

CONSISTENCY OF THE MODIFIED SEMI-PARAMETRIC MLE UNDER THE LINEAR REGRESSION MODEL WITH RIGHT-CENSORED DATA

QIQING YU

Department of Mathematical Sciences

Binghamton University

NY 13902

USA

e-mail: qyu@math.binghamton.edu

Abstract

Under the right censorship model and under the linear regression model $Y = \beta X + W$, where $E(W)$ may not exist, the modified semi-parametric MLE (MSMLE) was proposed by Yu and Wong [17]. The MSMLE $\hat{\beta}$ of β satisfying $P(\hat{\beta} \neq \beta \text{ infinitely often}) = 0$ if W is discontinuous, and the simulation study suggests that it is also consistent and efficient under certain regularity

2020 Mathematics Subject Classification: Primary 62N01; Secondary 62G08.

Keywords and phrases: linear regression model, modified semi-parametric maximum likelihood estimator, consistency, right-censored data.

Received June 07, 2021

© 2021 Scientific Advances Publishers

This work is licensed under the Creative Commons Attribution International License (CC BY 3.0).

http://creativecommons.org/licenses/by/3.0/deed.en_US



conditions. In this paper, we establish the consistency of the MSMLE under the necessary and sufficient condition that β is identifiable. Notice that under the latter assumption, the Buckley-James estimator and the median regression estimator can be inconsistent (see Yu and Dong [20]).

1. Introduction

We shall establish the consistency of the modified semi-parametric maximum likelihood estimator (MSMLE) proposed in Yu and Wong [17] under the linear regression model with right-censored data. We shall make the following assumptions.

A1. Let $(M_i, \delta_i, \mathbf{X}_i)$, $i = 1, \dots, n$, be i.i.d. observations from the random vector (M, δ, \mathbf{X}) , where $M = Y \wedge C$, C is a random censoring variable, $Y = \beta' \mathbf{X} + W$, $\mathbf{X} \in \mathcal{R}^p$ (the p -dimensional Euclidean space), W is the baseline random variable ($= Y | \mathbf{X} = 0$), W , C and \mathbf{X} are independent, $\delta = \mathbf{1}(Y \leq C)$, $\mathbf{1}(A)$ is the indicator function of the event A and $P(\delta = 1) \in (0, 1)$. Both β and S_o are unknown, where $S_o(y) = S_W(y) = P(W > y)$.

This is a semi-parametric set-up, as (β, F_o) is unknown ($F_o = 1 - S_o$). $E(W)$ may not exist.

Regression analysis is one of the most widely used statistical techniques. Its applications occur in almost every field, including engineering, economics, the physical sciences, management, life and biological sciences, and the social sciences.

To review available estimators for the regression problem, we first consider the case of complete data under the simple linear regression model. Suppose (X_i, Y_i) , $i = 1, \dots, n$, are i.i.d. observations from (X, Y) . There are several possible estimators for β , such as

- (1) the least squares estimator (LSE),
- (2) the Theil-Sen estimator (Theil [16] and Sen [15]),

(3) L -estimators and R -estimators (see, e.g., Montgomery and Peck [11]),

(4) adaptive estimators (Bickel [2]),

(5) various M -estimators (Huber [7]),

(6) the quantile (or median) regression estimator (see, e.g., He and Zhu [6]),

(7) the empirical likelihood estimator (Owen [12]),

(8) the semi-parametric maximum likelihood estimators (SMLE) (Yu and Wong [19]), and the modified SML (MSMLE) (Yu and Wong [17]).

Several of these semi-parametric estimators of β are a value of b that maximizes the generalized likelihood function

$$\mathbb{L} = \mathbb{L}(f, b) = \prod_{i=1}^n f(Y_i - bX_i), \tag{1.1}$$

where f belongs to a class of density functions.

If $f(t) = S(t-) - S(t)$, where $S(\cdot)$ is a survival function, it leads to the SMLE. If f in Equation (1.1) is a kernel estimator, it leads to M -estimators or the MSMLE. Various M -estimators have been proposed for finding a zero point (or zero-crossing point) of $\frac{\partial \ln \mathbb{L}(\hat{f}, b)}{\partial b}$, where \hat{f} is a kernel estimate of f . Zhang and Li [22] consider such an approach. Let ϕ

be the score function, that is, $\phi = (\ln f)' = \frac{f'}{f}$, where f' is the derivative

of f . Let $\hat{\phi}$ be an estimate of ϕ . A point x is said to be a zero-crossing point of a function g if $g(x-)g(x+) \leq 0$. Zhang and Li's M -estimate of β

is a zero-crossing point of a function $\Phi(\hat{\phi}, \cdot)$, where

$$\Phi(\hat{\phi}, b) = \sum_{i=1}^n \hat{\phi}(Y_i - \bar{Y} - b(X_i - \bar{X}))(X_i - \bar{X}). \tag{1.2}$$

An M -estimate can be obtained by iterative algorithms. Zhang and Li point out that the M -estimator with $\hat{\phi}(x) = x$ is the LSE, and thus is not efficient. They also show that the M -estimator with a suitable choice of $\hat{\phi}$ is efficient under certain regularity conditions.

