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 \( z \) be a \( p \times 1 \) covariate vector. We say that \( (z, Y) \) follows the PH (or Cox) model if the conditional hazard satisfies

\[
h(t|z) = h_Y|z(t|z) = h_0(t)e^{\beta z}, \quad \text{for } t < \tau, \tag{1.1}
\]

where \( \beta z = \beta'z, \beta' \) is the transpose of the \( p \times 1 \) vector \( \beta, \tau = \sup\{t : h_0(t) > 0\} \), and \( h_0 \) is the hazard function of \( Y|z = 0 \). The PH model has been extended to the time-dependent covariates PH (TDCPH) model (see [2] p.113). For instance, replace \( z \) in (1.1) by \( z = z(t) = u_t \beta g(t) \), where \( u \) is a time-independent covariate, and \( g(t) \) is a function of the time \( t \), e.g., \( g(t) = (t - a)1(t \geq a) \), where \( 1(A) \) 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

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

where \( a \) is a real number and both \( \beta \) and \( S_0 \) are unknown. As explained in their paper, this covariate \( Z(t) \) is very typical and share the light on how to estimate \((\beta, h_0)\) 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 \( \beta \) is not a vector, \( \beta \) 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 \( \beta \) is identifiable even if the support set contains only one point.
The generalized likelihood function with IC data \((L_i, R_i)\)'s is often given by
\[
L_* = \prod_{t=1}^{n} \mu_{S(t)}(L_i | Z_i), \quad \text{where} \quad \mu_{S(t)}(L_i | Z_i) = S(L_i | Z_i) - S(R_i | Z_i) \text{ and } L_i = (L_i, R_i). \quad (2.1)
\]

\(L_*\) depends on the survival function \(S(t)\). Yu and Diao [1] showed that if \(S(t)\) satisfies the PH model and is absolutely continuous, and \(Z = Z(t) = (t - a)U1(t \geq a)\), then
\[
S(t | z(t)) = \exp(-\int_0^t \exp(\beta u(x-a))1(t \geq a)h_o(x)dx)
\]
\[
= \begin{cases} 
S_o(t) & \text{if } t \leq a \text{ or } a = 0 \\
S_o(a) \exp(-\int_a^t \exp(\beta u(x-a))h_o(x)dx) & \text{if } t > a \text{ and } a \neq 0.
\end{cases} \quad (2.2)
\]

Hereafter, abusing notations, we write \(S(t) = S(t | z(t))\) and \(h_o(t) = h_o(t) \exp(\beta u(1(t \geq a)(t - a)).\)

Notice that the two hazard functions \(h_o(t) = 1(t \geq 0)\) and \(h_1(t \in (0, 1) \cup (2, \infty))\) 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 \(f_o(t) = \begin{cases} F_o^u(t) & \text{if } F_o^u(t) \text{ exists} \\
0 & \text{otherwise}
\end{cases}\) for identifiability of \(f_o\) and \(h_o\). Let \(S_F\) be the support set of the cdf \(F\) or a measure \(dF\), in the sense that \(x \in S_F\) iff \(F(x + \epsilon) - F(x - \epsilon) > 0, \forall \epsilon > 0\).

**Proposition 1.** (Yu and Diao [1]). The survival function \(S(t)\) is identifiable if \(t \in S_{F_L} \cup S_{R}\). An identifiability condition for \(\beta\) under the mixed case IC model is that \(S_{F_L} \cup S_{R}\) contains infinitely many points \(\{t_j\}_{j \geq 1}\) with a limiting point, say \(t_0 = \lim_{j \to \infty} t_j\) in \((a, \infty)\), provided that \(F_o(t_0) > 0\).

