Supplement to "Estimation in the Single Change-point Hazard Function for interval-censored Data with a Cure Fraction"

Bing Wang, Xiaoguang Wang and Lixin Song

ARTICLE HISTORY
Compiled March 7, 2019

This document provides detailed proofs for the asymptotic properties given in Section 4. Firstly, we demonstrate the proof of Theorem 1.

Proof of Theorem 1. Define $Z_n(t) = \sum_{i=1}^n I(R_i \geq t)$. When $t < \tau_H$, $\frac{Z_n(t)}{n} \to 1 - H(\infty, t)$ almost surely by the strong law of large numbers, and $1 - H(\infty, t) > 0$. Hence, $Z_n(t) \to \infty$, which implies $Y_{(n)} \to \tau_H$. In addition, we have that $F(\tau_H) \leq p < 1$. By the conclusion of [1], we obtain

$$\sup_{t \in R} |\hat{F}_n(t) - F(t)| \to 0$$

in probability. Since $Y_{(n)} \leq \tau_H$ almost surely, it follows that $|\hat{F}_n(Y_{(n)}) - F(Y_{(n)})| \to 0$ in probability. When $\tau_H < \infty$, $Y_{(n)} \to \tau_H$ in probability and F is continuous at τ_H. Thus

$$\hat{F}_n(Y_{(n)}) = F(Y_{(n)}) + o_p(1) \to F(\tau_H) = pF_0(\tau_H)$$

in probability. When $\tau_H = \infty$, $Y_{(n)} \to \infty$ in probability, so $\hat{F}_n(Y_{(n)}) \to p = pF_0(\tau_H)$ in probability, again. Note that

$$\tau_{F_0} = \sup\{t : F_0(t) < 1\} = \sup\{t : F(t) < p\}.$$

Hence $\hat{F}_n(Y_{(n)}) \to p$ if and only if $F_0(\tau_H) = 1$, that is $\tau_{F_0} \leq \tau_H$, and then the theorem follows. $\square$

Proof of Theorem 2. Define

$$X_n(t) = \left\{ \frac{\hat{\Lambda}_n^*(D) - \hat{\Lambda}_n^*(t)}{D - t} - \frac{\hat{\Lambda}_n^*(t) - \hat{\Lambda}_n^*(0)}{t} \right\} g\{t(D - t)\}, \ 0 < t < D,$$

and

$$X_n^0(t) = \left\{ \frac{\hat{\Lambda}_n^{*0}(D) - \hat{\Lambda}_n^{*0}(t)}{D - t} - \frac{\hat{\Lambda}_n^{*0}(t) - \hat{\Lambda}_n^{*0}(0)}{t} \right\} g\{t(D - t)\}, \ 0 < t < D,$$

where $\hat{\Lambda}_n^*(t) = -\log[-\frac{1}{\hat{p}}\{\exp(-\hat{\Lambda}_n(t)) - 1 + \hat{p}\}]$, $\hat{\Lambda}_n(t)$ is obtained by $\hat{F}_n$, and $\hat{\Lambda}_n^{*0}(t) = -\log[-\frac{1}{p}\{\exp(-\hat{\Lambda}_n) - 1 + p\}]$. Notice that $\Lambda^*(0) = \hat{\Lambda}_n^{*0}(0) = \Lambda(0) = 0$. Then $X(t) = t^{q-1}(D-t)^{q-1}\{t\Lambda^*(D) - D\Lambda^*(t)\}$. For any $\varepsilon > 0$, let $c_1 \in (0, \min\{X(\tau) - X(\tau - \varepsilon), X(\tau) - X(\tau+\varepsilon)\})$ relying on ε, τ_1, τ_2, p, θ. Then, if $|t-\tau| > \varepsilon$, we have $X(\tau) - X(\tau+\varepsilon) > c_1$. Noting that $X_n(t)$ attains its maximum at $\hat{\tau}_n$, for sufficiently large n, we have

