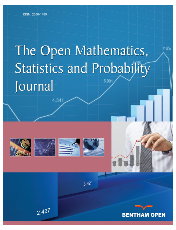
2666-1489/20



The Open Mathematics, Statistics and Probability Journal

Content list available at: https://openstatisticsandprobabilityjournal.com

RESEARCH ARTICLE

Consistency of the Semi-parametric MLE under the Cox Model with Right-Censored Data

Qiqing Yu[1,*]

[1] *Department of Mathematical Sciences, SUNY, Binghamton, NY, 13902, USA*

Abstract:

Objective:

We studied the consistency of the semi-parametric maximum likelihood estimator (SMLE) under the Cox regression model with right-censored (RC) data.

Methods:

Consistency proofs of the MLE are often based on the Shannon-Kolmogorov inequality, which requires finite E(lnL), where L is the likelihood function.

Results:

The results of this study show that one property of the semi-parametric MLE (SMLE) is established.

Conclusion:

Under the Cox model with RC data, E(lnL) may not exist. We used the Kullback-Leibler information inequality in our proof.

Keywords: Cox model, Maximum likelihood estimator, Consistency, Kullback-Leibler Inequality, Shannon-Kolmogorov inequality, Without loss of generality (WLOG).

Article History	Received: May 14, 2020	Revised: August 14, 2020	Accepted: August 19, 2020

1. INTRODUCTION

We studied the consistency of the semi-parametric maximum likelihood estimator (SMLE) under the Cox model with right-censored (RC) data.

Let Y be a random survival time, $\mathbf{X}$ a p-dimensional random covariate. Conditional on $\mathbf{X} = \mathbf{x}$, Y satisfies the Cox model if its hazard function satisfies

$$h(y|\mathbf{x}) = h_o(y)\, e^{\beta'\mathbf{x}}, \quad y < \tau_Y, \qquad \textbf{(1.1)}$$

where h_o is the baseline hazard function, *i.e.*, $h_o(y) = f_o(y) / S_o(y\text{-})$, f_o is a density function, S_o (y) = $S(y|0) \stackrel{\text{def}}{=} P(Y > y\,|\mathbf{X} =$), $F_o = 1 - S_o$, $\tau_Y = sup\{t{:}S_Y(t) > 0\}$, $h(y|\mathbf{x}) = \frac{f\,(y|\mathbf{x})}{S(y-|\mathbf{x})}$, $S(\cdot|\cdot)$ $f(\cdot|\cdot)$ **or**$F(\cdot|\cdot)$) is the conditional survival function (density function (df) or cumulative distribution function (cdf)) of Y given $\mathbf{X} = \mathbf{x}$. The restriction $y<\tau_Y$ is not in the original definition of the PH model, but is necessary if S_o is discontinuous at τ_Y (see Remark 1 [1])

* Address correspondence to this author at the Department of Mathematical Sciences, SUNY, Binghamton, NY, 13902, USA;
E-mail: qyu@math.binghamton.edu

2. METHODS

In this paper, we shall make use of the assumptions as follows:

AS1. Suppose that C is a random variable with the df $f_C(t)$ and the survival function $S_C(t)$, $\mathbf{X}$ takes at least p +1 values, say $\mathbf{0}$, $\mathbf{x}_1$, ..., $\mathbf{x}_p$, where $\mathbf{x}_1$, ..., $\mathbf{x}_p$ are linearly independent, (Y,$\mathbf{X}$) and C are independent. Let (Y_1,$\mathbf{X}_1$,C_1), ..., (Y_n,$\mathbf{X}_n$,C_n) be i.i.d. random vectors from (Y,$\mathbf{X}$,C). $M = min(Y,C)$ and $\delta = \mathbf{1}(Y \le C)$, where $\mathbf{1}(A)$ is the indicator function of the event A. Let (M_1, $\delta_1\mathbf{X}_1$), ..., (M_n, δ_n, $\mathbf{X}_n$) be i.i.d. RC observations from (M, δ, $\mathbf{X}$) with the df are as follows:

$$f_{M,\delta,X}(m,\delta,\mathbf{x}) = \big(S(m|\mathbf{x})\mathrm{f_c(m)}\big)^{1-\delta}\big(f(m|\mathbf{x})\mathrm{S_c(m)}\big)^{\delta} f_{\mathbf{X}}(\mathbf{x}), \quad \text{where } m \in \mathrm{D} \qquad \textbf{(1.2)}$$

$$\mathrm{D} = \begin{cases} (-\infty,\tau_M] & if\ P(Y=\tau_M|X=0)=0\ or\ P(C\ge\tau_M)>0 \\ (-\infty,\tau_M) & \text{otherwise,} \end{cases}$$

$$\tau_M = \sup\{x{:}\ S_M(x) > 0\}, S_0(\tau_M) < 1,$$

and $S(t|\mathbf{x})$ is a function of (S_o, β) (see Eq. (1.1)), but not $f_{\mathbf{x}}$ and f_C (the df's of $\mathbf{X}$ and C).