An interval \(A\) is called an innermost interval (II) if it is an intersection of the observed intervals \(I_1, \ldots, I_n\), and if \(A \cap I_i = A \cap I_j = 0\) for each \(I_i\). It is well known (see [8]) that under the PH model with time-independent covariates in order to maximize \(L_*\), 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 \(A_1, \ldots, A_m\) be all the II's induced by \(I_s\)'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 \(\xi_j = 1(R_i < w_{m+1})\) and define \(l_i\) and \(r_i\) by \(w_{r_i} \leq R_i < w_{r_i+1}\) and \(w_i \leq L_i < w_{i+1}\). Then the likelihood function in (2.1) becomes
\[
L_1(\beta, S_o) = \prod_{R_i \leq a \text{ or } u_i = 0} (S_o(w_{l_i}) - S_o(w_{r_i}))\]
\[
= \prod_{L_i < a < R_i, a \neq 0} \{S_o(w_{l_i}) - \xi_i S_o(a) \exp(-\sum_{w_j \in [a, R_i]} \int_{v_j}^{w_j} e^{\beta(x-a)} h_o(x)dx)\} \cdot \prod_{L_i > a, u_i \neq 0} S_o(a) \quad (2.3)
\]
\[
\cdot \left[\exp(-\sum_{w_j \in (a, R_i]} \int_{v_j}^{w_j} e^{\beta(x-a)} h_o(x)dx) - \xi_i \exp(-\sum_{w_j \in (a, R_i]} \int_{v_j}^{w_j} e^{\beta(x-a)} h_o(x)dx)\right].
\]

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 \((v_1, w_1], \ldots, (v_m, w_m]\) are all the II's. Moreover, notice that each \((v_k, w_k]\) will be contained by several modified observed intervals \(I_s\)'s (see Remark 2 in Yu and
Diao [1]) with $J \geq 1$ distinct values of $w_i$’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

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

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

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

where $v_k = v_{k1}, w_k = w_{k1}, w_{k2} = w_{k2}, \ldots, w_{kj} = w_{kj}, v_{kj} = v_{kj}$, 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 $h_o(x) = \sum h_j(x \in (a_j, b_j))$, where $h_j$ is a constant and $(a_j, b_j)$ may not be an $I$. Then $L_1$ in (2.3) becomes

$$\mathcal{L}(\beta, S_o) = \prod_{i=1}^{n} (S(L_i|U_i) - S(R_i|U_i))$$

$$= \prod_{R_i \leq a_i} \exp(- \sum_{b_j \leq L_i} h_j(b_j - a_j)) [1 - \exp(- \sum_{b_j \in (L_i, R_i]} h_j(b_j - a_j))] \prod_{L_i \geq a_i, a_i \neq 0} s_o(a) \exp(-\frac{e^{-\alpha u_i \beta}}{u_i \beta} \sum_{b_j \in (a_j, L_i]} h_j[e^{u_i \beta b_j} - e^{u_i \beta a_j}])$$

$$\times \left[1 - \exp(-\frac{e^{-\alpha u_i \beta}}{u_i \beta} \sum_{b_j \in (L_i, R_i]} h_j[e^{u_i \beta b_j} - e^{u_i \beta a_j}])\right] \prod_{L_i < a < R_i} \exp(-\sum_{b_j \leq L_i} h_j(b_j - a_j))$$

$$\times \left[1 - \exp(-\sum_{b_j \in (L_i, a_j]} h_j(b_j - a_j) - \frac{e^{-\alpha u_i \beta}}{u_i \beta} \sum_{b_j \in (L_i, R_i]} h_j[e^{u_i \beta b_j} - e^{u_i \beta a_j}])\right].$$

The MSMLE maximizes $\mathcal{L}$ over all $h_j$’s and $\beta$. 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, $(L, R) = (-\infty, C_1)1(Y \leq C_1) + (C_1, C_2)1(Y \in (C_1, C_2)) + (C_2, \infty)1(Y > C_2)$, $(C_1, C_2)$ and $(Y, U)$ are independent, and $P(C_1 < C_2) = 1$. Define a measure $\mu$ on the Borel σ-field $\mathcal{B}$ on $R^1$ by $\mu(B) = P(C_1 < B) + P(C_2 < B)$, $B \in \mathcal{B}$. Assume that $(L_i, R_i, U_i)$’s are i.i.d. from $F_{L,R,U}(\cdot \mid \hat{\mu})$. 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 $\int |\hat{S}_n - S_o| d\mu \rightarrow 0$ and $\hat{\beta}_n \xrightarrow{a.s.} \beta$, where $(\hat{S}_n, \hat{\beta}_n)$ is obtained by maximizing $\mathcal{L}$ in (2.5).

