Consistency of the Semi-parametric MLE under the Cox Model with Linearly Time-dependent Covariates and Interval-censored Data

Qiqing Yu* and Qinggang Diao

Department of Mathematical Sciences, SUNY, Binghamton, NY 13902, USA
Email: qyu@math.binghamton.edu

Abstract Yu and Diao [1] studied the estimation problem under the Cox model with linearly time-dependent covariates and with interval-censored (IC) data under the distribution-free set-up. They proposed a modified semi-parametric MLE (MSMLE) and the simulation results suggest that the MSMLE is consistent. In this paper, we establish the consistency of the MSMLE.

Keywords: Cox's model, time-dependent covariates, semi-parametric MLE, consistency.

1 Introduction

In this paper, we establish the consistency of the semiparametric estimator proposed in Yu and Diao [1] under the proportional hazards (PH) model with a special continuous time-dependent covariates and with interval-censored data.

Let Y be a continuous random variable, with the cumulative distribution function (cdf) F_Y, where $F_Y(t) = P(Y \leq t)$. Its survival function is denoted by $S_Y(t) = 1 - F_Y(t)$, its density function by $f_Y(t)$, and its hazard function by $h_Y(t) = \frac{f_Y(t)}{S_Y(t-)}$. Let $\mathbf{z}$ be a $p \times 1$ covariate vector. We say that $(\mathbf{z}, Y)$ follows the PH (or Cox) model if the conditional hazard satisfies

$$h(t|\mathbf{z}) = h_{Y|\mathbf{Z}}(t|\mathbf{z}) = h_o(t)e^{\beta\mathbf{Z}}, \text{ for } t < \tau, \tag{1.1}$$

where $\beta\mathbf{z} = \beta'\mathbf{z}$, β' is the transpose of the $p \times 1$ vector β, $\tau = \sup\{t : h_o(t) > 0\}$, and h_o is the hazard function of $Y|(\mathbf{z} = 0)$. The PH model has been extended to the time-dependent covariates PH (TDCPH) model (see [2] p.113). For instance, replace $\mathbf{z}$ in (1.1) by $\mathbf{z} = \mathbf{z}(t) = \mathbf{u}\beta g(t)$, where $\mathbf{u}$ is a time-independent covariate, and $g(t)$ is a function of the time t, *e.g.*, $g(t) = (t - a)\mathbf{1}(t \geq a)$, where $\mathbf{1}(A)$ is the indicator function of an event A (see [2] p.113). We shall call the latter case the PH model with linearly time-dependent covariates (LDCPH model). Interval-censored (IC) data are (L_i, R_i), $i = 1, ..., n$, where the true survival time $Y_i \in (L_i, R_i]$. A realistic model for the IC data is the mixed case interval censorship model (see [3]). One of its special case is the case 2 interval censorship model (C2 model) (see [4]).

The TDCPH model including the LDCPH model has been commonly used for right-censored (RC) data. The semi-parametric estimation with IC data under the LDCPH model was first studied by Yu and Diao [1], that is, $h(t|z(t))$ satisfies Eq. (1.1) with

$$\mathbf{Z}(t) = \mathbf{U} * (t - a)\mathbf{1}(t \geq a), \text{ where } \mathbf{U} \text{ is a time-independent covariate vector,} \tag{1.2}$$

a is a real number and both β and S_o are unknown. As explained in their paper, this covariate $\mathbf{Z}(t)$ is very typical and share the light on how to estimate (β, h_o) under the PH model with IC data and with other types of time-dependent covariates. Under the semi-parametric set-up, the typical estimation approach for right-censored data is the partial likelihood estimation. However, it is well known that this approach only works for right-censored data, but not for IC data (see, for example, Wong and Yu [5]).

The main findings about the LDCPH model in Yu and Diao [1] are as follows. (1) Even if the parameter β is not a vector, β may not be identifiable if the support set (of the observable random vector) contains only finitely many elements. This is quite different from the case of the PH model with time-independent covariates, under which β is identifiable even if the support set contains only one point.

(2) The generalized likelihood function needs to be modified, as it must be of the form of hazard functions under the semi-parametric set-up in (1.2). Otherwise, there is no consistent estimator of β. (3) Several naive modifications on the generalized likelihood function do not lead to consistent estimators. (4) A modified semi-parametric MLE (MSMLE) was proposed and their simulation studies suggested that the MSMLE of β is consistent.

In this paper, we prove the consistency of the MSMLE. The main difficulty in the proof is that there may not exist a convergent subsequence of the estimates of h_o in (1.1), but the MSMLE is in the form of the estimate of h_o. The paper is orgainized as follows. The MSMLE is introduced in Section 2. The consistency of the MSMLE is established in Section 3. The proofs of some lemmas are put in Section 4.