Due to (AS1) and Eq. (1.2), the generalized likelihood function can be written as:

$$L(S_0, \beta) = \prod_{i=1}^{n}[(S(M_i|X_i))^{(1-\delta_i)}(S(M_i-|X_i) - S(M_i|X_i))^{\delta_i}], \quad \textbf{(1.3)}$$

which coincides with the standard form of the generalized likelihood [2]. Eq. (1.3) is identical to the next expression:

$$L(S_0, \beta) = \prod_{i=1}^{n}[(S(M_i|X_i))^{1-\delta_i}(S(M_i-\eta_n|X_i) - S(M_i|X_i))^{\delta_i}] \quad \textbf{(1.4)}$$

where $\eta_n = \min\{|M_i-M_j|: M_i \neq M_j, i, j \in \{1,2, ..., n\}\}$. This form allows S_o to be arbitrary (discrete or continuous, or others), thus is more convenient in the later proofs. If Y is continuous then $S(t|\mathbf{x}) = (S(t|0))^{\exp(\mathbf{X}'\beta)} = (S_o(t))^{\exp(\mathbf{X}'\beta)}$, but

$S(t|\mathbf{x}) \neq (S(t|0))^{\exp(\mathbf{x}'\beta)}$ under the discrete Cox model ([3]). **(1.5)**

If Y is discrete then $S(t|\mathbf{x}) = \prod_{s\leq t}(1 - h(s|\mathbf{x})) = \prod_{s\leq t}(1 - h_0(s)e^{\mathbf{X}'\beta})$ If Y has a mixture distribution, then $S(t|\mathbf{x}) = p(S_{01}(t))^{\exp(\mathbf{X}'\beta)} + (1-p)\prod_{s\leq t}(1 - h_{02}(s)e^{\mathbf{X}'\beta}$ where $p \in (0,1)$, h_{01} and h_{02} are two hazard functions. $h_0(t) = ph_{01} + (1-p)h_{02}$ and $S_0(t) = pS_{01} + (1-p)S_{02}$

The SMLE of (S_o, β) maximizes L $(S, \mathbf{b})$ overall possible survival function S and $\mathbf{b} \in \boldsymbol{R}^p$, denoted by $(\hat{S}_o, \hat{\beta})$. The SMLE of $S(t|\mathbf{x})$ is denoted by $\hat{S}$ $(t|\mathbf{x})$, which is a function of $(\hat{S}_o, \hat{\beta})$. The computation issue of the SMLE under the Cox model has been studied, but its consistency has not been established under the model [3]. Their simulation results suggest that the SMLE is more efficient than the partial likelihood estimator under the Cox model.

The partial likelihood estimator is a common estimator under the Cox model, which maximizes the partial likelihood:

$L_0 = \prod_{i\in D}\frac{\exp(\beta'\mathbf{X}_i)}{\sum_{k\in\boldsymbol{R}_i}\exp(\beta'\mathbf{X}_k)}$, where D is the collection of indices of the exact observations and $\boldsymbol{R}_i$ is the risk set {j: $M_j \geq Y_j$}. The asymptotic properties of the estimator are well understood [4].

The consistency of the SMLE under the continuous Cox model with interval-censored (IC) data has been established, making use of the following result [5]:

The Shannon-Kolmogorov (S-K) inequality. *Let f_o and f be two densities with respect to (w.r.t.)* a measure μ and $\int f_0(t)\ln f_0(t)d\mu(t)$ is finite. Then, $\int f_0(t)\ln f_0(t)d\mu(t) \geq \int f_0(t)\ln f(t)d\mu(t)$, with equality *iff* $f = f_o$ a.e. w.r.t. μ.

Under the Cox model with IC data, the S-K inequality becomes $E(\ln L(S_o, \beta)) \geq E(\ln L(S, \mathbf{b}))$ $\forall$ $(S, \mathbf{b})$, where $L(\cdot,\cdot)$ is the likelihood function of the Cox model with IC data, which is different from L $(\cdot,\cdot)$ in Eq. (1.3) and S is a baseline survival function and $\mathbf{b} \in \boldsymbol{R}^p$. Their approach cannot be extended to the Cox model with RC data as the key assumption (in the S-K inequality) [3].

That is, finite $E(\ln L(S_o, \beta))$, may not hold. Indeed, if Y has a df $f_0(t) \propto \frac{\mathbf{1}(x\in\{2,3,4...\})}{x(\ln x)^2}$, $\delta_i \equiv 1$ and $\beta = 0$, then L

$$(S_0, \beta) = \prod_{i=1}^{n}(f_0(Y_i) \text{ and } E(\ln L)) = \sum_x f_0(x)\ln f_0(x) \propto \int_{x\geq 2}\frac{\ln x + 2\ln\ln x}{x(\ln x)^2} = -\infty.$$

A related inequality is as follows.

The Kullback-Leibler (K-L) information inequality. *Let f_o and f be two densities w.r.t. a measure μ. Then $\int f_0(t)\ln(f_0/f)(t)d\mu(t) \geq 0$, with equality iff $f = f_o$ a.e. w.r.t. μ.*

The K-L inequality says that $\int f_0(t)\ln(f_0/f)(t)d\mu(t)$ exists, though it maybe ∞. The two inequalities are not equivalent. In fact,

$$\int f_0(t)\ln f_0(t)d\mu(t) \geq 0 \text{ if } \int f_0(t)\ln f_0(t)d\mu(t) \geq \int f_0(t)\ln f(t)d\mu(t).$$

In this note, we show that the SMLE under the Cox model is consistent, making use of the Kullback-Leibler information inequality [6]

2. The Main Results. Notice that under the assumption that h_o exists, S_o, f_o, F_o and h_o are equivalent, in the sense that given one of them, the other 3 functions can be derived. Thus, the Cox model is applicable only to the distributions that the density functions exist, that is, Y is either continuous, or discrete, or the mixture of the previous two. Since the expression of $S(t|\mathbf{x})$ varies in these three cases, for simplicity, we only prove the consistency of the SMLE under the Cox model in the first two cases.