Under right censoring with $\mathbf{X} \in \mathcal{R}^P$, there are several extensions of the above estimators. The Buckley-James [3] estimator (BJE) is a modification of normal equations of the sum of least squares. Chatterjee and Mcleish [4] and Leurgans [10] propose several parametric and semi-parametric extensions of the LSE. Hillis [8], Ritov [13] and Zhang and Li [22] consider M -estimators and their modifications. Ireson and Rao [9] and Akritas et al. [1] consider extension of the Theil-Sen estimator. Since all these estimators are extensions, they inherit the properties of the estimators in the case of complete-data. Yu and Wong [17] propose the MSMLE of β , denoted by $\hat{\beta}$ or $\hat{\beta}_n$, which maximizes the likelihood

$\mathcal{L}(S, \mathbf{b}) = \prod_{i=1}^n [(f_{\mathbf{b}}(T_i(\mathbf{b})))^{\delta_i} (S_{\mathbf{b}}(T_i(\mathbf{b})))^{1-\delta_i}]$, where $S_{\mathbf{b}}$ is the product-limit-estimator (ple), $T_i(\mathbf{b}) = Y_i - \mathbf{b}\mathbf{X}_i$, and f is the kernel estimator with the rectangular kernel.

Remark 1. Yu and Wong show that $\hat{\beta}_n$ cannot be obtained by the algorithms for M -estimate, or by Newton-Raphson algorithm, or Monte Carlo method, as $\frac{d \ln \mathcal{L}(f, b)}{db} = 0$ a.e..

Yu and Wong [17] propose a feasible non-iterative algorithm for obtaining $\hat{\beta}_n$. They establish the consistency of the MSMLE under the following assumptions in addition to A1:

A2. $P(A) > 0$, where $A = \{(\mathbf{x}_1, \dots, \mathbf{x}_{p+1}) : \text{rank} \begin{pmatrix} 1 & \dots & 1 \\ \mathbf{x}_1 & \dots & \mathbf{x}_{p+1} \end{pmatrix} = p + 1, \mathbf{x}_i\text{'s are i.i.d. copies of } \mathbf{X} | \delta_1 = \dots = \delta_{p+1} = 1\}$.

A3. $F_o(t) = \int_{x \leq t} f_o(x) d\mu(x)$, where μ is the sum of the Lebesgue

measure on the real line and a counting measure on a countable set \mathcal{M} ,

$$\text{and } f_o(x) = \begin{cases} \frac{dF_o(x)}{dx}, & \text{if } x \notin \mathcal{M}, \\ F_o(x) - F_o(x-), & \text{if } x \in \mathcal{M}. \end{cases}$$

A4. $|E(\ln f_o(\epsilon))| < \infty$, $F_o(t+x) - F_o(t-x) = f_o(t)(1 + O(x))$ uniformly

for all $t \in \mathcal{M}$, $\frac{F_o(t+x) - F_o(t-x)}{2x} = f_o(t)(1 + O(x))$ uniformly for all

$t \notin \mathcal{M}$.

They also prove that

$$P\{\hat{\beta}_n = \beta \text{ for all large } n\} = 1 \text{ if } F_o \text{ has a discontinuity point and A2 holds.} \tag{1.3}$$

It is conjectured according to simulation results (see Yu and Wong [17])

that under certain regularity conditions, $\hat{\beta}_n$ is efficient.

An MSMLE solution is a maximizer of $\mathcal{L}(f, b)$, while an M -estimator solution is an approximation of a stationary point of $\mathcal{L}(f, b)$. Even though Zhang and Li [22] show that their M -estimator is consistent and efficient under certain regularity conditions, there are two drawbacks in their approach, in comparison to the MSMLE approach. One is in the assumption for consistency and the other is in computation.