Let $\Omega$ 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 \((A, \Sigma)\) such that \( \mu_n(B) \rightarrow \mu(B), \ \forall \ B \in \Sigma \). Let \( \{f_n\}_{n \geq 1} \) be a sequence of non-negative integrable functions. Then \( \int_A \lim_{n \rightarrow \infty} f_n \, d\mu_n \leq \lim_{n \rightarrow \infty} \int_A f_n \, d\mu_n \).

Lemma 2. \( P(\Omega_Q) = 1 \), where \( \Omega_Q = \{ \omega \in \Omega : \sup_{x,y,z} |Q(x,y,z)(\omega) - Q(x,y,z)(\omega)| \rightarrow 0 \} \).

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

Let \( \int \mu \cdot d\mu \rightarrow \mu \cdot d\mu \).

However, \( S \) that maximizes \( \int c \cdot |\cdot| \rightarrow 1 \) and a function \( F \rightarrow L \) with \( L \rightarrow 0 \) 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 \( S_{\delta}^{(c_1,c_2)}(t) = S_{\delta}(t|u) = \exp(- \int_0^t e^{u(s-a)} \mathbf{1}_{c<s} h_\delta(s) \, ds) \), where \( h_\delta \) is a hazard function and let \( S_{\delta} = S_{\delta}^{(0)} \) and \( F_{\delta} = 1 - S_{\delta} \). By Eq. (2.5), the normalized generalized log-likelihood is \( \mathcal{L}_\delta(S_{\delta}, b) = \frac{1}{n} \sum_{j=1}^n \log(S_{\delta}^{(c_1,c_2)}(L_j) - S_{\delta}^{(c_2,b)}(R_j)) \). By the strong law of large numbers (SLLN),

\[
\mathcal{L}_\delta(S_{\delta}, b) \xrightarrow{a.s.} \mathcal{L}_\delta(S_{\delta}, b) (\text{def}=E(\mathcal{L}_\delta(S_{\delta}, b))).
\]

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

\[
\mathcal{L}_\delta(S_{\delta}, b) = E(\mathcal{L}_\delta(S_{\delta}, b)) = E(\log(S_{\delta}(L|U) - S_{\delta}(R|U)))
\]

\[
= E(\mathcal{L}_\delta(S_{\delta}(L|U) - S_{\delta}(R|U)))
\]

\[
= E(E(w_\delta(C_1,C_2,U)|U)).
\]

Step 2 \( \implies \mathcal{L}_\delta(S_{\delta}, b) \) is maximized iff \( S_{\delta}(t|u) = S(t|u) \) for each \( (t,u) \in S_\mu \times S_{F_\delta} \) and \( b = \beta \).