2 The MSMLE

The generalized likelihood function with IC data (L_i, R_i)'s is often given by

$$\mathcal{L}_* = \prod_{i=1}^n \mu_{S(\cdot|\cdot)}(I_i|Z_i), \text{ where } \mu_{S(\cdot|\cdot)}(I_i|Z_i) = S(L_i|Z_i) - S(R_i|Z_i) \text{ and } I_i = (L_i, R_i]. \tag{2.1}$$

$\mathcal{L}_*$ depends on the survival function $S(t|z)$. Yu and Diao [1] showed that if $S(t|z)$ satisfies the PH model and is absolutely continuous, and $Z = Z(t) = (t-a)U\mathbf{1}(t \geq a)$, then

$$\begin{aligned} S(t|z(t)) &= \exp(-\int_0^t e^{\beta u(x-a)\mathbf{1}(t\geq a)} h_o(x)dx) \\ &= \begin{cases} S_o(t) & \text{if } t \leq a \text{ or } u = 0 \\ S_o(a)\exp(-\int_a^t e^{\beta u(x-a)} h_o(x)dx) & \text{if } t > a \text{ and } u \neq 0. \end{cases} \end{aligned} \tag{2.2}$$

Hereafter, abusing notations, we write $S(t|u) = S(t|z(t))$ and $h(t|u) = h_o(t)\exp(\beta u \mathbf{1}(t \geq a)(t-a))$. Notice that the two hazard functions $h_o(t) = \mathbf{1}(t > 0)$ and $h_1 = \mathbf{1}(t \in (0,2) \cup (2,\infty))$ both lead to $S_o(t) = \exp(-t)$ if $t > 0$. Since h_o (or f_o) can differ on a set A satisfying $\int_A dF_o(t) = 0$, where $F_o = 1 - S_o$, we define $f_o(t) = \begin{cases} F_o'(t) & \text{if } F_o'(t) \text{ exists} \\ 0 & otherwise \end{cases}$ for identifiability of f_o and h_o. Let $\mathcal{S}_F$ be the support set of the cdf F or a measure dF, in the sense that $x \in \mathcal{S}_F$ iff $F(x+\epsilon) - F(x-\epsilon) > 0$, $\forall\ \epsilon > 0$.

Proposition 1. (Yu and Diao [1]). The survival function $S(t|u)$ is identifiable if $t \in \mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}$. An identifiability condition for β under the mixed case IC model is that $\mathcal{S}_{F_L} \cup \mathcal{S}_{F_R}$ contains infinitely many points $\{t_j\}_{j\geq 1}$ with a limiting point, say $t_o = \lim_{j\to\infty} t_j$ in (a, ∞), provided that $F_o'(t_o) > 0$.

An interval A is called an innermost interval (II) if it is an intersection of the observed intervals I_1, ..., I_n, and if $A \cap I_i = A$ or $\emptyset$ for each I_i. It is well known (see [8]) that under the PH model with time-independent covariates in order to maximize $\mathcal{L}_*$, it suffices to put the weights of S_o to the II's. Moreover, the weight to each II is uniquely determined, but not the distribution of the weight within the II. Let A_1, ..., A_m be all the II's induced by I_i's and let (v_j, w_j) be the pair of endpoints of A_j, where $w_0 = -\infty < w_1 < w_2 < \cdots < w_m \leq \infty$. For each i, let $\xi_i = \mathbf{1}(R_i < w_m)$ and define l_i and r_i by $w_{r_i} \leq R_i < w_{r_i+1}$ and $w_{l_i} \leq L_i < w_{l_i+1}$. Then the likelihood function in (2.1) becomes

$$\begin{aligned} \mathcal{L}_1(\beta, S_o) = &\prod_{R_i\leq a \ or \ u_i=0} (S_o(w_{l_i}) - S_o(w_{r_i})) \\ &\cdot \prod_{L_i<a<R_i, u_i\neq 0} \{S_o(w_{l_i}) - \xi_i S_o(a)\exp(-\sum_{w_j\in[a,R_i]} \int_{v_j}^{w_j} e^{\beta(x-a)} h_o(x)dx)\} \cdot \prod_{L_i>a, u_i\neq 0} S_o(a) \\ &\cdot [\exp(-\sum_{w_j\in(a,L_i]} \int_{v_j}^{w_j} e^{\beta(x-a)} h_o(x)dx) - \xi_i \exp(-\sum_{w_j\in(a,R_i]} \int_{v_j}^{w_j} e^{\beta(x-a)} h_o(x)dx)]. \end{aligned} \tag{2.3}$$

Since h_o is a function of x and needs to be properly defined on $[v_j, w_j]$ for all $j < m$ (note that $S_o(w_m) = 0$), Yu and Diao proposed to modify the likelihood as follows.