(a) Zhang and Li's M -estimator can be inconsistent if $\Phi(\phi_o, b) = 0$ a.s. in b , where $\phi_o = f'_o/f_o$. The reason is as follows. (1) An M -estimate is a zero-crossing point of $\Phi(\hat{\phi}, b)$, and (2) one expects that the derivative of "the normalized log likelihood" $\frac{1}{n} \Phi(\hat{\phi}, b) \rightarrow E(\frac{1}{n} \Phi(\phi_o, b))$ a.s. for each b . Note that for uniform distributions or piece-wise uniform distributions (among other distributions), $\phi_o = 0$ a.s., and thus $\Phi(\phi_o, b) = 0$ a.s. in b .

Consequently, if the data are from the latter distributions and the score function f'/f is estimated by a kernel estimator, one cannot find a consistent M -estimate of β . On the contrary, we shall show in this paper that the MSMLE is consistent even if $\Phi(\phi_o, b) = 0$ a.s..

(b) Given a data set, even if the efficient and consistent M -estimate exists, the current algorithms (see Buckley and James [3] and Zhang and Li [22]) may not be able to find it, as there are often multiple solutions to the M -estimator and the algorithm can only find one of them. Yu and Wong ([18], Example 4.2 and Figure 1) present such an example for the BJE, which is also an M -estimator. Because Yu and Wong [18] propose an algorithm that can present all possible solutions of the BJE, we can examine which solution of the BJE can be obtained by the existing algorithms for the M -estimation. The algorithm for the BJE cannot be generalized to other M -estimators. Moreover, Zhang and Li did not show that every solution to their M -estimator is consistent, even under the set of regularity conditions imposed on the underlying distributions.

(c) An estimator of the parameter β is consistent only if β is identifiable. Yu and Dong establish the necessary and sufficient (NS) conditions for $(\beta, (S_o(t)))$ being identifiable. The consistency of all the existing estimators under the linear regression model (see A1) are all established under assumption A2, among other regularity conditions. In fact, both the BJE and the median regression estimator can be inconsistent under the NS condition for β being identifiable (see Yu and Dong [20]).

In this paper, we shall prove consistency of the MSMLE $\hat{\beta}_n$ without additional assumptions rather than A1 and the necessary and sufficient (NS) conditions for $(\beta, (S_o(t)))$ being identifiable (see Theorems 1 and 2). These assumptions allow F_o to be continuous or discontinuous. It is worth mentioning that the standard approach in proving consistency of

the MLE takes advantage of the normal equation $\frac{d \ln \mathfrak{L}}{d \mathbf{b}} \Big|_{\mathbf{b}=\hat{\beta}_n} = 0$. However, it does not work here, as $\frac{d \ln \mathfrak{L}}{d \mathbf{b}} = 0$ a.s. (see Remark 1). The asymptotic efficiency of the MSMLE under the assumption that F_o is continuous is still an open question. The paper is organized as follows. In Section 2, we present notations and introduce the MSMLE $\hat{\beta}_n$. In Section 3, we prove the consistency of the MSMLE.

2. Preliminary Results

In this section, we introduce some preliminary results. For $\mathbf{b} \in \mathcal{R}^p$, denote $T_i = T_i(\mathbf{b}) = M_i - \mathbf{b}'\mathbf{X}_i$. It follows from assumption (A1) that $(T_i(\beta), \delta_i)$'s are i.i.d. copies of $(W \wedge Z^c, \delta)$, where $Z^c = C - \beta'\mathbf{X}$. The generalized likelihood function is

$$\mathfrak{L}(S, \mathbf{b}, f) = \prod_{i=1}^n [(f(T_i(\mathbf{b})))^{\delta_i} (S(T_i(\mathbf{b})))^{1-\delta_i}],$$

where

$$f(t) = S(t-) - S(t), S \in \mathcal{F}, \tag{2.1}$$

$\mathcal{F} = \{H : H \text{ is a non-increasing function on } [-\infty, \infty], H(-\infty) = 1 \text{ and } H(\infty) = 0\}$. Yu and Wong [17] suggest to replace f by a kernel estimate,

$$f(x) = f_S(x) = - \int \frac{1}{\eta_n} K\left(\frac{x-t}{\eta_n}\right) dS(t), \text{ where } K(x) = \frac{1}{2} \mathbf{1}_{(|x| \leq 1)} \text{ and } \eta_n \rightarrow 0 \tag{2.2}$$

