/n}$ (horizontal dotted lines) at $\theta = \theta^{(j)}$, $j = 1, \dots, 5$, in ovarian cancer data (y-axis : exponential scale).

we set $n = 30, 100, 300, 500, 1000$, $k = 5n$ and $m = 30$. **Figure 3** shows

$$\bar{V}_{n,m}^{(k)}(\theta) / \bar{M}_{n,m}^{(k)}(\theta)^2 \text{ at } \theta = \theta^{(j)}, j = 2, 5, 8$$

computed under these settings. Although $\bar{V}_{n,30}^{(5n)}(\theta) / \bar{M}_{n,30}^{(5n)}(\theta)^2$ increases over an initial domain of n , **Figure 3** shows that the rate of such an increase is smaller as n increases. This provides a conjecture that $\bar{V}_n(\theta) / \bar{M}_n(\theta)^2$ may be bounded by some order smaller than $O(n^2)$, such as $O(n)$, for a sufficiently large n . We leave further investigation of this to future research.

5. Concluding Remarks

A main result of this paper was to show the almost sure convergence of the OPL constructed in incomplete data with two class possibilities. To obtain this result, we discussed the principle of formulating this type of structure of the OPL, and then developed the tools based on a partial-sum processes argument. The limit function of the OPL resulting finally (Corollary 1) is the essential supremum of partial likelihoods obtained based on all the forms of complete data included in incomplete data, which is similar to ℓ^∞ -norm on a Banach space. In Section 4.2, we showed numerically how an essential supremum approximates the OPL in real data for the Cox cure-mixture model.

Unfortunately, it will be difficult to show consistency and asymptotic normality of the maximum OPL estimator (MOPL) using the limit function of the OPL provided in Corollary 1. However, if the consistency is

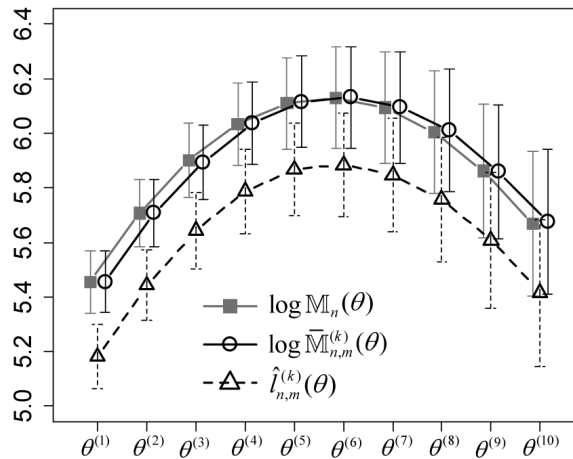


Figure 2. Plots of averages (polygonal lines) and SEs (horizontal whiskers) of $\log M_n(\theta)$, $\log(\bar{M}_{n,m}^{(k)}(\theta))$ and $\hat{I}_{n,m}^{(k)}(\theta)$ at $\theta = \theta^{(j)}$, $j = 1, \dots, 10$ obtained from 100 simulated data sets of $n = 30$ ($k = 5n$ and $m = 1000$).

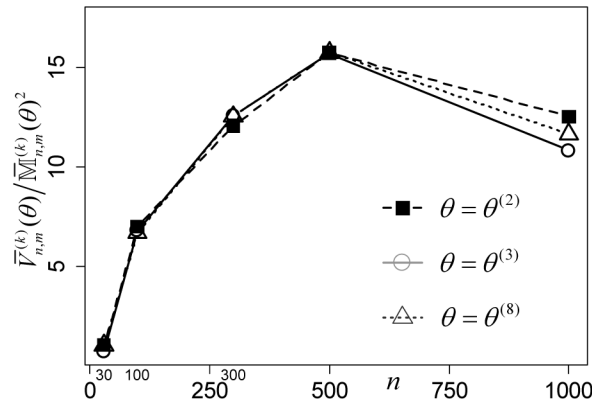


Figure 3. Plots of averages of $\bar{V}_{n,m}^{(k)}(\theta) / \bar{M}_{n,m}^{(k)}(\theta)^2$ at $\theta = \theta^{(j)}$ $j = 2, 5, 8$ obtained from 100 simulated data sets of $n = 30, 100, 300, 500$ and 1000 ($k = 5n$ and $m = 30$).

achieved (as almost expected), the global essential maximum will be accomplished around true complete data under a true regression parameter. On the other hand, for the purpose of showing the consistency of the MOPLE, there will be other convenient limit expressions, although not discussed in this paper. A future paper on this topic is based on an infinite-dimensional Laplace approximation for integral on the baseline hazard function [11]. However, in applying such a Laplace approximation to the OPL, a precondition that the OPL converges to a deterministic function is necessary. Hence, in order to obtain this precondition and for the reason that it is generally difficult to show the convergence result directly using the Laplace approximation, it is meaningful to discuss the asymptotic convergence of the OPL using the argument employed in this paper.