Theorem 1. *Under the Cox model with RC data, if Y is either continuous or discrete, and if $S_o(\tau_M) < 1$, then the SMLE* ($(\hat{S}_o(t), \hat{\beta})$) is consistent $\forall$ $t \in$ D (see Eq. (1.2)).

The proof of Theorem 1 makes use of a modified K-L inequality. K-L inequality requires that f_0 and f are both densities w.r.t. the measure μ. That is $\int f(t)d\mu(t) = 1$. However, in our case, we encounter the case that $\int f(t)d\mu(t) \in [0,1]$.

Lemma 1 (the modified K-L inequality). If $f_i \geq 0$, μ_1 is a measure, $\int f_1(t)d\mu_1(t) = 1$ and $\int f_2(t)d\mu_1(t) \leq 1$, then $\int f_1(t)\ln\frac{f_1(t)}{f_2(t)}d\mu_1(t) \geq 0$, with equality iff $f_1 = f_2$ a.e. w.r.t. μ_1.

Proof. In view of the K-L inequality, it suffices to prove the inequality $\int f_1(t)\ln\frac{f_1(t)}{f_2(t)}d\mu_1(t) \geq 0$ under the additional assumptions that $\int f_2(t)d\mu_1(t) < 1$, $\int f_1(t)d\mu_2(t) = 0$ and $\int f_2(t)d\mu(t) < 1$, where μ_2 is a measure and $\mu = \mu_1 + \mu_2$ Since $\int f_2(t)d\mu(t) = 1$, f_1 and f_2 are df's w.r.t. μ.

$$0 \leq \int f_1(t)\ln\frac{f_1(t)}{f_2(t)}d(\mu_1(t) + \mu_2(t)) \quad \text{(by the K-L inequality)}$$

$$= \int f_1(t)\ln\frac{f_1(t)}{f_2(t)}d\mu_1(t) + \int f_1(t)\ln\frac{f_1(t)}{f_2(t)}d\mu_2(t) = \int f_1(t)\ln\frac{f_1(t)}{f_2(t)}d\mu_1(t).$$

Proof of Theorem 1. Let Ω_0 be the subset of the sample space Ω such that the empirical distribution function (edf) $\hat{F}_n$, $(t, s, \boldsymbol{x})$ based on $(M_i, \delta_i, \boldsymbol{X}_i)$ converges to $F(t,s,\mathbf{x})$, the cdf of $(M, \delta, \mathbf{X})$. It is well-known that $P(\Omega,)$ =1. Notice that the SMLE ($\hat{S}_o, \hat{\beta}$) is a function of (ω, n), say ($\hat{S}_{o,n}(t)(\omega)$, $\hat{\beta}_{o,n}(t_n)(\omega)$, where $\omega \in \Omega$ and n is the sample size. Hereafter, fix an $\omega \in \Omega_0$, since $\hat{\beta}(=\hat{\beta}_n(\omega))$ is a sequence of vectors in $\boldsymbol{R}^p$, there is a convergent subsequence with the limit β^*, where the components of β_* can be $\pm\infty$. Moreover, S_o (= $S_{o,n}(\cdot)(\omega)$) is a sequence of bounded non-increasing functions, Helly's selection theorem ensures that given any subsequence of $\hat{S}_o$, there exists a further subsequence which is convergent. Without loss of generality (WLOG), we assume that $\hat{S}_o \to S_*$ and $\hat{\beta} \to \beta_*$. Of course, (β_*, S_*) depends on $\omega(\in \Omega_0)$. We prove in Theorem 2 for the discrete case and in Theorem 3 for the continuous case that:

$$(S_*(t), \beta_*) = (S_*(t), \beta_*)\ (\omega) = (S_0(t), \beta)\ \forall\ t \in \mathrm{D} \quad \textbf{(2.1)}$$

Since ω can be arbitrary in Ω_0 and $\mathrm{P}(\Omega_0) = 1$, the SMLE is consistent.

Before we prove Theorems 2 and 3, we present a preliminary result.

Lemma 2 (Proposition 17 in Royden (1968), page 231). *Suppose that* μ_n *is a sequence of measures on the measurable space* $(J, \mathscr{B})$ *such that* $\mu_n(B)\ \mu(B), \to B \in \mathscr{B}$, g_n and f_n *are non-negative* measurable functions, and $\lim_{n\to\infty} (f_n, g_n)(x) = (f, g)(x)$ *Then*,

(1) $\int f d\mu \leq \underline{lim}_{n\to\infty} \int f_n\, d\mu_n$;

(2) *if* $g_n \geq f_n\ (\geq 0)$ *and* $\lim_{n\to\infty} \int g_n d\mu_n = \int g d\mu$, *then* $\int f d\mu = \lim_{n\to\infty} \int f_n d\mu_n$.

Corollary 1. *Suppose that* μ_n *is a sequence of measures on the measurable space (J , B) such that* $\lim_{n\to\infty} \mu_n(B) \to \mu(B), \forall B \in \mathscr{B}$, f *and* f_n $(n \geq 1)$ *are integrable functions that are bounded below and* $f(x)_{n\to\infty} = \lim f_n(x)$. *Then* $\int f\, d\mu \leq \underline{\lim}_{n\to\infty} \int f_n\, d\mu_n$.