(e.g., $\eta_n = n^{-1/5}$), and replace S by the product-limit-estimate (PLE), denoted by $\hat{S}_{\mathbf{b}}$, based on observations $(T_i(\mathbf{b}), \delta_i)$'s. Let $F = 1 - S$ for $S \in \mathcal{F}$ and let $\hat{F}_{\mathbf{b}} = 1 - \hat{S}_{\mathbf{b}}$. Since $f_S(x) = \frac{S(x - \eta_n) - S(x + \eta_n)}{2\eta_n}$, (2.1)

becomes

$$\begin{aligned} l(\mathbf{b}) &= \mathbb{E}(\hat{S}_{\mathbf{b}}, \mathbf{b}, \hat{f}_{\hat{S}_{\mathbf{b}}}) = \prod_{i=1}^n [(f_{\hat{S}_{\mathbf{b}}}(T_i(\mathbf{b})))^{\delta_i} (\hat{S}_{\mathbf{b}}(T_i(\mathbf{b})))^{1-\delta_i}] \\ &= \prod_{i=1}^n \left[\frac{1}{2\eta_n} [\hat{S}_{\mathbf{b}}((T_i(\mathbf{b}) - \eta_n) -) - \hat{S}_{\mathbf{b}}(T_i(\mathbf{b}) + \eta_n)]^{\delta_i} (\hat{S}_{\mathbf{b}}(T_i(\mathbf{b})))^{1-\delta_i} \right]. \end{aligned} \quad (2.3)$$

Yu and Wong call a value of \mathbf{b} that maximizes $l(\cdot)$ over all $\mathbf{b} \in \mathcal{R}^p$ an MSMLE of β . The MSMLE of $S_o(t)$ based on $Y_1 - \mathbf{X}'_1 \hat{\beta}, \dots, Y_n - \mathbf{X}'_n \hat{\beta}$ is denoted by $\hat{S}(t)$.

It is obvious that the MSMLE of β is consistent only when the parameter is identifiable. Yu and Dong [20] establish the NS conditions for the parameter being identifiable under the semi-parametric linear regression model. Let \mathcal{D}_T be the support set of f_T , that is, $P(\|T - t\| < c) > 0 \quad \forall c > 0$ if $t \in \mathcal{D}_T$, where $\|\cdot\|$ denotes a norm. Here T can be W, Y, C or \mathbf{X} . Let $C_o = C - \mathbf{X}'\beta$ and $\tau_o = \tau_{C_o}$. Define

$$\mathcal{A} = \begin{cases} (-\infty, \tau_o), & \text{if } P(C = \tau_C) = 0 < S_W(\tau_o -), \\ (-\infty, \tau_o], & \text{if } P(C = \tau_C)S_W(\tau_o) > 0, \\ (-\infty, \infty), & \text{if } S_W(\tau_o -) = 0 \text{ or } P(C = \tau_C) \\ & \quad \times S_W(\tau_o -) > 0 = S_W(\tau_o). \end{cases} \quad (2.4)$$

Definition. Assume A1 holds. (S_o, β) is said to be *identifiable* if $(S_*(t - \mathbf{x}'\beta_*) = S_o(t - \mathbf{x}'\beta) \quad \forall t - \mathbf{x}'\beta \in \mathcal{A}$ implies that $(S_*(t), \beta_*) = (S_o(t), \beta) \quad \forall t \in \mathcal{A}$.

Theorem 1 ([20]). *Suppose that A1 holds and $\tau_o < \infty$. Then*

(a) *The survival function $S_W(t)$ is identifiable iff $t \in \mathcal{A}$.*

(b) *The parameter β is identifiable iff $\mathcal{B}_{\mathbf{x}_0} \neq \emptyset$, where $\mathbf{x}_0 \in \mathcal{D}_{\mathbf{x}}$ such that $\tau_C - \beta'_{\mathbf{x}_0} = \tau_o$, and $\mathcal{B}_{\mathbf{x}_0} = \{(w_1, \mathbf{x}_1, \dots, w_p, \mathbf{x}_p) : \mathbf{x}_1 - \mathbf{x}_0, \dots, \mathbf{x}_p - \mathbf{x}_0$ are linearly independent, $\mathbf{x}_i \in \mathcal{D}_{\mathbf{x}}$, $w_i \in \mathcal{D}_W$ and*

$$w_i + \beta'_{\mathbf{x}_i} \left\{ \begin{array}{ll} \leq \tau_C & \text{if } P(C = \tau_C) > 0 \\ < \tau_C & \text{otherwise} \end{array} \right\}.$$

Theorem 2 ([20]). *Suppose that A1 holds and $\tau_o = \infty$.*

(a) *The survival function $S_W(t)$ is identifiable for each t .*

(b) *The parameter β is identifiable iff $\exists \mathbf{x}_0 \in \mathcal{D}_{\mathbf{x}}$ such that $\mathcal{B}_{\mathbf{x}_0} \neq \emptyset$, where $\mathcal{B}_{\mathbf{x}_0} = \{(\mathbf{x}_1, \dots, \mathbf{x}_p) : \mathbf{x}_1 - \mathbf{x}_0, \dots, \mathbf{x}_p - \mathbf{x}_0$ are linearly independent, and $\mathbf{x}_i \in \mathcal{D}_{\mathbf{x}}\}$. Here $\mathbf{x}_0 = \mathbf{0}$ if $\mathbf{0} \in \mathcal{D}_{\mathbf{x}}$, otherwise $\mathbf{x}_0, \dots, \mathbf{x}_p$ are vectors belonging to $\mathcal{D}_{\mathbf{x}}$.*

Notice that if A1 holds, then A2 corresponds to $\mu(\mathcal{B}_{\mathbf{x}_0}) > 0$.