The results on the convergence of the exact OPL could easily suit the context of MC approximations. For example, at the end of Section 4.1 we show that, by applying Theorem 1 and Corollary 1, the size of the MC error is less than $O(n^2)/k$. This suggests that the MC method, for which the number is at most $O(n^2)$, achieves an appropriate approximation and can reduce the vast computational load of $O(2^n)$ implied by the exact method up to a feasible level. Further, in Section 4.2 we performed numerical experiments to investigate the practical size of the error in MC approximations using the Cox cure-mixture model. These experiments indicate that the exact OPL may be sufficiently approximated with the number of MC trials smaller than $O(n^2)$, such as $O(n)$, as n is larger.

In future study, it is important to derive the other expression of the limit function based on an infinite-dimensional Laplace approximation for integral on the baseline hazard and then to discuss the consistency and asymptotic normality of the MOPLE, since the asymptotic convergence of the OPL is given in this paper. Further, it is an interesting issue how the discussion of the OPL of the binary class as considered here could be extended to that under continuous class possibilities, such as the Cox frailty model.

Acknowledgements

The author is grateful to anonymous referees for their careful reading. This work is financially supported by JSPS KAKENHI grant number 23700336.

REFERENCES

- [1] D. R. Cox, "Regression Models and Life Tables (with Discussion)," *Journal of the Royal Statistical Society, Series B*, Vol. 34, No. 2, 1972, pp. 187-220.
- [2] J. S. Kim, "Maximum Likelihood Estimation for the Proportional Hazards Model with Partly Interval-Censored Data," *Journal of the Royal Statistical Society, Series B*, Vol. 65, No. 2, 2003, pp. 489-502. <http://dx.doi.org/10.1111/1467-9868.00398>
- [3] M. C. Paik and W.-Y. Tsai, "On Using the Cox Proportional Hazards Model with Missing Covariates," *Biometrika*, Vol. 84, No. 3, 1997, pp. 579-593. <http://dx.doi.org/10.1093/biomet/84.3.579>
- [4] H. Y. Chen and R. J. A. Little, "Proportional Hazards Regression with Missing Covariates," *Journal of the American Statistical Association*, Vol. 94, No. 447, 1999, pp. 896-908. <http://dx.doi.org/10.1080/01621459.1999.10474195>
- [5] A. H. Herring and J. G. Ibrahim, "Likelihood-Based Methods for Missing Covariates in the Cox Proportional Hazards Model," *Journal of the American Statistical Association*, Vol. 96, No. 453, 2001, pp. 292-302. <http://dx.doi.org/10.1198/016214501750332866>
- [6] S. A. Murphy and A. W. van der Vaart, "On Profile Likelihood (with Discussion)," *Journal of the American Statistical Association*, Vol. 95, No. 450, 2000, pp. 449-465. <http://dx.doi.org/10.1080/01621459.2000.10474219>
- [7] D. R. Cox, "Partial Likelihood," *Biometrika*, Vol. 62, No. 2, 1975, pp. 269-276. <http://dx.doi.org/10.1093/biomet/62.2.269>
- [8] R. Gill, "Marginal Partial Likelihood," *Scandinavian Journal of Statistics*, Vol. 19, No. 2, 1992, pp. 133-137.
- [9] M. R. Kosorok, "Introduction to Empirical Processes and Semiparametric Inference," Springer, Berlin, 2008. <http://dx.doi.org/10.1007/978-0-387-74978-5>
- [10] J. P. Sy and J. M. G. Taylor, "Estimation in a Cox Proportional Hazards Cure Model," *Biometrics*, Vol. 56, No. 1, 2000, pp. 227-236. <http://dx.doi.org/10.1111/j.0006-341X.2000.00227.x>
- [11] T. Sugimoto, "A Large Sample Study of Marginal Partial Likelihood in a Cox Cure-Mixture Regression Model," Unpublished.
- [12] A. Y. C. Kuk and C.-H. Chen, "A Mixture Model Combining Logistic Regression with Proportional Hazards Regression," *Biometrika*, Vol. 79, No. 3, 1992, pp. 531-541. <http://dx.doi.org/10.1093/biomet/79.3.531>
- [13] Y. Peng and K. B. G. Dear, "A Nonparametric Mixture Model for Cure Rate Estimation," *Biometrics*, Vol. 56, No. 1, 2000, pp. 237-243. <http://dx.doi.org/10.1111/j.0006-341X.2000.00237.x>