First, let $h_o(x) = 0$ if $x \notin \cup_k (v_k, w_k]$, where $(v_1, w_1]$, ..., $(v_m, w_m]$ are all the II's. Moreover, notice that each $(v_k, w_k]$ will be contained by several modified observed intervals I_i's (see Remark 2 in Yu and

Diao [1]) with J (≥ 1) distinct values of u_i's, where J depends on k. There are two types of (v_k, w_k): (1) $w_k - v_k \approx 0$, or $w_k \leq a$, or $w_k - a \approx 0$ and $a \in (v_k, w_k]$; (2) otherwise. For $k < m$, define

$$h_o(x) = \begin{cases} \text{constant on } (v_k, w_k] & \text{if } (v_k, w_k) \text{ belongs to type (1)} \\ J\text{-piece-wise constant on } (v_k, w_k] & \text{if } (v_k, w_k) \text{ belongs to type (2)} \end{cases}$$

(in particular, if (v_k, w_k) belongs to type (2), then

$$h_o(x) = \sum_{j=1}^{J} h_{kj}\mathbf{1}(x \in (v_{kj}, w_{kj}]) \text{ for } x \in (v_k, w_k], \tag{2.4}$$

where $v_k = v_{k1}$, $w_{k1} = v_{k2}$, $w_{k2} = v_{k3}$, ..., $w_{kJ} = w_k$, $w_{kj} - v_{kj} = \begin{cases} \frac{w_k - v_k}{J} & \text{if } a \in (v_k, w_k],\ j \in \{1, ..., J\} \\ \frac{w_k - v_k}{J-1} & \text{if } a \notin (v_k, w_k],\ j \in \{2, ..., J\} \end{cases}$ and h_{kj}'s are constant.

If $k = m$, simply define $S_o(w_m) = 0$ (h_o can be arbitrary on $(v_m, w_m]$, provided that $h_o \geq 0$ and $\int_{v_m}^{w_m} h_o(x)dx = \infty$). Abusing notations, let $(a_j, b_j]$ be the interval in which h_o is constant, as specified in (2.4). Then $h_o(x) = \sum_j h_j \mathbf{1}(x \in (a_j, b_j])$, where h_j is a constant and $(a_j, b_j]$ may not be an II. Then $\mathcal{L}_1$ in (2.3) becomes

$$\begin{aligned} \mathcal{L}(\beta, S_o) &= \prod_{i=1}^{n}(S(L_i|U_i) - S(R_i|U_i)) \\ &= \prod_{R_i \leq a,\ or\ u_i = 0} \left\{ \exp(-\sum_{b_j \leq L_i} h_j[b_j - a_j])[1 - \exp(-\sum_{b_j \in (L_i, R_i]} h_j[b_j - a_j])]^{\xi_i} \right\} \\ &\cdot \prod_{L_i \geq a, u_i \neq 0} \left\{ S_o(a)\exp(-\frac{e^{-au_i\beta}}{u_i\beta}\sum_{b_j \in (a, L_i]} h_j[e^{u_i\beta b_j} - e^{u_i\beta a_j}]) \right. \\ &\left. \cdot [1 - \exp(-\frac{e^{-au_i\beta}}{u_i\beta}\sum_{b_j \in (L_i, R_i]} h_j[e^{u_i\beta b_j} - e^{u_i\beta a_j}])] \right\} \\ &\cdot \prod_{L_i < a < R_i, u_i \neq 0} \left\{ \exp(-\sum_{b_j \leq L_i} h_j[b_j - a_j]^{\xi_i}) \right. \\ &\left. \cdot [1 - \exp(-\sum_{b_j \in (L_i, a]} h_j(b_j - a_j) - \frac{e^{-au_i\beta}}{u_i\beta}\sum_{b_j \in (L_i, R_i]} h_j[e^{u_i\beta b_j} - e^{u_i\beta a_j}])]^{\xi_i} \right\}. \end{aligned} \tag{2.5}$$

The MSMLE maximizes $\mathcal{L}$ over all h_j's and β. It is well known that the Newton Raphson method does not work for deriving the SMLE or the MSMLE under the PH model with IC data (see Wong and Yu [5]). Yu and Diao [1] suggested to use the steepest decent method.