We shall make use of the modified Kullback-Leibler (KL) inequality as follows.

Proposition 1 ([21]). *If $f_i \geq 0$, $\int f_1(t) d\mu_1(t) = 1$ and $\int f_2(t) d\mu_1(t) \leq 1$,*

where μ_1 is a measure, 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 .

Denote $S(t|\mathbf{x}) = S_o(t - \mathbf{x}'\beta)$, $\hat{S}(t|\mathbf{x}) = \hat{S}_{\hat{\beta}}(t - x\hat{\beta})$, $\hat{S}_{\beta}(t|\mathbf{x}) = \hat{S}_{\beta}(t - x\beta)$
and $S_*(t|\mathbf{x}) = S_*(t - \mathbf{x}'\beta_*)$, where $S_* \in \mathcal{F}$ and $\beta_* \in \mathcal{R}^P$. In view of
Equation (2.3), one may write the measure w.r.t. the cdf's $F(m, s, \mathbf{x})$
($\stackrel{\text{def}}{=} F_{M, \delta, \mathbf{x}}(m, s, \mathbf{x})$) and $F_*(m, s, \mathbf{x})$ as

$$\begin{aligned} dF(m, s, \mathbf{x}) &= \mathbf{1}(s = 0)dF(m, 0, \mathbf{x}) + \mathbf{1}(s = 1)dF(m, 1, \mathbf{x}), \\ dF(m, 0, \mathbf{x}) &= S(m|\mathbf{x})dF_C(m)dF_{\mathbf{x}}(x), \\ dF(m, 1, \mathbf{x}) &= S_C(m)dF(m|\mathbf{x})dF_{\mathbf{x}}(\mathbf{x}), \\ dF_*(m, s, \mathbf{x}) &= \mathbf{1}(s = 0)S_*(m|\mathbf{x})dF_C(m)dF_{\mathbf{x}}(\mathbf{x}) \\ &\quad + \mathbf{1}(s = 1)S_C(m)dF_*(m|\mathbf{x})dF_{\mathbf{x}}(\mathbf{x}). \end{aligned} \quad (2.5)$$

In view of Equation (2.5), the Proposition 1 under the LR model is modified as follows.

Proposition 2. Let $S(t|\mathbf{x})$ be the true conditional survival function and $g(t|\mathbf{x}) = 1$. Let

$$g_*(t|\mathbf{x}) = \begin{cases} 0 & \text{if } t - \mathbf{x}'\beta \notin D_W, \\ \frac{S_*(t - |\mathbf{x}|) - S_*(t + |\mathbf{x}|)}{S(t - |\mathbf{x}|) - S(t + |\mathbf{x}|)} & \text{if } S(t - |\mathbf{x}|) - S(t + |\mathbf{x}|) > 0, \\ \frac{S'_*(t|\mathbf{x})}{S'(t|\mathbf{x})} & \text{if } S'(t|\mathbf{x}) > 0 \text{ and } S'_*(t|\mathbf{x}) \text{ exist,} \\ \limsup_{s \downarrow 0} \frac{S_*(t - s|\mathbf{x}) - S_*(t + s|\mathbf{x})}{S(t - s|\mathbf{x}) - S(t + s|\mathbf{x})} & \text{otherwise } (\frac{0}{0} \stackrel{\text{def}}{=} 0). \end{cases} \quad (2.6)$$

Then (1) $\int \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x}) + \int \ln \frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})} dF(t, 1, \mathbf{x}) \geq 0$, $\forall S_*(t|\mathbf{x})$; and

(2) the equality holds iff $S_*(t|\mathbf{x}) = S(t|\mathbf{x})$, $\forall t \in \mathcal{A}$.