Proof. Let $k = \inf_n \inf f_n(x)$. If $k \geq 0$ then the corollary follows from Lemma 2. Otherwise, let $f_n^-(x) = 0 \wedge f_n(x)$, $f_n^+(x) = 0 \vee f_n(x)$, $f^-(x) = 0 \wedge f(x)$ and $f^+(x) = 0 \vee f(x)$. Then, $f_n^+ \to f^+$ and $f_n^- \to f^-$ point wisely, as, $f_n \to f$

$\underline{\lim}_{n\to\infty} \int f_n\, d\mu_n = \underline{\lim}_{n\to\infty} \int (f_n^+ + f_n^-) d\mu_n = \underline{\lim}_{n\to\infty} [\int f_n^+\, d\mu_n + f_n^-\, d\mu_n] \geq \int \underline{\lim}_{n\to\infty} f_n^+\, d\mu + \int \underline{\lim}_{n\to\infty} f_n^-\, d\mu$ (by Lemma2, as $f_n^+(x)$ is nonnegative and $|f^-(x)| \leq k$) $= \int f^+\, d\mu + \int f^-\, d\mu = \int (f^+ + f^-) d\mu = \int f\, d\mu$.

Theorem 2. Under the discrete Cox model with RC data, Eq. (2.1) holds.

Proof. For the given $\omega \in \Omega_0$ and (S_*, β_*) in the proof of Theorem 1, as assumed, $(\hat{S}_o, \hat{\beta})(\omega) \to (S_*, \beta_*)$. Defining $h_*(t) = \frac{S_*(t-) - S_*(t)}{S_*(t-)}$ and $h_*(t|\boldsymbol{x}) = h_*(t)^{e^{\beta_* x}}$ (for $S_*(t-) > 0$) yeilds $S_*(t|\boldsymbol{x})$ and $f_*(t|\boldsymbol{x})$, which are continuous functions of S_* and β_*. Consequently, $\hat{S}(\cdot|\cdot) \to S_*(\cdot|\cdot)$.

Let $G_n(S_0, \beta) = \ln L(S_0, \beta)/n$ (see Eq.(1.3)). Then, the SMLE $(\hat{S}_o, \hat{\beta})$ satisfies

$$\begin{aligned} G_n(\hat{S}o, \hat{\beta}) &= \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i) ln\hat{S}(M_i|\mathbf{X}_i) + \frac{1}{n}\sum_{i=1}^{n}\delta_i \ln(\hat{S}(M_i - |\mathbf{X}_i) - \hat{S}(M_i|\mathbf{X}_i)) \\ &= \int ln\hat{S}(t|\mathbf{x}) d\hat{F}_n(t,0,\mathbf{x}) + \int \ln(\hat{S}(t-|\mathbf{x}) - \hat{S}(t|\mathbf{x})) d\hat{F}_n(t,1,\mathbf{x}) \geq G_n(\mathrm{S}_0, \beta). \end{aligned} \quad \textbf{(2.2)}$$

$$=> 0 \geq \int ln \frac{s(t|\mathbf{X})}{\hat{s}(t|\mathbf{X})}\, d\hat{F}_n(t,0,\mathbf{x}) + \int ln \frac{s(t-|\mathbf{X}) - s(t|\mathbf{X})}{\hat{s}(t-|\mathbf{X}) - \hat{s}(t|\mathbf{X})}\, d\hat{F}_n(t,1,\mathbf{x}).$$

Let $\mu_n(B) = \int_B \frac{\hat{s}(t|\mathbf{X})}{s(t|\mathbf{X})} d\hat{F}_n(t,0,\mathbf{x})$, .

where B is a measurable set in $\boldsymbol{R}^{p+1}$. To apply Lemma 2,

$$\text{Let } K(t,0,\mathbf{x}) \stackrel{\text{def}}{=} \frac{1}{s(t|\mathbf{X})} \left(\geq \frac{\hat{s}(t|\mathbf{X})}{s(t|\mathbf{X})}\right), \text{ then} \quad \textbf{(2.3)}$$

$$\begin{aligned} \int K(t,0,\mathbf{x}) d\hat{F}_n(t,0,\mathbf{x}) &= \int \frac{1}{s(t|\mathbf{X})} d\hat{F}_n(t,0,\mathbf{x}) \\ &\to \int \frac{1}{s(t|\mathbf{X})} dF(t,0,\mathbf{x}) \quad (\text{as } \omega \in \Omega_0) \\ &= \int \frac{1}{s(t|\mathbf{X})} S(t|\mathbf{x}) dF_C(t)\, dF_{\mathbf{X}}(\mathbf{x}) \quad (by(1.2)); \end{aligned} \quad \textbf{(2.4)}$$

$$\underline{lim}_{n\to\infty}\mu_n(B) = \underline{lim}_{n\to\infty}\int_B \frac{\hat{S}(t|\mathbf{X})}{S(t|\mathbf{X})} d\hat{F}_n(t,0,\mathbf{x}) \quad \textbf{(2.5)}$$

$$= \int_B \underline{lim}_{n\to\infty}\frac{\hat{S}(t|\mathbf{X})}{S(t|\mathbf{X})} dF(t,0,\mathbf{x}) \text{ (by statement (2) of Lemma 2, (2.3)and (2.4))}$$

$$= \int_B \frac{S_*(t|\mathbf{X})}{S(t|\mathbf{X})} dF(t,0,\mathbf{x}) \ \left(= \int_B \frac{S_*(t|\mathbf{X})}{S(t|\mathbf{X})} S(t|\mathbf{x})dF_C(t)dF_{\mathbf{X}}(\mathbf{x}) \quad (see\ Eq.(1.2))\right)$$

$$= \int_B dF_*(t,0,\mathbf{x}) \overset{\text{def}}{=} \mu(B). \quad \textbf{(2.6)}$$