3 Consistency

In order to simplify the presentation of the proof of consistency, we shall only make use of the C2 model and assume that $\beta \in (-\infty, \infty)$. The C2 model assumes that C_1 and C_2 are two random follow-up times, $(L, R) = (-\infty, C_1)\mathbf{1}(Y \leq C_1) + (C_1, C_2)\mathbf{1}(Y \in (C_1, C_2]) + (C_2, \infty)\mathbf{1}(Y > C_2)$, (C_1, C_2) and (Y, U) are independent, and $P(C_1 < C_2) = 1$. Define a measure μ on the Borel σ-field $\mathcal{B}$ on R^1 by $\mu(B) = P(C_1 \in B) + P(C_2 \in B)$, $B \in \mathcal{B}$. Assume that (L_i, R_i, U_i)'s are i.i.d. from $F_{L,R,U}$ ($\stackrel{def}{=} Q$). Denote $\hat{Q} = \hat{Q}(x, y, z) = \sum_{i=1}^{n} \mathbf{1}(L_i \leq x, R_i \leq y, U_i \leq z)/n$.

Theorem 1. *Assume that the censoring satisfies the C2 model, $S(t|u)$ satisfies (2.2), the identifiable condition in Proposition 1 holds, and U takes on at least two distinct values. Then $\int |\hat{S}_n - S_o|d\mu \stackrel{a.s.}{\to} 0$ and $\hat{\beta}_n \stackrel{a.s.}{\to} \beta$, where $(\hat{S}_n, \hat{\beta}_n)$ is obtained by maximizing $\mathcal{L}$ in (2.5).*

Let Ω be the sample space. In order to prove Theorem 1, we shall make use of the next 3 lemmas.

Lemma 1. *Suppose that* $\{\mu_n\}_{n\geq 1}$ *is a sequence of measures on the measurable space* $(\mathcal{A}, \Sigma)$ *such that* $\mu_n(B) \to \mu(B),\ \forall\ B \in \Sigma$. *Let* $\{f_n\}_{n\geq 1}$ *be a sequence of non-negative integrable functions. Then* $\int_{\mathcal{A}} \varliminf_{n\to\infty} f_n\, d\mu \leq \varliminf_{n\to\infty} \int_{\mathcal{A}} f_n\, d\mu_n$.

Lemma 2. $P(\Omega_{\hat{Q}}) = 1$, *where* $\Omega_{\hat{Q}} = \{\omega \in \Omega :\ \sup_{x,y,z} |\hat{Q}(x, y, z)(\omega) - Q(x, y, z)(\omega)| \to 0\}$.

Lemma 3. *Given* $\omega \in \Omega$, *for each sub-sequence of* $\{\hat{S}_n(\cdot|\cdot)\}_{n\geq 1}$, *there exists a further subsequnce, say* $\{\hat{S}_{n_j}(\cdot|\cdot)\}_{j\geq 1}$ *and a function* $S_*(\cdot|\cdot)$ *such that* $\hat{S}_{n_j}(t|u) \to S_*(t|u)$ *for each* (t, u).

Lemma 1 is Fatou's Lemma with varying measures. It is almost the same as Proposition 17 in [7] (page 231), and so is its proof. Thus we skip the proof of Lemma 1. Lemma 2 is the multvariate version of the Glivenko - Cantelli theorem. Eddy and Hartigan [8] presented a similar version of Lemma 2 with certain addtional regularity conditions such as $P((L, R, U) \in B) = 0$, where B is the boundary of $\mathcal{S}_{F_{L,R,U}}$. However, $P(B) > 0$ under the assumptions in this paper. Thus their result is not applicable here. Lemma 3 is a variation of Helly's selection theorem. The proofs of Lemmas 2 and 3 are relegated to Section 4 for a better presentation.

Proof of Theorem 1. We shall give the proof in 3 steps.

Step 1 (preliminary). Let $S_*^{(ub)}(t) = S_*(t|u) = \exp(-\int_0^t e^{u(s-a)b\mathbf{1}(s>a)} h_*(s)ds)$, where h_* is a hazard function and let $S_* = S_*^{(0)}$ and $F_* = 1 - S_*$. By Eq. (2.5), the normalized generalized log-likelihood is $\mathcal{L}_n(S_*, b) = \frac{1}{n}\sum_{j=1}^{n} \log(S_*^{(u_j b)}(L_j) - S_*^{(u_j b)}(R_j))$. By the strong law of large numbers (SLLN),

$$\mathcal{L}_n(S_*, b) \overset{a.s.}{\to} \mathcal{L}_o(S_*, b) (\overset{def}{=} E(\mathcal{L}_n(S_*, b))).$$

Let $w_{S_*}(c_1, c_2, U) = F(c_1|U)\log F_*(c_1|U) + S(c_2|U)\log S_*(c_2|U) + (S(c_1|U) - S(c_2|U))\log(S_*(c_1|U) - S_*(c_2|U))$. Then

$$\begin{aligned}\mathcal{L}_o(S_*, b) =& E(\mathcal{L}_n(S_*, b)) = E(\log(S_*(L|U) - S_*(R|U)))\\ =& E(E(\log(S_*(L|U) - S_*(R|U)))|U)\\ =& E(E(w_{S_*}(C_1, C_2, U)|U)).\end{aligned}$$

Step 2 $\vdash$: $\mathcal{L}_o(S_*, b)$ is maximized iff $S_*(t|u) = S(t|u)$ for each $(t, u) \in \mathcal{S}_\mu \times \mathcal{S}_{F_U}$ and $b = \beta$.