Proof. Let $h_*(t, s|\mathbf{x}) = \mathbf{1}(s = 0) \frac{S_*(t|\mathbf{x})}{S(t|\mathbf{x})} + \mathbf{1}(s = 1) \frac{g_*(t|\mathbf{x})}{g(t|\mathbf{x})}$. One can treat h_* the density of $F_*(t, s|\mathbf{x})$ w.r.t. the measure $dF(t, s|\mathbf{x})$, where $F(t, s|\mathbf{x}) = F(t, s|\mathbf{x})F_{\mathbf{x}}(\mathbf{x})$. Then the density of $S(t|\mathbf{x})$ w.r.t. the measure $dF(t, s|\mathbf{x})$ is $h(t, s|\mathbf{x}) = \mathbf{1}(s = 0) \frac{S(t|\mathbf{x})}{S(t|\mathbf{x})} + \mathbf{1}(s = 1) \frac{g(t|\mathbf{x})}{g(t|\mathbf{x})} = 1$. Given $S_*(t|\mathbf{x})$, by Proposition 1, $0 \leq \int h(t, s|\mathbf{x}) \ln \frac{h(t, s|\mathbf{x})}{h_*(t, s|\mathbf{x})} dF(t, s|\mathbf{x})$, and the equality holds iff $S_*(t|\mathbf{x}) = S(t|\mathbf{x})$, $\forall t \in \mathcal{A}$. Thus $0 \leq \iint h(t, s|\mathbf{x}) \ln \frac{h(t, s|\mathbf{x})}{h_*(t, s|\mathbf{x})} dF(t, s|\mathbf{x}) dF_{\mathbf{x}}(\mathbf{x})$ and the equality holds iff $S_*(t|\mathbf{x}) = S(t|\mathbf{x})$, $\forall t \in \mathcal{A}$. \square

3. The Main Results

We establish the consistency of the MSMLE in this section. When we say that $\hat{S}(t)$ is consistent, we mean that $t \in \mathcal{A}$.

Theorem 3. *The MSMLE $(\hat{S}, \hat{\beta})$ is consistent if the identifiability conditions stated in Theorems 1 and 2 hold.*

Proof. Let Ω_o be the subset of the sample space Ω such that the empirical distribution function (edf) $\hat{F}_n(t, s, \mathbf{x})$ based on $(M_i, \delta_i, \mathbf{X}_i)$'s converges to $F(t, s, \mathbf{x})$, the cdf of (M, δ, \mathbf{X}) . It is well-known that $P(\Omega_o) = 1$. Notice that the MSMLE $(\hat{S}, \hat{\beta})$ is a function of (ω, n) , say $(\hat{S}(\cdot)(\omega, n), \hat{\beta}_n(\omega))$, where ω belongs to the sample space and n is the sample size. Hereafter, fix an $\omega \in \Omega_o$, since $\hat{\beta}(= \hat{\beta}_n(\omega))$ is a sequence of vectors in \mathcal{R}^p , there is a convergent subsequence with the limit β_* , where the components of β_* can be $\pm\infty$. For simplicity, we shall suppress (ω, n) hereafter. Moreover, \hat{S} is a sequence of bounded non-increasing functions, Helly's selection theorem ensures that given any subsequence

of \hat{S} , there exists a further subsequence which is convergent. By taking convergent subsequence, without loss of generality (WLOG), we can assume that $\hat{\beta} \rightarrow \beta_*$ and $\hat{S} \rightarrow S_* \in \mathcal{F}$. It is well-known that $\sup_{t \in \mathcal{A}} |\hat{S}_\beta(t) - S_o(t)| \rightarrow 0$ a.s..

Since $(\hat{S}, \hat{\beta})$ is the MSMLE, $\frac{1}{n} \ln l(\hat{\beta}) \geq \frac{1}{n} \ln l(\beta)$ (see Equation (2.3)), thus

$$\begin{aligned} & \int \ln \hat{S}(t|\mathbf{x}) d\hat{F}_n(t, 0, \mathbf{x}) + \int \ln(\hat{S}(t - \eta_n|\mathbf{x}) - \hat{S}(t + \eta_n|\mathbf{x})) d\hat{F}_n(t, 1, \mathbf{x}) \\ & \geq \int \ln \hat{S}_\beta(t|\mathbf{x}) d\hat{F}_n(t, 0, \mathbf{x}) + \int \ln(\hat{S}_\beta(t - \eta_n|\mathbf{x}) \\ & \quad - \hat{S}_\beta(t + \eta_n|\mathbf{x})) d\hat{F}_n(t, 1, \mathbf{x}). \quad (\text{see (2.3)}) \end{aligned}$$