It is easy to check that the expression \( w_\delta(c_1,c_2,u) \) is maximized by a conditional survival function \( S_{\delta}(\cdot|\cdot) \), if and only if \( S_{\delta}(c_i|u) = S(c_i|u), \ \text{for} \ i \in \{1,2\} \). Since \( \sup\{|p \log p| : 0 \leq p \leq 1\} \leq 1 \), \( w_\delta(c_1,c_2,u) \) is bounded by 3, we see that \( \mathcal{L}_\delta(S_{\delta}, b) \) is finite. Thus \( \mathcal{L}(\cdot|\cdot) \) maximizes \( \mathcal{L}_\delta(S_{\delta}, b) \). Moreover, by Eq. (2.2), \( S(t|u) = \exp(- \int_0^t e^{u(s-a)} \mathbf{1}_{c<s} h_\delta(s) \, ds) \) is continuous in \( (t,u) \), thus any other \( S_{\delta}(\cdot|\cdot) \) that maximizes \( \mathcal{L}_\delta(S_{\delta}, b) \) satisfies \( S_{\delta}(\cdot|u) = \mathcal{L}(\cdot|u) \) on \( S_\mu \times S_{F_\delta} \). Furthermore, notice that one can write \( S(t|u) = \exp(- \int_0^t e^{u(s-a)} \mathbf{1}_{c<s} h_\delta(s) \, ds) \), where \( u_0 \in S_{F_\delta} \) and \( h_1(s) = e^{u_0 \beta(s-c)} \mathbf{1}_{c<s} h_0(s) \). Without loss of generality (WLOG), we can assume that \( 0 \in S_{F_\delta} \). Thus \( S_{\delta}(t|0) = S(t|0) \) for \( t \in S_\mu \). Now, \( S_{\delta}(t|u) = \exp(- \int_0^t e^{u(s-a)} \mathbf{1}_{c<s} h_\delta(s) \, ds) \). We shall show that \( b = \beta \). Since the identifiable conditions in Proposition 1 hold, for \( t_o \) and \( k \in S_{F_{c_1}} \cup S_{F_{c_2}} \) satisfying \( k \rightarrow t_o \) and \( F_{\delta}(k) > 0 \),

\[
h_\delta(t_o) = \lim_{t_k \rightarrow t_o} \frac{-\log(S_{\delta}(t_k|0)) - (-\log(S_{\delta}(t_o|0)))}{t_k - t_o}
\]

\[
= \lim_{t_k \rightarrow t_o} \frac{-\log(S_{\delta}(t_k)) - (-\log(S_{\delta}(t_o)))}{t_k - t_o}
\]

\[
= \lim_{t_k \rightarrow t_o} \frac{\int_{t_o}^{t_k} h_\delta(s) \, ds - \int_{t_o}^{t_k} h_\delta(s) \, ds}{t_k - t_o} = h_\delta(t_o).
\]
By assumption, there exists some \( u \neq 0 \), then
\[
\lim_{t_k \to t_o} e^{u(t_o-a)} 1_{(t_o \geq a)} h_u(t_o) = \lim_{t_k \to t_o} \frac{-\log(S(t_k))-(-\log(S(t_o)))}{t_k - t_o} = \lim_{t_k \to t_o} \frac{-\log(S^{(ub)}(t_k))-(-\log(S^{(ub)}(t_o)))}{t_k - t_o} = \lim_{t_k \to t_o} \int_0^{t_k} e^{u(s-a)} 1_{(s>\alpha)} h_u(s) ds - \int_0^{t_o} e^{u(s-a)} 1_{(s>\alpha)} h_u(s) ds
\]
Thus, we must have \( e^{u(t_o-a)} 1_{(t_o \geq a)} h_u(t_o) = e^{u(t_o-a)} 1_{(t_o \geq a)} h_o(t_o) \) and \( b(t_o) = h_o(t_o) \), as \( F'_0(t_o) > 0 \).
Then \( h_o(t_o) > 0 \), which means \( b = \beta \). That is, \( L_o(S, b) \) is maximized by \( (S_o, \beta) \). This proves the claim.

**Step 3** (final conclusion). Let \( \Omega_o = \{ \omega : \lim_{n \to \infty} L_n(S_o, \beta) = L_o(S_o, \beta) \} \cap \Omega_Q \). 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 further subsequence, say itself such that \( \hat{S}_{n_i}(l(w)) \to S_{\omega}(l(w)) \) for each \( (l, w) \). Moreover, since \( \beta_{n_i} \in (-\infty, \infty) \), there is a subsequence which converges to \( b \in [-\infty, \infty] \). WLOG, we can assume that \( \hat{S}_{n_i}(|\cdot|) \to S_{\omega}(|\cdot|) \) and \( \beta_{n_i} \to b \). Then
\[
\lim_{n \to \infty} L_n(\hat{S}_{n_i}, \hat{\beta}_{n_i}) = \lim_{n \to \infty} \int_{\Delta} \log(\hat{S}_{n_i}(l|u|) - \hat{S}_{n}(r|u|))d\hat{Q}(l, r, u)
\]