It is easy to check that the expression $w_{S_*}(c_1, c_2, u)$ is maximized by a conditional survival function $S_*(\cdot|\cdot)$, if and only if $S_*(c_i|u) = S(c_i|u)$, $i \in \{1, 2\}$. Since $\sup\{|p \log p| : 0 \leq p \leq 1\} \leq 1$, $w_{S_*}(c_1, c_2, u)$ is bounded by 3, we see that $\mathcal{L}_o(S_*, b)$ is finite. Thus $S(\cdot|\cdot)$ maximizes $\mathcal{L}_o(S_*, b)$. Moreover, by Eq. (2.2), $S(t|u) = \exp(-\int_0^t e^{\beta u(x-a)\mathbf{1}(t\geq a)} h_o(x)dx)$ is continuous in (t, u), thus any other $S_*(\cdot|\cdot)$ that maximizes $\mathcal{L}_o(S_*, b)$ satisfies $S_*(\cdot|u) = S(\cdot|u)$ on $\mathcal{S}_\mu \times \mathcal{S}_{F_U}$. Furthermore, notice that one can write $S(t|u) = \exp\{-\int_0^t e^{(u-u_o)\beta(s-c)\mathbf{1}(s>c)} h_1(s)ds\}$, where $u_o \in \mathcal{S}_{F_U}$ and $h_1(s) = e^{u_o\beta(s-c)\mathbf{1}(s>c)} h_o(s)$. Without loss of generality (WLOG), we can assume that $0 \in \mathcal{S}_{F_U}$. Thus $S_*(t|0) = S(t|0)$ for $t \in \mathcal{S}_\mu$. Now, $S_*(t|u) = \exp\{-\int_0^t e^{ub(s-c)\mathbf{1}(s>c)} h_o(s)ds\}$, We shall show that $b = \beta$. Since the identifiable conditions in Proposition 1 hold, for t_o and $t_k \in \mathcal{S}_{F_{C_1}} \cup \mathcal{S}_{F_{C_2}}$ satisfying $t_k \to t_o$ and $F_o'(t_o) > 0$,

$$\begin{aligned}h_*(t_o) &= \lim_{t_k\to t_o} \frac{-\log(S_*(t_k|0)) - (-\log(S_*(t_o|0))}{t_k - t_o}\\ &= \lim_{t_k\to t_o} \frac{-\log(S_o(t_k)) - (-\log(S_o(t_o))}{t_k - t_o}\\ &= \lim_{t_k\to t_o} \frac{\int_0^{t_k} h_o(s)ds - \int_0^{t_o} h_o(s)ds}{t_k - t_o} = h_o(t_o).\end{aligned}$$

By assumption, there exists some $u \neq 0$, then

$$\begin{aligned}
e^{ub(t_o-a)\mathbf{1}(t_o\geq a)}h_*(t_o) &= \lim_{t_k\to t_o} \frac{-\log(S(t_k|0)) - (-\log(S(t_o|0))}{t_k - t_o} \\
&= \lim_{t_k\to t_o} \frac{-\log(S_o^{(ub)}(t_k)) - (-\log(S_o^{(ub)}(t_o))}{t_k - t_o} \\
&= \lim_{t_k\to t_o} \frac{\int_0^{t_k} e^{u(s-a)\beta\mathbf{1}(s>a)}h_o(s)ds - \int_0^{t_o} e^{u(s-a)\beta\mathbf{1}(s>a)}h_o(s)ds}{t_k - t_o} \\
&= e^{u(t_o-a)\beta\mathbf{1}(t_o\geq a)}h_o(t_o).
\end{aligned}$$

Thus, we must have $e^{u(t_o-a)b\mathbf{1}(t_o\geq a)}h(t_o) = e^{u(t_o-a)\beta\mathbf{1}(t_o\geq a)}h_o(t_o)$ and $h(t_o) = h_o(t_o)$, as $F_o'(t_o) > 0$. Then $h_o(t_o) > 0$, which means $b = \beta$. That is, $\mathcal{L}_o(S, b)$ is maximized by (S_o, β). This proves the claim.
Step 3 (final conclusion). Let $\Omega_o = \{\omega : \lim_{n\to\infty} \mathcal{L}_n(S_o, \beta) = \mathcal{L}_o(S_o, \beta)\} \cap \Omega_{\hat{Q}}$. It follows from the SLLN and Lemma 2 that $P(\Omega_o) = 1$. Hereafter, fix an $\omega \in \Omega_o$.