Verify that $\int ln\frac{S(t|\boldsymbol{X})}{\hat{S}(t|\boldsymbol{X})} d\hat{F}_n(t,0,\mathbf{x}) = \int H(\frac{S(t|\boldsymbol{X})}{\hat{S}(t|\boldsymbol{X})})\frac{\hat{S}(t|\boldsymbol{X})}{S(t|\boldsymbol{X})} d\hat{F}_n(t,0,\mathbf{x}),\ \ where$

$$H(t) = tlogt \geq -1/e \text{ for } t > 0 \text{ and } H\left(S(t|\boldsymbol{X})/\hat{S}(t|\boldsymbol{X})\right) \geq -1/e \quad \textbf{(2.7)}$$

$$\underline{lim}_{n\to\infty}\int \frac{S(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t,0,\mathbf{x}) = \underline{lim}_{n\to\infty}\int H(\frac{S(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})})\frac{\hat{S}(t|\mathbf{x})}{S(t|\mathbf{x})} d\hat{F}_n(t,0,\mathbf{x})$$

$$= \underline{lim}_{n\to\infty}\int H(\frac{S(t|\mathbf{X})}{\hat{S}(t|\mathbf{X})})d\mu_n(t,\mathbf{x}) \qquad \text{(see (2.5))}$$

$$\geq \int \underline{lim}_{n\to\infty}H(\frac{S(t|\mathbf{X})}{\hat{S}(t|\mathbf{X})})d\mu(t,\mathbf{x}) \qquad \text{(by (2.6), (2.7) and Corollary 1)}$$

$$= \int \underline{lim}_{n\to\infty}H(\frac{S(t|\mathbf{X})}{\hat{S}(t|\mathbf{X})})\, dF_*(t,0,\mathbf{x}) \qquad \text{(see (2.6))}$$

$$= \int \frac{S(t|\mathbf{X})}{S_*(t|\mathbf{X})} ln\frac{S(t|\mathbf{X})}{S_*(t|\mathbf{X})} dF_*(t,0,\mathbf{x}) \quad \textbf{(2.8)}$$

$$= \int\int \frac{S(t|\mathbf{X})}{S_*(t|\mathbf{X})} ln\frac{S(t|\mathbf{X})}{S_*(t|\mathbf{X})} S_*(t|\mathbf{x})dF_C(t)dF_{\mathbf{x}}(\mathbf{x}) \qquad \text{(by Eq. (1.2))}$$

$$= \int ln\frac{S(t|\mathbf{X})}{S_*(t|\mathbf{X})} dF(t,0,\mathbf{x})$$

Similarly, since $\frac{\hat{S}(t-|\mathbf{X})-\hat{S}(t|\mathbf{X})}{S(t-|\mathbf{X})-S(t|\mathbf{X})} \leq \frac{1}{S(t-|\mathbf{X})-S(t|\mathbf{X})} \overset{\text{def}}{=} K(y,1,\mathbf{x})$ and

$$\int_B K(t,1,\mathbf{x})d\hat{F}_n(t,1,\mathbf{x}) = \int_B \frac{1}{S(t-|\mathbf{X})-S(t|\mathbf{X})} d\hat{F}_n(t,1,\mathbf{x})$$

$$\to \int_B \frac{1}{S(t-|\mathbf{X})-S(t|\mathbf{X})} S_C(t)dF(t|\mathbf{x})dF_X(\mathbf{x}),$$

letting $v_n(B) \overset{\text{def}}{=} \int_B \frac{\hat{S}(t-|\mathbf{X})-\hat{S}(t|\mathbf{X})}{S(t-|\mathbf{X})-S(t|\mathbf{X})} d\hat{F}_n(t,1,\mathbf{x}),$

$$\underline{lim}_{n\to\infty}v_n(B) = \int_B \underline{lim}_{n\to\infty}\frac{\hat{S}(t-|\mathbf{X})-\hat{S}(t|\mathbf{X})}{S(t-|\mathbf{X})-S(t|\mathbf{X})} dF(t,1,\mathbf{x}) \qquad \text{(by statement (2)of Lemma 2)}$$

$$= \int_B \frac{S_*(t-|\mathbf{x})-S_*(t|\mathbf{x})}{S(t-|\mathbf{x})-S(t|\mathbf{x})} dF(t,1,\mathbf{x})$$

$$= \int \sum_t \mathbf{1}((t,\mathbf{x}) \in B) \frac{f_*(t|\mathbf{x})}{f(t|\mathbf{x})} f(t|\mathbf{x}) S_C(t) dF_X(\mathbf{x}) \qquad \text{(see Eq. (1.2))}$$

$$= \int_B 1 dF_*(t,1,\mathbf{x}) \stackrel{\text{def}}{=} v(B). \tag{2.9}$$

Since $\int ln \frac{S(t-|\mathbf{x})-S(t|\mathbf{x})}{\hat{S}(t-|\mathbf{x})-\hat{S}(t|\mathbf{x})} d\hat{F}_n(t,1,\mathbf{x})$

$$= \int H\left(\frac{S(t-|\mathbf{x})-S(t|\mathbf{x})}{\hat{S}(t-|\mathbf{x})-\hat{S}(t|\mathbf{x})}\right) \frac{\hat{S}(t-|\mathbf{x})-\hat{S}(t|\mathbf{x})}{S(t-|\mathbf{x})-S(t|\mathbf{x})} d\hat{F}_n(t,1,\mathbf{x}),$$