The last inequality yields

$$0 \geq \int \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) + \int \ln \frac{\hat{S}_\beta(t - \eta_n|\mathbf{x}) - \hat{S}_\beta(t + \eta_n|\mathbf{x})}{\hat{S}(t - \eta_n|\mathbf{x}) - \hat{S}(t + \eta_n|\mathbf{x})} d\hat{F}_n(t, 1, \mathbf{x}). \quad (3.1)$$

By assumption, $\hat{F}_n(\cdot, \cdot, \cdot)(\omega) \rightarrow F(\cdot, \cdot, \cdot)$ on Ω_o . We shall prove in Lemmas 2 and 3 that

$$\varliminf_{n \rightarrow \infty} \int \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) \geq \int \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x}) \quad (\text{in Lemma 2}), \quad (3.2)$$

$$\varliminf_{n \rightarrow \infty} \int \ln \frac{\hat{S}_\beta(t - \eta_n|\mathbf{x}) - \hat{S}_\beta(t + \eta_n|\mathbf{x})}{\hat{S}(t - \eta_n|\mathbf{x}) - \hat{S}(t + \eta_n|\mathbf{x})} d\hat{F}_n(t, 1, \mathbf{x}) \geq \int \ln \frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})} dF(t, 1, \mathbf{x}). \quad (3.3)$$

$$\text{Then } 0 \geq \int \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x}) + \int \ln \frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})} dF(t, 1, \mathbf{x})$$

(by Equations (3.1), (3.2) and (3.3))

$$\geq 0. \quad (\text{by Proposition 2})$$

The last two inequalities imply that $\int \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x}) + \int \ln \frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})} dF(t, 1, \mathbf{x}) = 0$. Thus $S_*(t|\mathbf{x}) = S(t|\mathbf{x})$ (i.e., $S_*(t - \beta'_*\mathbf{x}) = S_o(t - \beta'\mathbf{x})$) $\forall (t, \mathbf{x}) \in D_{M, \mathbf{X}}$ by statement (2) of Proposition 2. Since the NS assumptions of the identifiability of $(S_o(t), \beta)$ in Theorems 1 and 2 hold, $(S_*(t), \beta_*) = (S_o(t), \beta)$, $\forall t \in \mathcal{A}$ by Theorems 1 and 2. Recall $P(\Omega_o) = 1$, thus the MSMLE $(\hat{S}_o(t), \hat{\beta})$ is consistent for $t \in \mathcal{A}$. \square

We shall make use of Fatou's Lemma with varying measures (see Lemma 1 below) in the proofs of Lemmas 2 and 3.

Lemma 1 (Proposition 17 in Royden [14], page 231). *Suppose that μ_n is a sequence of measures on the measurable space $(\mathcal{S}, \mathcal{B})$ such that $\mu_n(B) \rightarrow \mu(B)$, $\forall B \in \mathcal{B}$, g_n and f_n are non-negative measurable functions, and $\lim_{n \rightarrow \infty} (f_n, g_n)(\mathbf{x}) = (f, g)(\mathbf{x})$. Then*

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

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

Corollary 1. *Suppose that μ_n is a sequence of measures on the measurable space $(\mathcal{S}, \mathcal{B})$ such that $\lim_{n \rightarrow \infty} \mu_n(B) \rightarrow \mu(B)$, $\forall B \in \mathcal{B}$, f and f_n are integrable functions, $n \geq 1$.*

(1) If f_n are bounded below and $f(x) = \lim_{n \rightarrow \infty} f_n(x)$, then $\int f d\mu \leq \lim_{n \rightarrow \infty} \int f_n d\mu_n$.

(2) If f_n are bounded below then $\int \lim_{n \rightarrow \infty} f_n d\mu \leq \lim_{n \rightarrow \infty} \int f_n d\mu_n$.

Lemma 2. Under the assumptions in the proof of Theorem 3,

$$\lim_{n \rightarrow \infty} \int \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) \geq \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x})$$

(i.e., inequality (3.2) holds).

Proof. For the given $\omega \in \Omega_o$ and (S_*, β_*) in Equation (3.2), as assumed, $\hat{S}(t) \rightarrow S_*(t)$, $\hat{\beta}(\omega) \rightarrow \beta_*$, and $S_*(t)$ is continuous a.e.. $S_*(t|\mathbf{x}) = \overline{\lim}_{n \rightarrow \infty} \hat{S}(t - \mathbf{x}'\hat{\beta})$, which equals $S_*(t - \mathbf{x}'\beta_*)$ a.e. in $t - \mathbf{x}'\beta_*$ w.r.t. the measure induced by $F_{M, \delta, \mathbf{X}}$, except perhaps at the discontinuous point of S_* , say w and $S_*(t|\mathbf{x}) = S_*(w -)$, where $w = t - \mathbf{x}'\beta_*$. Let $\alpha = \sup_{t \in \mathcal{A}} S_o(t)$, where \mathcal{A} is as in Equation (2.4). Then either (1) $\alpha > 0$ or (2) $\alpha = 0$.