- [14] W. Lu and Z. Ying, "On Semiparametric Transformation Cure Models," *Biometrika*, Vol. 91, No. 2, 2004, pp. 331-343. <http://dx.doi.org/10.1093/biomet/91.2.331>
- [15] T. Sugimoto, T. Hamasaki and M. Goto, "Estimation from Pseudo Partial Likelihood in a Semiparametric Cure Model," *Journal of the Japanese Society of Computational Statistics*, Vol. 18, No. 1, 2005, pp. 33-46.
- [16] B. W. Turnbull, "Nonparametric Estimation of a Survivorship Function with Doubly Censored Data," *Journal of the American Statistical Association*, Vol. 69, No. 345, 1974, pp. 169-173. <http://dx.doi.org/10.1080/01621459.1974.10480146>
- [17] B. W. Turnbull, "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data," *Journal of the Royal Statistical Society, Series B*, Vol. 38, No. 3, 1976, pp. 290-295.
- [18] J. D. Kalbfleisch and R. L. Prentice, "Marginal Likelihoods Based on Cox's Regression and Life Model," *Biometrika*, Vol. 60, No. 2, 1973, pp. 267-278. <http://dx.doi.org/10.1093/biomet/60.2.267>
- [19] P. K. Andersen and R. D. Gill, "Cox's Regression Model for Counting Processes: A Large Sample Study," *Annals of Statistics*, Vol. 10, No. 3, 1982, pp. 1100-1120. <http://dx.doi.org/10.1214/aos/1176345976>
- [20] D. Collett, "Modelling Survival Data in Medical Research," 2nd Edition, Chapman & Hall/CRC, London, 2003.
- [21] A. W. van der Vaart and J. A. Wellner, "Weak Convergence and Empirical Processes," Springer-Verlag, New York, 1996. <http://dx.doi.org/10.1007/978-1-4757-2545-2>

Appendix

A.1. Proof of Lemma 3

For notational simplicity, we let $\tau = (t, \theta) \in [0, \tau_e] \times \Theta$, then write

$$\tilde{Y}_i(t; \theta) = \tilde{Y}_i(\tau) \quad \text{and} \quad n^{-1}\mathbb{T}_n(\tau, \mathbf{q}) = n^{-1}\mathbb{T}_n(\tilde{Y}(\tau), \mathbf{q}).$$

It is immediately clear that $n^{-1}\mathbb{T}_n(\tau, \mathbf{q})$ is almost surely bounded. In fact,

$$n^{-1}\mathbb{T}_n(\tau, \mathbf{q}) \leq n^{-1}\mathbb{T}_n(|\tilde{Y}(\tau)|, \mathbf{q}) \leq n^{-1} \sum_{i=1}^n |\tilde{Y}_i(\tau)| \rightarrow_{\text{as}} \mathbb{E} \left[|\tilde{Y}_i(\tau)| \right] < \infty,$$

where the last part is obvious from Conditions (i) and (ii).

Therefore, under a condition that $n^{-1}\mathbb{T}_n(\tau, \mathbf{q})$ is bounded, to prove the weak convergence of this lemma, the following results are needed (see [21]):

- **s1:** $n^{-1}\mathbb{T}_n$ is asymptotically tight,
- **s2:** The marginals $(n^{-1}\mathbb{T}_n(\tau_1, \mathbf{q}_1), \dots, n^{-1}\mathbb{T}_n(\tau_k, \mathbf{q}_k))$ converge weakly to a limit for every finite subset $(\tau_1, \mathbf{q}_1), \dots, (\tau_k, \mathbf{q}_k)$.