$H\left(\frac{S(t-|\mathbf{x})-S(t|\mathbf{x})}{\hat{S}(t-|\mathbf{x})-\hat{S}(t|\mathbf{x})}\right) \geq -\frac{1}{e}$ and v_n converges set wisely to a finite measure v (see (2.9)), by a similar argument as in (2.4), (2.6), (2.7) and (2.8), we have:

$$\underline{lim}_{n\to\infty} \int ln \frac{S(t-|\mathbf{x})-S(t|\mathbf{x})}{\hat{S}(t-|\mathbf{x})-\hat{S}(t|\mathbf{x})} d\hat{F}_n(t,1,\mathbf{x})$$

$$= \underline{lim}_{n\to\infty} \int H\left(\frac{S(t-|\mathbf{x})-S(t|\mathbf{x})}{\hat{S}(t-|\mathbf{x})-\hat{S}(t|\mathbf{x})}\right) \frac{\hat{S}(t-|\mathbf{x})-\hat{S}(t|\mathbf{x})}{S(t-|\mathbf{x})-S(t|\mathbf{x})} d\hat{F}_n(t,1,\mathbf{x})$$

$$\geq \int \underline{lim}_{n\to\infty} H\left(\frac{S(t-|\mathbf{x})-S(t|\mathbf{x})}{\hat{S}(t-|\mathbf{x})-\hat{S}(t|\mathbf{x})}\right) dF_*(t,1,\mathbf{x})$$

$$= \int \ln \frac{f(t|\mathbf{x})}{f_*(t|\mathbf{x})} dF(t,1,\mathbf{x}) \quad \text{(as } Y \text{ is discrete).} \tag{2.10}$$

$$0 \geq \int ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t,0,\mathbf{x}) + \int ln \frac{S(t-|\mathbf{x})-S(t|\mathbf{x})}{S_*(t-|\mathbf{x})-S_*(t|\mathbf{x})} dF(t,1,\mathbf{x}) \text{ (by Eq. (2.2))}$$

$$= \int ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t,0,\mathbf{x}) + \int ln \frac{f(t|\mathbf{x})}{f_*(t|\mathbf{x})} dF(t,1,\mathbf{x}) \quad \text{(by 2.8)and (2.10))}$$

$$\geq 0 \qquad \text{(by Lemma 1, the modified K-L inequality).}$$

Thus, $\int ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t,0,\mathbf{x}) + \int ln \frac{f_*(t|\mathbf{x})}{f(t|\mathbf{x})} dF(t,1,\mathbf{x}) = 0$. Hence, $(S_0(t),\beta) = (S_*(t),\beta) \forall t \in D$ by the 2nd statement of the K-L inequality.

Theorem 3. *Under the Cox model with RC data, if Y is continuous then Eq. (2.1) holds.*

Proof. For the given $\omega \in \Omega$ and (S_*,β_*) in the proof of Theorem 1, as well as $\hat{\beta}(\omega)$ and $\hat{S}(t|\mathbf{x})(\omega)$, we have $S_*(t|\mathbf{x}) = (S_*(t))^{\exp(\beta'\mathbf{x})}$. By a similar argument as in proving Eq. (2.8), we can show:

$$\underline{lim}_{n\to\infty} \int ln \frac{S(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t,0,\mathbf{x}) = \underline{lim}_{n\to\infty} \int H\left(\frac{S(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})}\right) \frac{\hat{S}(t|\mathbf{x})}{S(t|\mathbf{x})} d\hat{F}_n(t,0,\mathbf{x}) \geq \int ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t,0,\mathbf{x}) \tag{2.11}$$

In view of Eq. (1.4) due to Y is continuous, we denote:

$$G(t,\mathbf{x},n) = \frac{\hat{S}(t-\eta_n|\mathbf{x})-\hat{S}(t|\mathbf{x})}{S(t-\eta_n|\mathbf{x})-S(t|\mathbf{x})}, \quad A_k = \{G(t,\mathbf{x},n) \leq k, \forall n\} \text{ and } B_k = A_k \backslash A_{k-1} \tag{2.12}$$

$$G(t,\mathbf{x},n) = \frac{\hat{S}(t-\eta_n|\mathbf{x})-\hat{S}(t|\mathbf{x})}{S(t-\eta_n|\mathbf{x})-S(t|\mathbf{x})} = \frac{\hat{S}(t-\eta_n|\mathbf{x})-\hat{S}(t|\mathbf{x})/\eta_n}{S(t-\eta_n|\mathbf{x})-S(t|\mathbf{x})/\eta_n} \to \frac{F'_*(t|\mathbf{x})}{F'(t|\mathbf{x})} \text{ a.e.,} \tag{2.13}$$

as S_* is a monotone function, S_*' exists a.e., and so do $S_*'(t|\mathbf{x})$ and $F_*'(t|\mathbf{x})$. We have

$$\int \mathbf{1}(\cup_{k\geq 1} B_k)\, dF(t,s,\mathbf{x}) = 1. \qquad \textbf{(2.14)}$$