$$\begin{aligned}
& P(|\hat{\tau} - \tau| > \varepsilon) \\
\leq\ & P(X(\tau) - X(\hat{\tau}) > c_1) \\
\leq\ & P(X(\tau) - X(\hat{\tau}) + X_n(\hat{\tau}) - X_n(\tau) > c_1) \\
\leq\ & P(|X_n(\hat{\tau}) - X(\hat{\tau})| + |X(\tau) - X_n(\tau)| > c_1) \\
=\ & P(|X_n(\hat{\tau}) - X(\hat{\tau})| + |X(\tau) - X_n(\tau)| > c_1, \sup_{\tau_1<t<\tau_2} |X_n(t) - X(t)| > \frac{c_1}{2}) \\
& + P(|X_n(\hat{\tau}) - X(\hat{\tau})| + |X(\tau) - X_n(\tau)| > c_1, \sup_{\tau_1<t<\tau_2} |X_n(t) - X(t)| \leq \frac{c_1}{2}) \\
\leq\ & P(\sup_{\tau_1<t<\tau_2} |X_n(t) - X(t)| > \frac{c_1}{2}) + P(\emptyset) \\
\leq\ & P(\sup_{\tau_1<t<\tau_2} |X_n(t) - X_n^0(t)| > \frac{c_1}{4}) + P(\sup_{\tau_1<t<\tau_2} |X_n^0(t) - X(t)| > \frac{c_1}{4}) \\
\leq\ & P(D\tau_1^{q-1}(D-\tau_2)^{q-1} \sup_{\tau_1<t<\tau_2} |U_n^0(t)| + \tau_2^q(D-\tau_2)^{q-1}U_n^0(D) > \frac{c_1}{4}) \\
& + P(\sup_{\tau_1<t<\tau_2} |X_n(t) - X_n^0(t)| > \frac{c_1}{4}).
\end{aligned}$$

We can obtain the last inequality by

$$\begin{aligned}
X_n^0(t) - X(t) &= t^{q-1}(D-t)^{q-1}[t\{\hat{\Lambda}_n^{*0}(D) - \Lambda_n^*(D)\} - D\{\hat{\Lambda}_n^{*0}(D) - \Lambda_n^*(D)\}] \\
&= t^{q-1}(D-t)^{q-1}U_n^0(D) - t^{q-1}(D-t)^{q-1}U_n^0(t),
\end{aligned}$$

where $U_n^0 = \hat{\Lambda}_0^*(t) - \Lambda^*(t)$. Consequently, there exist $c_2 > 0$ and $c_3 > 0$, depending on c_1, τ_1, τ_2, D and q, such that

$$\begin{aligned}
& P(|\hat{\tau} - \tau| > \varepsilon) \\
\leq\ & P(\sup_{\tau_1<t<\tau_2} |U_n^0(t)| > c_2) + P(U_n^0(D) > c_3) + P(\sup_{\tau_1<t<\tau_2} |X_n(t) - X_n^0(t)| > \frac{c_1}{4}) \\
=\ & I_1 + I_2 + I_3. \qquad (1)
\end{aligned}$$

From the definition of $\Lambda^*(t)$, we find that

$$\begin{aligned}
|U_n^0(t)| &= |\log(-\frac{1}{p}\{e^{-\hat{\Lambda}_n(t)} - 1 + p\}) - \log(-\frac{1}{p}\{e^{-\Lambda(t)} - 1 + p\})| \\
&= \left| \frac{e^{-\alpha(t)}}{e^{-\alpha(t)} - 1 + p} \right| \cdot |\hat{\Lambda}_n(t) - \Lambda(t)|, \qquad (2)
\end{aligned}$$

where $\alpha(t)$ is between $\hat{\Lambda}(t)$ and $\Lambda(t)$. Thus, $\exp(-\alpha(t))$ lies on the segment between $\hat{S}(t) = 1 - \hat{F}_n(t)$ and $S(t) = 1 - F(t) = 1 - pF_0(t)$. For interval-censored data, according to [1], $\sup_{t\in[0,\tau_{F_0}]} |\hat{F}_n(t) - F(t)| \to 0$ almost surely for $\tau_{F_0} \leq \tau_G$. Thus for any $\alpha < 1 - pF_0(D)$,

$$\exp(-\alpha(t)) > [1 - F(D)] - \alpha = [1 - pF_0(D)] - \alpha = \phi(D),$$

provided that $\tau_{F_0} > D$. It follows (2) that

$$|U_n^0(t)| \leq \frac{1}{\phi(D) - 1 + p}|\hat{\Lambda}_n(t) - \Lambda(t)| = \frac{1}{\phi(D) - 1 + p}|U_n(t)|.$$