(Proof of s2). First, we show the pointwise version of the convergence. If (τ, \mathbf{q}) is fixed at one point on $[0, \tau_e] \times \Theta \times \{0, 1\}^\infty$, $n^{-1}\mathbb{T}_n(\tau, \mathbf{q})$ is a random variable on \mathbf{R} . Let $e(\mathbf{q}) \subset \{0, 1\}^\infty$ be an equivalent class of \mathbf{q} such that a success rate $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n q_{li}$ obtained by every $\mathbf{q}_1 \in e(\mathbf{q})$ is the same quantity as $\lim_n \sum_{i=1}^n q_i/n$ gained from \mathbf{q} (exactly as $\lim_n \sum_{i, q_i=0} q_i/n$ and $\lim_n \sum_{i, q_i=1} q_i/n$ which depend on two populations). Then, \mathbf{q} is regarded as a random element sampled from $e(\mathbf{q})$: all elements (q_1, q_2, \dots) included in \mathbf{q} ($\in e(\mathbf{q})$) are binominal random variables with the success rate $\lim_n \sum_{i=1}^n q_i/n$. Consequently, by applying the simple SLLN based on the facts that $\mathbb{E} \left[|\tilde{Y}_i(\tau) q_i| \right] < \infty$ and $\tilde{Y}_i(\tau) q_i, i=1, \dots, n$, are i.i.d. samples of two known distributions with zero means, we have $|n^{-1}\mathbb{T}_n(\tau, \mathbf{q}) - 0| \rightarrow_{\text{as}} 0$ as $n \rightarrow \infty$. From this result,

$$1 = \lim_n \Pr \left(|n^{-1}\mathbb{T}_n(\tau, \mathbf{q})| = 0 \mid e(\mathbf{q}) \right) = \lim_n \sum_{\mathbf{q}_1 \in e(\mathbf{q})} \Pr(\mathbf{q}_1 \mid e(\mathbf{q})) \Pr \left(|n^{-1}\mathbb{T}_n(\tau, \mathbf{q}_1)| = 0 \mid \mathbf{q}_1 \right)$$

is satisfied, which provides

$$\lim_{n \rightarrow \infty} \Pr \left(|n^{-1}\mathbb{T}_n(\tau, \mathbf{q})| = 0 \mid \mathbf{q} \right) = 1 \tag{A1}$$

by two properties of the probability measure, $\sum_{\mathbf{q}_1 \in e(\mathbf{q})} \Pr(\mathbf{q}_1 \mid e(\mathbf{q})) = 1$ and $\Pr(\cdot \mid \mathbf{q}_1) \leq 1$. Thus, (A.1) is the weak convergence result under (τ, \mathbf{q}) fixed at one point.

Once the result of pointwise convergence at each point is obtained, by the ε - δ method we can show how the \mathbf{R}^k -variate $n^{-1}\mathbb{T}_n$ converges on every $((\tau_1, \mathbf{q}_1), \dots, (\tau_k, \mathbf{q}_k))$ because k is finite. Using the pointwise convergence under a fixed $(\tau_l, \mathbf{q}_l) \in [0, \tau_e] \times \Theta \times \{0, 1\}^\infty$, there exists a sufficiently large M_l for $\forall \varepsilon_l > 0$ such that

$$\Pr \left(|n^{-1}\mathbb{T}_n(\tau_l, \mathbf{q}_l)| \leq \varepsilon_l \mid \mathbf{q}_l \right) \leq 1 - \varepsilon_l \quad \text{for } n \geq M_l.$$

Letting $M^k = \max(M_1, \dots, M_k)$ and $\varepsilon_+^k = \max(\varepsilon_1, \dots, \varepsilon_k)$, it is then satisfied that

$$\text{for } n \geq M^k, \Pr \left(\max_{l=1, \dots, k} |n^{-1}\mathbb{T}_n(\tau_l, \mathbf{q}_l)| \leq \varepsilon_+^k \mid \mathbf{q}_1, \dots, \mathbf{q}_k \right) \leq 1 - \varepsilon_+^k$$

on $((\tau_1, \mathbf{q}_1), \dots, (\tau_k, \mathbf{q}_k))$. The more M^k increases, the closer ε_+^k approaches to zero. Then, by tanking $M^k \rightarrow \infty$ on both sides, we obtain

$$\lim_{n \rightarrow \infty} \Pr \left(\max_{l=1, \dots, k} |\mathbb{T}_n(\tau_l, \mathbf{q}_l)|/n \leq \varepsilon_+^k \downarrow 0 \mid \mathbf{q}_1, \dots, \mathbf{q}_k \right) = 1.$$

Hence, the marginals $(|n^{-1}\mathbb{T}_n(\tau_1, \mathbf{q}_1)|, \dots, |n^{-1}\mathbb{T}_n(\tau_k, \mathbf{q}_k)|)$ converge weakly (in probability) to zeros.