The reason is as follows. For each $(t, \mathbf{x})$ such that $F'(t|\mathbf{x}) > 0$ and Eq. (2.13) holds, $F_*'(t|\mathbf{x}) / F'(t|\mathbf{x})$ $(= f_*(t|\mathbf{x}) / f(t|\mathbf{x}))$ is finite. Then, there exists n_o such that $G(t, \mathbf{x}, n) < 1 + F_*'(t|\mathbf{x}) / F'(t|\mathbf{x})$ for $n \geq n_o$. On the other hand, $G(t, \mathbf{x}, n)$ is finite for $n = 1, ..., n_o$. Thus, $G(t, \mathbf{x}, n) < k$ for some k. Since Eq. (2.1) holds a.e. and $\int 1 dF(t, s, \mathbf{x}) = 1$, Eq. (2.14) holds.

We shall prove in Lemma 3 that

$$\underline{lim}_{n\to\infty} \int_{B_k} ln \frac{s(t-\eta_n|\mathbf{x}) - s(t|\mathbf{x})}{\hat{s}(t-\eta_n|\mathbf{x}) - \hat{s}(t|\mathbf{x})} d\hat{F}_n(t,1,\mathbf{x}) \geq \int_{B_k} ln \frac{f(t|\mathbf{x})}{f_*(t|\mathbf{x})} dF(t,1,\mathbf{x}) \quad \text{for } k \geq 1. \qquad \textbf{(2.15)}$$

Then,

$$\begin{aligned}
&\underline{lim}_{n\to\infty} \int ln \frac{S(t-\eta_n|\mathbf{x}) - S(t|\mathbf{x})}{\hat{S}(t-\eta_n|\mathbf{x}) - \hat{S}(t|\mathbf{x})}\, d\hat{F}_n(t,1,\mathbf{x}) \\
&= \underline{lim}_{n\to\infty} \int -lnG(t,\mathbf{x},n) d\hat{F}_n(t,1,\mathbf{x}) \\
&= \underline{lim}_{n\to\infty} \textstyle\sum_{k\geq 1} \int_{B_k} -lnG(t,\mathbf{x},n) d\hat{F}_n(t,1,\mathbf{x}) \qquad \text{(by (2.14))} \\
&= \underline{lim}_{n\to\infty} \int_{k\geq 1}\int_{B_k} -lnG(t,\mathbf{x},n) d\hat{F}_n(t,1,\mathbf{x})\, dv(k) \qquad (dv \text{ is a counting measure}) \\
&= \underline{lim}_{n\to\infty} \int_{k\geq 1}\int_{B_k} H((G(t,\mathbf{x},n))^{-1}) G(t,\mathbf{x},n) d\hat{F}_n(t,1,\mathbf{x})\, dv(k) \qquad (H(t) = tlnt) \\
&\geq \int_{k\geq 1} \underline{lim}_{n\to\infty} \int_{B_k} H((G(t,\mathbf{x},n))^{-1}) G(t,\mathbf{x},n) d\hat{F}_n(t,1,\mathbf{x}) dv(k) \qquad (\text{by Corollary 1, as} \\
&\qquad\qquad H(t) \geq -\frac{1}{e} \text{ and } \int_{B_k} \ln(G(t,\boldsymbol{x},n))^{-1} d\hat{F}_n(t,1,\boldsymbol{x}) \quad \text{is bounded below by} - 1/\mathrm{e} \\
&\geq \textstyle\sum_{k\geq 1} \underline{lim}_{n\to\infty} \int_{B_k} -lnG(t,\mathbf{x},n) d\hat{F}_n(t,1,\mathbf{x}) \qquad (\text{ by (2.15)}) \qquad \textbf{(2.16)} \\
&= \sum_{k\geq 1} \int_{B_k} ln \frac{f(t|\mathbf{x})}{f_*(t|\mathbf{x})} dF(t,1,\mathbf{x}) \\
&= \int ln \frac{f(t|\mathbf{x})}{f_*(t|\mathbf{x})} dF(t,1,\mathbf{x})
\end{aligned}$$

Since $\hat{S}(t|\mathbf{x})$ is the SMLE,

$$\begin{aligned}
0 &\geq \underline{lim}_{n\to\infty} \left[\int ln \frac{S(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t,0,\mathbf{x}) + \int ln \frac{S(t-\eta_n|\mathbf{x}) - S(t|\mathbf{x})}{\hat{S}(t-\eta_n|\mathbf{x}) - \hat{S}(t|\mathbf{x})} d\hat{F}_n(t,1,\mathbf{x}) \right] \\
&\geq \int ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t,0,\mathbf{x}) + \int ln \frac{f(t|\mathbf{x})}{f_*(t|\mathbf{x})} dF(t,1,\mathbf{x}) \qquad \text{(by (2.11) and (2.16))} \\
&\geq 0 \qquad \big(\text{by Lemma 1 (the modified K} - \text{L inequality)}\big).
\end{aligned}$$

3. RESULTS

The last inequality further implies that $\int ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} d\, F(t,0,\mathbf{x})$ + $\int ln \frac{f(t|\mathbf{x})}{f_*(t|\mathbf{x})} d\, F(t,1,\mathbf{x}) = 0$. Thus, $(S_0(t),\beta) = (S_*(t),\beta_*)\ \forall t \in$ D by the 2nd statement of the K-L inequality and by the assumption ASI.