(3.1)
\[
\leq \lim_{n \to \infty} \int_{\Delta} \log(\hat{S}_{n_i}(l|u|) - \hat{S}_{n}(r|u|))d\hat{Q}(l, r, u) \quad \text{(by Lemma 1,)}
\]
\[
\text{as (1) } \hat{Q}(\eta) \to Q(\eta) \text{ for every Borel subset } \eta \text{ of } \Delta = \{(l, r, u) : -\infty \leq l < r \leq \infty, u \in (-\infty, \infty)\} \text{ and (2)}
\]
\[-\log(\hat{S}_{n_i}(l|u|) - \hat{S}_{n}(r|u|)) \geq 0 \]. Consequently,
\[
L_o(S_o, \beta) \leq \lim_{n \to \infty} L_n(\hat{S}_{n_i}, \hat{\beta}_{n_i}) \quad \text{(as } L_n(S_o, \beta) \leq L_n(\hat{S}_{n}, \hat{\beta}_{n_i}))\]
\[
\leq \int_{\Delta} \log(S_{\omega}(l|u|) - S_{\omega}(r|u|))dQ(l, r, u) \quad \text{(as } \hat{S}_{n_i}(t|u|) \to S_{\omega}(t|u)|)}
\]
\[
= L_o(S_{\omega}, b) \quad \text{(by the claim in Step 2)}
\]
\[
\forall \omega \in \Omega_o
\]
Notice that \( P(\Omega_o) = 1 \). Thus \( S_{\omega} = S_o \) a.s. \( \mu \) and \( b = \beta \) by the claim in Step 2. \( \Box \)

4 Proofs

**Proof of Lemma 2.** Given \( \epsilon = 1/k \), there exist \( x_1, \ldots, x_m, y_1, \ldots, y_m, z_1, \ldots, z_m \in (-\infty, \infty] \) such that \( \max\left( P(L \in (x_{i-1}, x_i)) \right) < \epsilon \) for \( i \in \{1, \ldots, m\} \), where \( x_0 = y_0 = z_0 = -\infty \) and \( x_m = y_m = z_m = \infty \). Let \( \Omega_k \) be the event that \( |\hat{Q}(x_i, y_j, z_h)| + |\hat{F}_L(x_i -) - F_L(x_i -)| + |\hat{F}_R(y_j -) - F_R(y_j -)| + |\hat{F}_U(z_h -) - F_U(z_h -)| \to 0 \). Since \( m \) is finite, \( P(\Omega_k) = 1 \). Let \( \omega \in \Omega_k \). There exists \( n_o \) such that
\[
|\hat{Q}(x_i, y_j, z_h) - Q(x_i, y_j, z_h)| < \epsilon, \]
\[
|\hat{Q}(x_i, y_j, z_h) - Q(x_i, y_j, z_h)| < \epsilon, \]
\[
Q(\infty, y_j, -\infty) - Q(\infty, y_j, -\infty) < \epsilon, \]
\[
|\hat{Q}(\infty, y_j, -\infty) - Q(\infty, y_j, -\infty)| < \epsilon, \]
\[
|Q(\infty, y_j, -\infty) - Q(\infty, y_j, -\infty)| < \epsilon, \]
\[
Q(\infty, y_j, -\infty) - Q(\infty, y_j, -\infty)| < \epsilon, \]
\[
\forall i, j, h \in \{0, 1, 2, \ldots, m\}. \]
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 \(\bar{Q}(x, y, z) = Q(x, y, z) < \epsilon\). 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 \(\bar{Q}(x, y, z) - Q(x, y, z) < \epsilon\). For instance, in Case (1),

\[
\bar{Q}(x, y, z) - Q(x, y, z) \\
\leq |Q(x_i - x_{i-1}, y_j - y_{j-1}, z_h - z_{h-1}) - Q(x_i - x_{i-1}, y_{j-1} - y_j, z_h - z_{h-1})| \\
\leq |Q(x_i - x_{i-1}, y_j - y_{j-1}, z_h - z_{h-1}) - Q(x_i - x_{i-1}, y_{j-1} - y_j, z_{h-1} - z_h) + Q(x_i - x_{i-1}, y_{j-1} - y_j, z_{h-1})| \\
\leq |Q(x_i - x_{i-1}, y_{j-1}, z_{h-1}) - Q(x_i - x_{i-1}, y_{j-1}, z_{h-1}) + Q(x_i - x_{i-1}, y_{j-1}, z_{h-1})| \\
\leq \epsilon.
\]