By the assumption

$$h(l, r) > 0, \text{ if } 0 < F_0(l) < F_0(r) < 1. \tag{3}$$

and the likelihood function $L(\beta, \theta, p, \tau|\boldsymbol{O}_i,\ i = 1, \ldots, n)$, there exists $c_4 > 0$ relying on $c_1, c_2, \tau_1, \tau_2, D, q, p$ and F_0, satisfying

$$\begin{aligned} I_1 &\leq P(\sup_{\tau_1 < t < \tau_2} |U_n(t)| > c_4, \tau_2 \leq Y_{(n)}) + P(Y_{(n)} < \tau_2) \\ &\leq P(\sup_{\tau_1 < t < \tau_2} |U_n(t \wedge Y_{(n)})| > c_4) + \prod_{i=1}^{n} P(R_i < \tau_2). \end{aligned} \tag{4}$$

We know that

$$\hat{\Lambda}_n(t \wedge Y_{(n)}) - \Lambda(t \wedge Y_{(n)}) = \log \frac{1 - \hat{F}_n(t \wedge Y_{(n)})}{1 - F(t \wedge Y_{(n)})}. \tag{5}$$

By $P(\lim_{n\to\infty} \sup_{t\in\mathbb{R}} |\hat{F}_n(t) - F(t)| = 0) = 1$ obtained by [1], the first term on the right side of last inequality of (4) converges to 0 as $n \to \infty$. Next, by (1), $I_2 \leq P(|U_n(D)| > c_3, Y_{(n)} \geq D) + P(Y_{(n)} < D)$. Similarly, I_2 converges to 0 as $n \to \infty$.

In order to prove $I_3 \to 0$, we rewrite $X_n(t)$ and $X_n^0(t)$ as

$$X_n(t) = t^{q-1}(D - t)^{q-1}[t\{\hat{\Lambda}_n^*(t) - \hat{\Lambda}_n^*(t)\} - (D - t)\hat{\Lambda}_n^*(t)], \tag{6}$$

and

$$X_n^0(t) = t^{q-1}(D - t)^{q-1}[t\{\hat{\Lambda}_n^{*0}(D) - \hat{\Lambda}_n^{*0}(t)\} - (D - t)\hat{\Lambda}_n^{*0}(t)]. \tag{7}$$

By (4) and (5),

$$\begin{aligned} I_3 &\leq P(\sup_{\tau_1 < t < \tau_2} |X_n(t) - X_n^0(t)| > \frac{c_1}{4}, Y_{(n)} \geq D) + P(Y_{(n)} < D) \\ &\leq P(2 \sup_{\tau_1 < t < \tau_2} |\hat{\Lambda}_n^{*0}(t) - \hat{\Lambda}_n^{*0}(t)|\tau_2^q(D - \tau_2)^{q-1} > \frac{c_4}{8}) \\ &\quad + P(\sup_{\tau_1 < t < \tau_2} |\hat{\Lambda}_n^*(t) - \hat{\Lambda}_n^{*0}(t)|\tau_2^{q-1}(D - \tau_2)^q > \frac{c_4}{8}) + P(Y_{(n)} < D) \\ &= I_{31} + I_{32} + \prod_{i=1}^{n} P(R_i < D). \end{aligned}$$

We can see that

$$I_{31} \leq P(|\log(\hat{p}) - \log(p)| + \sup_{\tau_1 < t < \tau_2} \left| \log \frac{e^{\hat{\Lambda}_n(t)} - 1 + \hat{p}}{e^{\hat{\Lambda}_n(t)} - 1 + p} \right| > \frac{c_4}{8}).$$

Since $\hat{p}$ converges to p in probability, and $\sup_{0<t<D} |\hat{\Lambda}_n(t) - \Lambda(t)| \to 0$, we have $I_{31} \to 0$. Similarly, $I_{32} \to 0$. This completes the proof of Theorem 2. □

In order to show the proof of Theorems 3-5, we need some lemmas. We first state some conditions from [2].

Condition 1. $\sqrt{n} P_n \dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_0, \boldsymbol{\nu}_0) = O_{p^*}(1)$.
For i.i.d. observations, Condition 2 holds automatically if $P\dot{l}^2_{\boldsymbol{\mu}}(\boldsymbol{\mu}_0, \boldsymbol{\nu}_0) < \infty$ by the central limit theorem.