(Proof of s1). The asymptotically tightness is shown by the following results (see [21]):

- **s11:** $\mathbb{T}_n(\tau, \mathbf{q})$ is asymptotically tight in \mathbf{R} for every (τ, \mathbf{q}) ,
- **s12:** \mathbb{T}_n is asymptotically uniformly ρ -equicontinuous in probability,
- **s13:** $([0, \tau_e] \times \Theta \times \{0, 1\}^\infty, \rho)$ is totally bounded.

(Proof of s11). In the proof of **s2**, we show that $\left|n^{-1}\mathbb{T}_n(\tau, \mathbf{q})/n\right|$ is almost surely bounded. Because this boundedness is equivalent to the compact property in \mathbf{R} , $\left|n^{-1}\mathbb{T}_n(\tau, \mathbf{q})\right|$ is covered almost surely by a compact set.

(Proof of s12). We define a semimetric ρ as

$$\rho_{12} = \rho((\tau_1, \mathbf{q}_1), (\tau_2, \mathbf{q}_2)) = \mathbb{E}\left[\left|\tilde{Y}_i(\tau_1) - \tilde{Y}_i(\tau_2)\right|\right] + \rho_H(\mathbf{q}_1, \mathbf{q}_2),$$

where $\rho_H(\mathbf{q}_1, \mathbf{q}_2) = \lim_n \rho_{H_n}(\mathbf{q}_1, \mathbf{q}_2) = \lim_n n^{-1} \sum_{i=1}^n |q_{1i} - q_{2i}|$ is the Hamming-type metric and we note that

$$\mathbb{E}\left[\left|\tilde{Y}_i(\tau)\right|\right] = \mathbb{E}\left[\Pr(q_i^* = 0) \mathbb{E}\left[\tilde{Y}_i(\tau) \mid q_i^* = 0\right] + \Pr(q_i^* = 1) \mathbb{E}\left[\tilde{Y}_i(\tau) \mid q_i^* = 1\right]\right].$$

By a triangle inequality, we have

$$\begin{aligned} & \sup_{\rho((\tau_1, \mathbf{q}_1), (\tau_2, \mathbf{q}_2)) < \delta} \left|n^{-1}\mathbb{T}_n(\tau_1, \mathbf{q}_1) - n^{-1}\mathbb{T}_n(\tau_2, \mathbf{q}_2)\right| \\ & \leq \sup_{\rho_{12} < \delta} \left|n^{-1}\mathbb{T}_n(\tau_1, \mathbf{q}_1) - n^{-1}\mathbb{T}_n(\tau_2, \mathbf{q}_1)\right| + \sup_{\rho_{12} < \delta} \left|n^{-1}\mathbb{T}_n(\tau_2, \mathbf{q}_1) - n^{-1}\mathbb{T}_n(\tau_2, \mathbf{q}_2)\right| \\ & = EQ_1 + EQ_2, \end{aligned}$$

where $EQ_1 = \sup_{\rho_{12} < \delta} \left|n^{-1}\mathbb{T}_n(\tilde{Y}(\tau_1) - \tilde{Y}(\tau_2), \mathbf{q}_1)\right|$ and $EQ_2 = \sup_{\rho_{12} < \delta} \left|n^{-1}\mathbb{T}_n(\tau_2, \mathbf{q}_1 - \mathbf{q}_2)\right|$. To evaluate EQ_1 , it should be noted that

$$EQ_1 \leq \sup_{\rho_{12} < \delta} n^{-1}\mathbb{T}_n\left(\left|\tilde{Y}(\tau_1) - \tilde{Y}(\tau_2)\right|, \mathbf{q}_1\right) \leq \sup_{\rho_{12} < \delta} n^{-1} \sum_{i=1}^n \left|\tilde{Y}_i(\tau_1) - \tilde{Y}_i(\tau_2)\right|$$