By Lemma 3, given any subsequence $\{n_i\}_{i\geq 1}$ of $\{n\}_{n\geq 1}$, there exist a futher subsequence, say itself such that $\hat{S}_{n_i}(t|u) \to S_*(t|u)$ for each (t, u). Moreover, since $\hat{\beta}_{n_i} \in (-\infty, \infty)$, there is a subsequence which converges to $b \in [-\infty, \infty]$. WLOG, we can assume that $\hat{S}_n(\cdot|\cdot) \to S_*(\cdot|\cdot)$ and $\hat{\beta}_n \to b$. Then

$$\begin{aligned}
\overline{\lim_{n\to\infty}} \mathcal{L}_n(\hat{S}_n, \hat{\beta}_n) &= \overline{\lim_{n\to\infty}} \int_\Delta \log(\hat{S}_n(l|u) - \hat{S}_n(r|u))d\hat{Q}(l, r, u) \\
&= -\underline{\lim_{n\to\infty}} \int_\Delta -\log(\hat{S}_n(l|u) - \hat{S}_n(r|u))d\hat{Q}(l, r, u)) \qquad (3.1) \\
&\leq -\int_\Delta \underline{\lim_{n\to\infty}} -\log(\hat{S}_n(l|u) - \hat{S}_n(r|u))d\hat{Q}(l, r, u) \text{ (by Lemma 1,}
\end{aligned}$$

as (1) $\hat{Q}(\eta) \to Q(\eta)$ for every Borel subset η of $\Delta = \{(l, r, u) : -\infty \leq l < r \leq \infty, u \in (-\infty, \infty)\}$ and (2) $-\log(\hat{S}_n(l|u) - \hat{S}_n(r|u)) \geq 0$). Consequently,

$$\begin{aligned}
\mathcal{L}_o(S_o, \beta) &\leq \overline{\lim_{n\to\infty}} \mathcal{L}_n(\hat{S}_n, \hat{\beta}_n) && (\text{as } \mathcal{L}_n(S_o, \beta) \leq \mathcal{L}_n(\hat{S}_n, \hat{\beta}_n)) \\
&\leq -\int_\Delta \underline{\lim_{n\to\infty}} -\log(\hat{S}_n(l|u) - \hat{S}_n(r|u))d\hat{Q}(l, r, u) && (\text{by (3.1)}) \\
&= \int_\Delta \log(S_*(l|u) - S_*(r|u))dQ(l, r, u) && (\text{as } \hat{S}_n(t|u) \to S_*(t|u)) \\
&= \mathcal{L}_o(S_*, b) \\
&\leq \mathcal{L}_o(S_o, b) && (\text{by the claim in Step 2}), \forall\, \omega \in \Omega_o.
\end{aligned}$$

Notice that $P(\Omega_o) = 1$. Thus $S_* = S_o$ a.s. μ and $b = \beta$ by the claim in Step 2. □.

4 Proofs

Proof of Lemma 2. Given $\epsilon = 1/k$, there exist x_1, ..., x_m, y_1, ..., y_m, z_1, ..., $z_m \in (-\infty, \infty]$ such that $\max\{P(L \in (x_{i-1}, x_i)), P(R \in (y_{i-1}, y_i))P(U \in (z_{i-1}, z_i))\} < \epsilon$ for $i \in \{1, ..., m\}$, where $x_0 = y_0 = z_0 = -\infty$ and $x_m = y_m = z_m = \infty$. Let Ω_k be the event that $\max_{i,j,h}\{|\hat{Q}(x_i, y_j, z_h) - Q(x_i, y_j, z_h)| + |\hat{F}_L(x_i-) - F_L(x_i-)| + |\hat{F}_R(y_j-) - F_R(y_j-)| + |\hat{F}_U(z_h-) - F_U(z_h-)|\} \to 0$. Since m is finite, $P(\Omega_k) = 1$. Let $\omega \in \Omega_k$. There exists n_o such that

$|\hat{Q}(x_i, y_j, z_h) - Q(x_i, y_j, z_h)| < \epsilon$,
$|\hat{Q}(x_i-, \infty, \infty) - Q(x_i-, \infty, \infty)| < \epsilon$,
$|\hat{Q}(\infty, y_j-, \infty) - Q(\infty, y_j-, \infty)| < \epsilon$,
$|\hat{Q}(\infty, \infty, z_h-) - Q(\infty, \infty, z_h-)| < \epsilon$, $\forall\, i, j, h \in \{0, 1, 2, ..., m\}$.