Condition 2.

$$\frac{|\sqrt{n}(P_n - P)\dot{l}_{\boldsymbol{\mu}}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\nu}}) - \sqrt{n}(P_n - P)\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_0, \boldsymbol{\nu}_0)|}{1 + \sqrt{n}|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0|} = o_{p^*}(1),$$

where $|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0| = o_{p^*}(1)$ and $|\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_0| = o_{p^*}(1)$.

Condition 3. $\sqrt{n} P\ddot{l}_{\boldsymbol{\mu}\boldsymbol{\nu}}(\boldsymbol{\mu}_0, \boldsymbol{\nu}_0)|\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_0| = O_p(1)$.
When $\hat{\boldsymbol{\nu}}$ is a $\sqrt{n}$-consistent, this condition holds automatically.

Condition 4. (Smoothness Condition) For $(\boldsymbol{\mu}, \boldsymbol{\nu}) \in D_n$,

$$|P\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\nu}) - P\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_0, \boldsymbol{\nu}_0) - P\ddot{l}_{\boldsymbol{\mu}\boldsymbol{\mu}}(\boldsymbol{\mu}_0, \boldsymbol{\nu}_0)(\boldsymbol{\mu} - \boldsymbol{\mu}_0) - P\ddot{l}_{\boldsymbol{\mu}\boldsymbol{\nu}}(\boldsymbol{\mu}_0, \boldsymbol{\nu}_0)(\boldsymbol{\nu} - \boldsymbol{\nu}_0)| = o(|\boldsymbol{\mu} - \boldsymbol{\mu}_0|) + o(|\boldsymbol{\nu} - \boldsymbol{\nu}_0|),$$

where $D_n = \{(\boldsymbol{\mu}, \boldsymbol{\nu}) : |\boldsymbol{\mu} - \boldsymbol{\mu}_0| \leq \eta_n \downarrow 0, |\boldsymbol{\nu} - \boldsymbol{\nu}_0| \leq cn^{1/2}\}$ for some constant c.

Condition 5. Under the probability P,

$$\sqrt{n} \begin{bmatrix} (P_n - P)\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_0, \boldsymbol{\nu}_0) \\ \hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_0 \end{bmatrix} \xrightarrow{d} \boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda_1} \\ \boldsymbol{\Lambda_2} \end{bmatrix}, \tag{8}$$

where $\boldsymbol{\Lambda} \sim N_4(0, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ being a 4×4 positive definite matrix.

The following Lemmas 1-4 are due to [2], which also correspond to Theorem 6.1 in [3] for the semi-parametric model with a infinite-dimensional parameter space.

Lemma 1. *Suppose that $\boldsymbol{\mu}_0$ is the unique solution to $P\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\nu}_0) = 0$ and $\hat{\boldsymbol{\nu}}$ such that $|\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}_0| = o_{p^*}(1)$. If*

$$\sup_{\boldsymbol{\mu} \in \Theta_1, |\boldsymbol{\nu} - \boldsymbol{\nu}_0| \leq \eta_n} \frac{|P_n \dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\nu}) - P\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\nu}_0)|}{1 + |P_n \dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\nu})| + |P\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\nu}_0)|} = o_{p^*}(1)$$

for every sequence $\{\eta_n\} \downarrow 0$, then $\hat{\boldsymbol{\mu}}$ almost surely solving $P_n \dot{l}_{\boldsymbol{\mu}}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\nu}}) = o_{p^*}(1)$ converges in outer probability to $\boldsymbol{\mu}_0$.

Lemma 2. *Suppose that the class of functions* $\{\psi(\boldsymbol{\mu},\boldsymbol{\nu}) : |\boldsymbol{\mu}-\boldsymbol{\mu}_0| < \gamma, |\boldsymbol{\nu}-\boldsymbol{\nu}_0| < \gamma\}$ *is P-Donsker for some* $\gamma > 0$*, and that* $P|\psi(\boldsymbol{\mu},\boldsymbol{\nu}|X)-\psi(\boldsymbol{\mu}_0,\boldsymbol{\nu}_0|X)|^2 \to 0$*, as* $|\boldsymbol{\mu}-\boldsymbol{\mu}_0| \to 0$ *and* $|\boldsymbol{\nu}-\boldsymbol{\nu}_0| \to 0$*. If* $\hat{\boldsymbol{\mu}} \xrightarrow{p^*} \boldsymbol{\mu}_0$ *and* $\hat{\boldsymbol{\nu}} \xrightarrow{p^*} \boldsymbol{\nu}_0$*, then*

$$|\sqrt{n}(P_n - P)(\psi(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\nu}}) - \psi(\boldsymbol{\mu}_0,\boldsymbol{\nu}_0))| = o_{p^*}(1).$$

We should note that the conditions of Lemma 2 imply Condition 2. However, there is a set of simple sufficient conditions for Condition 2, thus we will verify the conditions of Lemma 2 in the proof of Theorem 5 below.