and $\mathbb{E}\left[\left|\tilde{Y}_i(\tau_1) - \tilde{Y}_i(\tau_2)\right|\right] < \delta$ under $\rho_{12} < \delta$. Thus, using Markov's inequality, we have

$$\Pr(EQ_1 > \varepsilon) < \varepsilon^{-1} \mathbb{E}[EQ_1] \leq \varepsilon^{-1} n^{-1} \sum_{i=1}^n \mathbb{E}\left[\sup_{\tau_1 \in K_{\tau_2}^\delta} \left|\tilde{Y}_i(\tau_1) - \tilde{Y}_i(\tau_2)\right|\right] \leq \varepsilon^{-1} \delta O(1),$$

where the final inequality follows from Condition (iii). To evaluate $EQ_2 = \sup_{\rho_{12} < \delta} \left|n^{-1} \sum_{i=1}^n \tilde{Y}_i(\tau_2)(q_{1i} - q_{2i})\right|$, let $I(\delta_n)$ be an index set that holds

$$q_{1i} - q_{2i} \neq 0 \text{ if } i \in I(\delta_n) \text{ and } q_{1i} - q_{2i} = 0 \text{ if } i \notin I(\delta_n),$$

and let $|I(\delta_n)|$ be the number of elements of $I(\delta_n)$. We have $|I(\delta_n)| \leq n\delta$, because $\rho_H(\mathbf{q}_1, \mathbf{q}_2) < \delta$ is satisfied if $\rho_{12} < \delta$. Then, using Markov's inequality, we have

$$\begin{aligned} \Pr(EQ_2 > \varepsilon) & \leq \varepsilon^{-1} \mathbb{E}[EQ_2] = \varepsilon^{-1} n^{-1} \mathbb{E}\left[\sup_{\rho_{12} < \delta} \left|\sum_{i \in I(\delta_n)} \tilde{Y}_i(\tau_2)(q_{1i} - q_{2i})\right|\right] \\ & \leq \varepsilon^{-1} n^{-1} \mathbb{E}\left[\sum_{i \in I(\delta_n)} \left|\tilde{Y}_i(\tau_2)\right|\right] \leq \varepsilon^{-1} O(|I(\delta_n)|)/n = \varepsilon^{-1} \delta O(1). \end{aligned}$$

For example, we can take $\delta < \eta\varepsilon/(1+O(1))$ independently of n , (τ_1, τ_2) and $(\mathbf{q}_1, \mathbf{q}_2)$ for every $\eta, \varepsilon > 0$. Therefore, since EQ_1 and EQ_2 are asymptotically uniformly ρ -equicontinuous in probability, \mathbb{T}_n bounded by these sums is also so.

(Proof of s13). Here we show that $\left([0, \tau_e] \times \Theta \times \{0, 1\}^\infty, \rho\right)$ is covered by a finite number of ε -balls. First, under a given $\tau \in [0, \tau_e] \times \Theta$, we show that $\left(\tau \times \{0, 1\}^\infty, \rho\right)$ is totally bounded.

Let D_q^n be the collection of all the functions such that

$$\pi(s) = q_i \text{ if } s \in ((i-1)/n, i/n], \quad i = 1, \dots, n,$$

which has either 0 or 1 at each n partition point i/n on $[0, 1]$. The space $\{0, 1\}^\infty$ is identifiable to D_q^n , which is easier to see geometrically than $\{0, 1\}^\infty$. Let Φ be the onto-mapping map $\Phi : \mathbf{q} \in \{0, 1\}^\infty \mapsto \pi \in D_q$, where D_q is D_q^n obtained as $n \rightarrow \infty$. Since $\rho_{H_n}(\mathbf{q}_1, \mathbf{q}_2)$ is identical to

$$\int_0^1 |\pi_1(x) - \pi_2(x)| d\mu(x)$$

using the Lebesgue measure μ on $[0, 1]$, $\rho_H (= \lim_{n \rightarrow \infty} \rho_{H_n})$ can be expressed as the L_1 -norm in D_q ,

$$\rho_H(\mathbf{q}_1, \mathbf{q}_2) = \int_0^1 |\pi_1(x) - \pi_2(x)| d\mu(x).$$

To consider the number of partitions of D_q with ε -band, let $g_l, l = 1, 2, \dots, 1/\varepsilon$ be representative elements such that