In Case (4),

\[
\bar{Q}(x, y, z) - Q(x, y, z) \\
\leq |Q(x_i - x_{i-1}, y_j - y_{j-1}, z_h) - Q(x_i - x_{i-1}, y_{j-1} - y_j, z_h)| \\
\leq |Q(x_i - x_{i-1}, y_j - y_{j-1}, z_h) - Q(x_i - x_{i-1}, y_j - y_{j-1}, z_{h-1}) + Q(x_i - x_{i-1}, y_j - y_{j-1}, z_{h-1})| \\
\leq |Q(x_i - x_{i-1}, y_j, z_{h-1}) - Q(x_i - x_{i-1}, y_{j-1}, z_{h-1}) + Q(x_i - x_{i-1}, y_{j-1}, z_{h-1})| \\
\leq \epsilon.
\]

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 \(\{u_j\}_{j \geq 1}\) be the collection of all rational numbers. Since \(\hat{S}_n(t|u_j)\) is a sequence of bounded decreasing functions in \(t\), by Helly’s selection theorem, given any subsequence of \(\{n_i\}_{i \geq 1}\), there exists a further subsequence, say \(\{n_{i_1}\}_{i_1 \geq 1}\) such that \(\hat{S}_{n_{i_1}}(t|u_j)\) converges for each \(t\). Moreover, for \(j \geq 2\), there exists a subsequence, say \(\{n_{j_1}\}_{j_1 \geq 1}\) of \(\{n_{(j-1)1}\}_{j \geq 1}\) such that \(\hat{S}_{n_{j_1}}(t|u_j)\) converges for each \(t\). It is easy to verify that \(\hat{S}_{n_{i_1}}(t|u_j)\) converges for each \((t, u_j)\), say \(\hat{S}_{n_{i_1}}(t|u_j) \to \hat{S}_1(t|u_j)\).

In view of Eq. (2.2), write \(\hat{S}_n(t|u) = \hat{S}_{n}^{(u\beta_u|n)}(t) = \exp(-\int_0^t e^{u\beta_u(s-u)}1^{(s-u)}X(s)ds)\). \(\hat{S}_n(t|u)\) is a bounded decreasing function of \(u\beta_u\). Thus \(\hat{S}_1(t|u_j)\) is a bounded decreasing function of \(bu_j\). Since the set of rational numbers \(\{u_j\}_{j \geq 1}\) is dense in \((-\infty, \infty)\), \(\hat{S}_1(t|u_j)\) can be extended to \(\hat{S}_1(t|u)\) for each \(u \in (-\infty, \infty)\) by letting \(\hat{S}_1(t|u) = \lim_{i \to \infty} \hat{S}_1(t|u_j)\) if \(u\) is not a rational number. WLOG, we can assume \(b \geq 0\), then \(\hat{S}_1(t|u)\) is a bounded decreasing function in \(u\) and is thus continuous except on a countable subset, say \(D\). We shall show that

\[
\hat{S}_{n_{i_1}}(t|u) \to \hat{S}_1(t|u) \forall t \in \mathbb{R} \ \forall u \notin D.
\]
Since $\mathcal{D}$ is countable, using the argument similar to showing that $\hat{S}_{n_i}(t|u_j)$ converges for each $(t, u_j)$, we can show that there is a subsequence of $\{n_i\}_{i \geq 1}$, say $\{m_k\}_{k \geq 1} \subset \{n_i\}_{i \geq 1}$ 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.