Lemma 3. *Suppose that* $\hat{\boldsymbol{\mu}}$ *satisfies* $P_n\dot{l}_{\boldsymbol{\mu}}(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\nu}}) = o_{p^*}(n^{-1/2})$ *and is a consistent estimator of* $\boldsymbol{\mu}$*, which is the unique point for which* $P\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu}_0) = 0$*, and* $\hat{\boldsymbol{\nu}}$ *is an estimator of* $\boldsymbol{\nu}_0$ *satisfying* $|\hat{\boldsymbol{\nu}}-\boldsymbol{\nu}_0| = O_{p^*}(n^{-1/2})$*. Then under Conditions 1-4,* $\sqrt{n}(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}_0) = O_{p^*}(1)$*.*

Lemma 4. *Suppose that* $\boldsymbol{\mu}_0$ *is the unique solution to* $P\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu}_0) = 0$ *and* $\hat{\boldsymbol{\nu}}$ *is an estimator of* $\boldsymbol{\nu}_0$ *satisfying* $|\hat{\boldsymbol{\nu}}-\boldsymbol{\nu}_0| = O_{p^*}(1)$*. Then under Conditions 2-5,* $\sqrt{n}(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}_0) \xrightarrow{d} \{-P\ddot{l}_{\boldsymbol{\mu}\boldsymbol{\mu}}(\boldsymbol{\mu}_0,\boldsymbol{\nu}_0)\}^{-1}N_4(0,\boldsymbol{V})$*, where* $\boldsymbol{V} = Var(\Lambda_1 + P\ddot{l}_{\boldsymbol{\mu}\boldsymbol{\nu}}(\boldsymbol{\mu}_0,\boldsymbol{\nu}_0)\Lambda_2)$*.*

Lemma 5. *For* $\dot{l}_\beta(\boldsymbol{\mu},\boldsymbol{\nu}|\boldsymbol{O})$ *and* $\dot{l}_\theta(\boldsymbol{\mu},\boldsymbol{\nu}|\boldsymbol{O})$ *defined in (12) and (13), if* $|\boldsymbol{\mu}-\boldsymbol{\mu}_0| \le \eta_n \downarrow 0$ *and* $|\boldsymbol{\nu}-\boldsymbol{\nu}_0| \le cn^{-1/2}$*, then* $P(\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu}) - \dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_0,\boldsymbol{\nu}_0))^2 = o_{p^*}(1)$*.*

Proof. See Lemma 5 of [5].

Proof of Theorem 3. To prove the consistency of the pseudo estimator $\hat{\boldsymbol{\mu}}$, we mainly need

$$\sup_{\boldsymbol{\mu}\in\boldsymbol{\Theta}_1, |\boldsymbol{\nu}-\boldsymbol{\nu}_0|\le\eta_n} |P_n\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu}) - P\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu}_0)| = o_{p*}(1)$$

for every sequence $\{\eta_n\} \downarrow 0$. Then the consistency of $\hat{\boldsymbol{\mu}}$ follows from Lemma 1. Since

$$|P_n\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu}) - P\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu}_0)| \le |(P_n - P)\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu})| + |P(\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu}) - \dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu}_0))|,$$