Now $\forall$ (x, y, z), $\exists$ some (i, j, h) such that $(x, y, z) \in (x_{i-1}, x_i] \times (y_{j-1}, y_j] \times (z_{h-1}, z_h]$. If $(x, y, z) = (x_i, y_j, z_h)$, then $|\hat{Q}(x, y, z) - Q(x, y, z)| < \epsilon$. Otherwise, there are 12 disjoint cases such as:

(1) $x \in (x_{i-1}, x_i)$, $y \in (y_{j-1}, y_j)$ and $z \in (z_{h-1}, z_h)$;
(2) $x \in (x_{i-1}, x_i)$, $y = y_j$ and $z \in (z_{h-1}, z_h)$;
(3) $x \in (x_{i-1}, x_i)$, $y \in (y_{j-1}, y_j)$ and $z = z_h$;
(4) $x \in (x_{i-1}, x_i)$, $y = y_j$ and $z = z_h$; etc.

In each of the 12 cases, we can also show that $|\hat{Q}(x, y, z) - Q(x, y, z)| < \epsilon$. For instance, in Case (1),

$$\begin{aligned}
&|\hat{Q}(x, y, z) - Q(x, y, z)| \\
\leq&|\hat{Q}(x_i-, y_j-, z_h-) - Q(x_{i-1}, y_{j-1}, z_{h-1})| + |Q(x_i-, y_j-, z_h-) - \hat{Q}(x_{i-1}, y_{j-1}, z_{h-1})| \\
\leq&|\hat{Q}(x_i-, y_j-, z_h-) - Q(x_i-, y_j-, z_h-) + Q(x_i-, y_j-, z_h-) - Q(x_{i-1}, y_{j-1}, z_{h-1})| \\
&+ |\hat{Q}(x_{i-1}, y_{j-1}, z_{h-1}) - Q(x_{i-1,z_{h-1}}, y_{j-1}, z_{h-1}) + Q(x_{i-1}, y_{j-1}, z_{h-1}) - Q(x_i-, y_j-, z_h-)| \\
\leq& \underbrace{|\hat{Q}(x_i-, y_j-, z_h-) - Q(x_i-, y_j-, z_h-)|}_{\leq\epsilon} + \underbrace{|Q(x_i-, y_j-, z_h-) - Q(x_{i-1}, y_{j-1}, z_{h-1})|}_{\leq(F_L(x_i-)-F_L(x_{i-1}))+(F_R(y_j-)-F_R(y_{j-1}))+(F_U(z_h-)-F_U(z_{h-1}))} \\
&+ |\hat{Q}(x_{i-1}, y_{j-1}, z_{h-1}) - Q(x_{i-1}, y_{j-1}, z_{h-1})| + |Q(x_{i-1}, y_{j-1}, z_{h-1}) - Q(x_i-, y_j-, z_j-)| \\
\leq&8\epsilon, \text{ (as } \omega \in \Omega_k).
\end{aligned}$$

In Case (4),

$$\begin{aligned}
&|\hat{Q}(x, y, z) - Q(x, y, z)| \\
\leq&|\hat{Q}(x_i-, y_j, z_h) - Q(x_{i-1}, y_j, z_h)| + |Q(x_i-, y_j, z_h) - \hat{Q}(x_{i-1}, y_j, z_h)| \\
\leq&|\hat{Q}(x_i-, y_j, z_h) - Q(x_i-, y_j, z_h) + Q(x_i-, y_j, z_h) - Q(x_{i-1}, y_j, z_h)| \\
&+ |\hat{Q}(x_{i-1}, y_j, z_h) - Q(x_{i-1,z_h}, y_j, z_h) + Q(x_{i-1}, y_j, z_h) - Q(x_i-, y_j, z_h)| \\
\leq& \underbrace{|\hat{Q}(x_i-, y_j, z_h) - Q(x_i-, y_j, z_h)|}_{\leq\epsilon} + \underbrace{|Q(x_i-, y_j, z_h) - Q(x_{i-1}, y_j, z_h)|}_{\leq(F_L(x_i-)-F_L(x_{i-1}))} \\
&+ |\hat{Q}(x_{i-1}, y_j, z_h) - Q(x_{i-1}, y_j, z_h)| + |Q(x_{i-1}, y_j, z_h) - Q(x_i-, y_j, z_j)| \\
\leq&4\epsilon.
\end{aligned}$$

The proofs for the other 10 cases are similar and are skipped. Notice $\epsilon = 1/k \to 0$, $P(\Omega_k) = 1$ and thus $P(\cap_{k=1}^{\infty}\Omega_k) = 1$. □

Proof of Lemma 3. Let $\{u_j\}_{j\geq1}$ be the collection of all rational numbers. Since $\hat{S}_n(t|u_1)$ is a sequence of bounded decreasing functions in t, by Helly's selection theorem, given any subsequence of $\{n\}_{n\geq1}$, there exists a further subsequence, say $\{n_{1i}\}_{i\geq1}$ such that $\hat{S}_{n_{1i}}(t|u_1)$ converges for each t. Moreover, for $j \geq 2$, there exists a subsequence, say $\{n_{ji}\}_{i\geq1}$ of $\{n_{(j-1)i}\}_{i\geq1}$ such that $\hat{S}_{n_{ji}}(t|u_j)$ converges for each t. It is easy to verify that $\hat{S}_{n_{ii}}(t|u_j)$ converges for each (t, u_j), say $\hat{S}_{n_{ii}}(t|u_j) \to S_1(t|u_j)$.