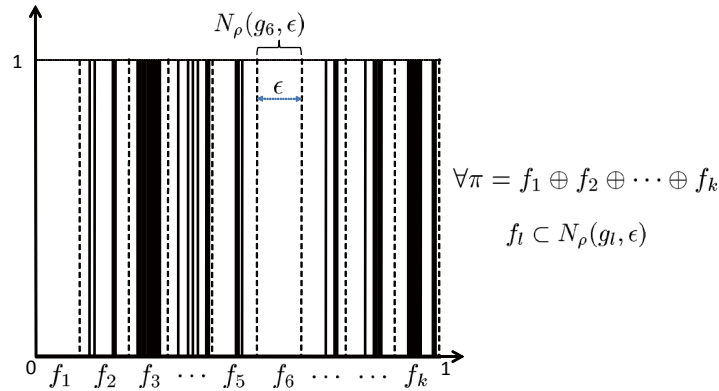


Figure 4. An illustration of the neighbourhoods $N_{\rho_H}(g_l, \varepsilon)$ and their elements f_l in D_q .

$$g_l(x) = \begin{cases} 1, & x \in ((l-1)\varepsilon, l\varepsilon] \\ 0, & x \in \text{otherwise} \end{cases}$$

All elements π of D_q are included in the combination $\bigcup_{l=1}^{1/\varepsilon} N_{\rho_H}(g_l, \varepsilon)$, where

$$N_{\rho_H}(g_l, \varepsilon) = \{f : \rho_H(f, g_l) < \varepsilon/2 \text{ and } f(x) = 0 \text{ for } x \notin ((l-1)\varepsilon, l\varepsilon]\}$$

As illustrated in **Figure 4**, for every π , we can put $\pi = f_1 \oplus f_2 \oplus \dots \oplus f_{1/\varepsilon}$ by selecting every element f_l included in a cylinder-type neighbourhood $N_{\rho_H}(g_l, \varepsilon)$ ($f_l \in N_{\rho_H}(g_l, \varepsilon)$, $l = 1, \dots, 1/\varepsilon$) and connecting such f_l 's. These facts imply

$$D_q = \Phi(\{0, 1\}^\infty) = \bigcup_{l=1}^{1/\varepsilon} N_{\rho_H}(g_l, \varepsilon) \Leftrightarrow \{0, 1\}^\infty = \bigcup_{l=1}^{1/\varepsilon} \Phi^{-1}\{N_{\rho_H}(g_l, \varepsilon)\},$$

so that $(\{0, 1\}^\infty, \rho_H)$ is totally bounded.

The remainder of the discussion is that $([0, \tau_e] \times \Theta \times \mathbf{q}, \rho) = ([0, \tau_e] \times \Theta, \rho_G)$ is totally bounded under a given $\mathbf{q} \in \{0, 1\}^\infty$, where $\rho_G(\tau_1, \tau_2) = E[\|\tilde{Y}_i(\tau_1) - \tilde{Y}_i(\tau_2)\|]$. Because $\rho_G(\tau_1, \tau_2)$ is the same as the L_1 -norm based on the probability measure, the result to be proved here is equivalent to verifying whether the class of functions $\{\tilde{Y}_i(t; \theta) : (t, \theta) \in [0, \tau_e] \times \Theta\}$ is Glivenko-Cantelli. Consequently, Condition (iii) provides that $([0, \tau_e] \times \Theta, \rho_G)$ is totally bounded.

Because the projection spaces $(\{0, 1\}^\infty, \rho_H)$ and $([0, \tau_e] \times \Theta, \rho_G)$ are totally bounded, the product space $([0, \tau_e] \times \Theta \times \{0, 1\}^\infty, \rho)$ is also so. \square

A.2. Proof of Lemma 4

(Proof of Lemma 4). Since the convergence in probability at each point is obtained similarly to Lemma 3 (proof of **s2**), a weak convergence of the arbitrary marginals on $\mathbf{q} \in \{0, 1\}^\infty$ is also shown by the ε - δ method similar to Lemma 3. Further, the asymptotic tightness is also shown similarly to the proof of **s1** in Lemma 3. In the latter, a more simple proof may be provided based on the fact that $\{0, 1\}^\infty$ is homomorphism to the Cantor set. \square