and $P\ddot{l}_{\boldsymbol{\mu}\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu}|\boldsymbol{O})$ obviously tends to zero when $|\boldsymbol{\nu}-\boldsymbol{\nu}_0| \le \alpha \downarrow 0$. We need to show that the class of the functions $F_\alpha \equiv \{\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu}) : \boldsymbol{\mu} \in \boldsymbol{\Theta}_1 \subset \mathbb{R}^2, |\boldsymbol{\nu}-\boldsymbol{\nu}_0| \le \eta_n\}$ is a VC-class for some $\eta_n > 0$, where $\boldsymbol{\Theta}_1 = \{\boldsymbol{\mu} = (\beta,\theta)' : \beta \ge A_1, \theta \ge A_2\}$. This implies that the uniform strong law of large numbers holds, i.e., $\sup_{f\in F_n}(P_n - P)f \xrightarrow{p} 0$; See [4], Chap. 2.6-2.7, for details. Let $F_{1\alpha} = \{I_{(-\infty,-\tau]}(R) : |\tau - \tau_0| \le \alpha_1\}$, and $F_{2\alpha} = \{I_{(-\infty,-\tau]}(L) : |\tau - \tau_0| \le \alpha_1\}$. Then the VC-indexes of the class of functions $F_{1\alpha}$ and $F_{2\alpha}$ are both 2 by Example 2.6.1 of [4]. Thus the class of functions

$$\{I_{(-\infty,\pi]}(L)I_{(\pi,\infty)}(R)\frac{Re^{-\beta R-\theta(R-\tau)} - Le^{-\beta L}}{e^{-\beta L} - e^{-\beta R-\theta(R-\tau)}} : \boldsymbol{\mu} \in \boldsymbol{\Theta}_1 \subset \mathbb{R}^2, |\boldsymbol{\nu}-\boldsymbol{\nu}_0| \le \alpha\}$$

is Donsker by Lemma 2.6.18 and Example 2.10.8 of [4], because $(Re^{-\beta R-\theta(R-\tau)} - Le^{-\beta L})/(e^{-\beta L} - e^{-\beta R-\theta(R-\tau)})$ is bounded. It is similar to show that the other classes of functions are also Donsker. Thus the class of functions of F_α is VC-class by applying Example 2.10.7 and Theorem 2.10.6 of [4]. Finally, by Lemma 1, $\hat{\boldsymbol{\mu}}$ is consistent. □

Proof of Theorem 4. We first verify the stochastic equicontinuity condition:

$$|\sqrt{n}(P_n - P)\{\dot{l}_{\boldsymbol{\mu}}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\nu}}) - \dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_0, \boldsymbol{\nu}_0)\}| = o_{p^*}(1). \tag{9}$$

Let $F_\gamma = \{\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\nu}) - \dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_0, \boldsymbol{\nu}_0) : |\boldsymbol{\mu} - \boldsymbol{\mu}_0| \leq \gamma, |\boldsymbol{\nu} - \boldsymbol{\nu}_0| \leq \gamma\}$. Similar to the proof of Theorem 1 we can show that F is a VC-class. Thus (9) follows from Lemma 2 together with Lemma 5. Next, the smoothness Condition 4 holds by $P\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\nu}|\boldsymbol{O}) < \infty$, $P\ddot{l}_{\boldsymbol{\mu}\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\nu}|\boldsymbol{O}) < \infty$, and Lemma 5, and $P_n\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_0, \boldsymbol{\nu}_0)$ converges in distribution to a normal random variable by the central limit theorem. Thus $\sqrt{n}|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}| = O_{p^*}(1)$ by Lemma 3. $\square$

Proof of Theorem 5. By the consistency of $\hat{p}$ and $\hat{\tau}$ together with the Slutsky's theorem and the central limit theorem, we can show that (8) holds for normally distributed Λ_1 with mean zero and positive variance. Hence by Lemma 4, $\sqrt{n}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)$ is asymptotically normal with mean 0 and variance $\{P\ddot{l}_{\boldsymbol{\mu}\boldsymbol{\mu}}(\boldsymbol{\mu}_0, \boldsymbol{\nu}_0)\}^{-2}\boldsymbol{V}$. $\square$

References

[1] P. Groeneboom and J. Wellner, *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Oberwolfach Seminars, Birkhäuser Basel, 1992.

[2] H. Hu, *Large sample theory for pseudo-maximum likelihood estimates in semiparametric models*, Ph.D. diss., University of Washington, 1998.

[3] J. Huang, *Efficient estimation for the proportional hazards model with interval censoring*, The Annals of Statistics 24 (1996), pp. 540–568.

[4] A.W. Van Der Vaart and J.A. Wellner, *Weak Convergence and Empirical Processes*, Springer, 1996.

[5] X. Zhao, X. Wu, and X. Zhou, *A change-point model for survival data with long-term survivors*, Statistica Sinica 19 (2009), pp. 377–390.