Lemma 3. Inequality (2.15) holds.

Proof. Let $k \geq 1$ and $\nu_n(B) \overset{\text{def}}{=} \int_{B \cap \mathrm{B}_n} \mathrm{G}_n(t,\mathbf{x},n) d\hat{F}_n(t,1,\mathbf{x})$, where B is a measurable set and $G(t,\mathbf{x},n) = \frac{\hat{s}(t-|\mathbf{x})-\hat{s}(t|\mathbf{x})}{s(t-|\mathbf{x})-s(t|\mathbf{x})} \in [0,k]$ on B_k,

$$\lim_{n\to\infty} \nu_n(B) = \lim_{n\to\infty} \int_{B\cap \mathrm{B}_k} \mathrm{G}_n(t,\mathbf{x},n) d\hat{F}_n(t,1,\mathbf{x})$$
$$= \int_{B\cap \mathrm{B}_k} \lim_{n\to\infty} G(t,\mathbf{x},n) dF(t,1,\mathbf{x}) \quad \text{(by Lemma 2, as } G(t,\mathbf{x},n) \in [0,k])$$
$$= \iint \mathbf{1}((t,\mathbf{x}) \in B\cap \mathrm{B}_k) \frac{f_*(t|\mathbf{x})}{f(t|\mathbf{x})} f(t|\mathbf{x}) \mathrm{S}_c(t) dt F_{\mathbf{x}}(\mathbf{x}) \quad \text{(see Eq. (1.2))}$$
$$= \int_{B\cap \mathrm{B}_k} \mathrm{dF}_*(t,1,\mathbf{x}) \overset{\text{def}}{=} d\nu(B) \quad \text{(see Eq. (1.2))}$$

CONCLUSION

Since $H((S(t\text{-}\eta_n|\mathbf{x}) - (S(t|\mathbf{x}))/((\hat{S}(t\text{-}\eta_n|\mathbf{x}) - (\hat{S}(t|\mathbf{x}))) \geq -1/e$ and v_n converges set wisely to a finite measure v by a similar argument as in (2.4), (2.6), (2.7) and (2.8), we can show that:

$$\underline{lim}_{n\to\infty} \int_{\mathrm{B}_k} ln \frac{S(t-\eta_n|\mathbf{x}) - S(t|\mathbf{x})}{\hat{S}(t-\eta_n|\mathbf{x}) - \hat{S}(t|\mathbf{x})} d\hat{F}_n(t,1,\mathbf{x})$$
$$= \underline{lim}_{n\to\infty} \int_{\mathrm{B}_k} H\left(\frac{s(t-\eta_n|\mathbf{x})-s(t|\mathbf{x})}{\hat{s}(t-\eta_n|\mathbf{x})-\hat{s}(t|\mathbf{x})}\right) \frac{\hat{s}(t-\eta_n|\mathbf{x})-\hat{s}(t|\mathbf{x})}{s(t-\eta_n|\mathbf{x})-s(t|\mathbf{x})} d\hat{F}_n(t,1,\mathbf{x})$$
$$\geq \int_{\mathrm{B}_k} \underline{lim}_{n\to\infty} H\left(\frac{s(t-\eta_n|\mathbf{x})-s(t|\mathbf{x})}{\hat{s}(t-\eta_n|\mathbf{x})-\hat{s}(t|\mathbf{x})}\right) dF_*(t,1,\mathbf{x})$$
$$= \int_{\mathrm{B}_k} ln \frac{f(t|\mathbf{x})}{f_*(t|\mathbf{x})} dF(t,1,\mathbf{x}) \quad \text{for } k \geq 1. \quad \square$$

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

Not applicable.

FUNDING

None.

CONFLICT OF INTEREST

The author declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

The author would like to thank the editor and two referees for their invaluable comments.

REFERENCES

[1] Yu. Qiqing, "A note on the proportional hazards model with discontinuous data", *Stat. Probab. Lett.*, vol. 77, no. 7, pp. 735-739, 2007. [http://dx.doi.org/10.1016/j.spl.2006.11.008]

[2] J. Kiefer, "J and J. Wolfowitz. "Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters", *Ann. Math. Stat.*, vol. 1, pp. 887-906, 1956. [http://dx.doi.org/10.1214/aoms/1177728066]

[3] G.Y.C. Wong, M.P. Osborne, Q.G. Diao, and Q.Q. Yu, "The piece-wise cox model with right-censored data", *Comm. Statist. Comput. Simul.*, vol. 46, pp. 7894-7908, 2017. [http://dx.doi.org/10.1080/03610918.2016.1255968]

[4] D.R. Cox, and D. Oakes, *Analysis of Survival Data.*, Chapman & Hall NY, 1984.

[5] Q.Q. Yu, and Q.G Diao, "Consistency of the semi-parametric MLE under the Cox model with linearly time-dependent covariates and interval-censored data", *J. Adv.Stat.*, vol. 4, no. 1, .

[6] S. Kullback, and R.A. Leibler, "On information and sufficiency", *Ann. Math. Stat.*, vol. 22, pp. 79-86, 1951. [http://dx.doi.org/10.1214/aoms/1177729694]