In view of Eq. (2.2), write $\hat{S}_n(t|u) = \hat{S}_n^{(u\hat{\beta}_n)}(t) = \exp(-\int_0^t e^{u\hat{\beta}_n(s-a)\mathbf{1}(s>a)}\hat{h}_n(s)ds)$. $\hat{S}_n(t|u)$ is a bounded decreasing function of $u\hat{\beta}_n$. Thus $S_1(t|u_j)$ is a bounded decreasing function of bu_j. Since the set of rational numbers $\{u_j\}_{j\geq1}$ is dense in $(-\infty, \infty)$, $S_1(t|u_j)$ can be extended to $S_1(t|u)$ for each $u \in (-\infty, \infty)$ by letting $S_1(t|u) = \lim_{u_j\downarrow u} S_1(t|u_j)$ if u is not a rational number. WLOG, we can assume $b \geq 0$, then $S_1(t|u)$ is a bounded decreasing function in u and is thus continuous except on a countable subset, say $\mathcal{D}$. We shall show that

$$S_{n_{ii}}(t|u) \to S_1(t|u) \ \forall\ t \text{ and } \forall\ u \notin \mathcal{D}. \qquad (A.1)$$

It suffices to consider the case that $u \notin \mathcal{D}$ and u is not a rational number. Let w and v be two rationals such that $w < u < v$. Since the set of rationals are dense, w and v can be selected as close to u as possible. Then $S_{n_{ii}}(t|v) \leq S_{n_{ii}}(t|u) \leq S_{n_{ii}}(t|w)$. Verify that $S_{n_{ii}}(t|v)^{as} \overset{i\to\infty}{\to} S_1(t|v)^{as} \overset{v\to u}{\to} S_1(t|u)$, as $x = u$ is a continuous point of $S(t|x)$. Moreover, $S_{n_{ii}}(t|w)^{as} \overset{i\to\infty}{\to} S_1(t|w)^{as} \overset{w\to u}{\to} S_1(t|u)$. By the sandwich theorem, $S_{n_{ii}}(t|u) \to S_1(t|u)$. That is, (A.1) holds.

Since $\mathcal{D}$ is countable, using the arguement similar to showing that $\hat{S}_{n_{ii}}(t|u_j)$ converges for each (t, u_j), we can show that there is a subsequence of $\{n_{ii}\}_{i\geq 1}$, say $\{m_k\}_{k\geq 1} \subset \{n_{ii}\}_{i\geq 1}$ such that $\hat{S}_{m_k}(t|u)$ converges for each $u \in \mathcal{D}$. Consequently, $\hat{S}_{m_k}(t|u)$ converges for each (t, u). □

Acknowledgments. The authors thank the editor and a referee for their invaluable comments.

References

1. Q.Q. Yu, and Q.G. Diao, "The Proportional Hazards Model with linearly time-dependent covariates and interval-censored data", *Journal of Advanced Statistics* vol. 3, no. 4, 58-70, 2018.
2. D.R. Cox, and D. Oakes, *Analysis of Survival Data,* Chapman & Hall, 1984.
3. A. Schick, and Q.Q. Yu, "Consistency of the GMLE with mixed case interval-censored data", *Scandinavian Journal of Statistics*, vol. 27, pp. 45-55, 2000.
4. P. Groeneboom, and J.A. Wellner, *Information bounds and nonparametric maximum likelihood estimation*, Birkhäuser Verlag, Basel, 1992
5. G.Y.C. Wong, and Q.Q. Yu, "Estimation under the Lehmann regression model with interval-censored data", *Communications in Statistics – Simulation and Computation*, vol. 41, 1489-1500, 2012.
6. G.Y.C. Wong, Q.G. Diao, and Q.Q. Yu, "Piecewise proportional hazards models with interval-censored data", *Journal of Statistical Computation and Simulation*, vol. 88, pp.140-155, 2018.
7. H.L. Royden, *Real Analysis*, Second edition, New York: Macmillan, 1968.
8. W. F. Eddy, and J. A. Hartigan, "Uniform Convergence of the Empirical Distribution Function Over Convex Sets", *The Annals of Statistics* vol. 5, no. 2, pp. 370-374, 1977.