Suppose that $\alpha > 0$. Then $\exists n_o$ such that $\hat{S}_\beta(t) \geq \alpha/2 \forall t \in \mathcal{A}$ and $n \geq n_o$, as $\sup_{t \in \mathcal{A}} |\hat{S}_\beta(t) - S_W(t)| \rightarrow 0$ a.s.. Denote $G(t, \mathbf{x}, n) = \frac{\hat{S}(t|\mathbf{x})}{\hat{S}_\beta(t|\mathbf{x})}$,

we have

$$\overline{\lim}_{n \rightarrow \infty} G(t, \mathbf{x}, n) = \frac{S_*(t|\mathbf{x})}{S(t|\mathbf{x})} \text{ if } S(t|\mathbf{x}) > 0. \quad (3.4)$$

Let $A_k = \{(t, \mathbf{x}) : G(t, \mathbf{x}, n) \leq k, \forall n \geq n_o\}$ and $B_k = A_k \setminus A_{k-1}$, $k \geq 1$. Since $G(t, \mathbf{x}, n)$ is finite for each $n \geq n_o$, provided that $S(t|\mathbf{x}) > 0$, we have

$$\cup_{k \geq 1} B_k = \mathcal{U} \stackrel{\text{def}}{=} \{(t, \mathbf{x}) : S(t|\mathbf{x}) \geq \alpha\} \text{ (and } \int \mathbf{1}((t, \mathbf{x}) \in \cup_{k \geq 1} B_k) dF(t, s, \mathbf{x}) = 1). \quad (3.5)$$

For each $k \geq 1$, let $a_k \stackrel{\text{def}}{=} \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} \mathbf{1}((t, \mathbf{x}) \in B_k)$.

$$\begin{aligned}
 & \underline{\lim}_{n \rightarrow \infty} \int_{B_k} \ln \frac{\hat{S}_\beta(t|x)}{\hat{S}(t|x)} d\hat{F}_n(t, 0, \mathbf{x}) \\
 & \geq \int_{B_k} \underline{\lim}_{n \rightarrow \infty} \ln \left(\frac{S(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} \right) dF(t, 0, \mathbf{x}) \quad (\text{by (2) of Corollary 1} \\
 & \quad \text{as } a_1 \in [0, \infty), a_k \in [\ln(1/k), \ln(1/(k-1))], k \geq 2) \\
 & = \int_{B_k} \ln \left(\underline{\lim}_{n \rightarrow \infty} \frac{(\hat{S}_\beta(t|\mathbf{x}))}{(\hat{S}(t|\mathbf{x}))} \right) dF(t, 0, \mathbf{x}) \quad (\text{as } \ln(x) \text{ is continuous}) \\
 & = \int_{B_k} \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x}) \quad (\text{see (3.4)}) \\
 & = \int_{B_k} H \left(\frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} \right) \frac{S_*(t|\mathbf{x})}{S(t|\mathbf{x})} dF(t, 0, \mathbf{x}) \quad (\text{where } H(t) \stackrel{\text{def}}{=} t \ln t \geq -1/e) \\
 & = \int_{B_k} H \left(\frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} \right) dF_*(t, 0, \mathbf{x}) \quad (\text{see (2.5)}) \\
 & \geq \int_{B_k} (-1/e) dF_*(t, 0, \mathbf{x}) = (-1/e) \int_{B_k} 1 dF_*(t, 0, \mathbf{x}) \\
 & \geq (-1/e) \int 1 dF_*(t, s, \mathbf{x}) \geq -1/e.
 \end{aligned}$$

$$\text{That is, for } k \geq 1, \underline{\lim}_{n \rightarrow \infty} \int_{B_k} \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) \geq \int_{B_k} \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x}) \quad (3.6)$$

$$\geq -1/e. \quad (3.7)$$

$$\begin{aligned}
\text{Then } & \liminf_{n \rightarrow \infty} \int \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) \\
&= \liminf_{n \rightarrow \infty} \sum_{k \geq 1} \int_{B_k} \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) && \text{(by (3.5))} \\
&= \liminf_{n \rightarrow \infty} \int_{k \geq 1} \int_{B_k} \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) d\nu(k) \\
&&& (d\nu \text{ is a counting measure}) \\
&\geq \int_{k \geq 1} \liminf_{n \rightarrow \infty} \int_{B_k} \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) d\nu(k) \\
&&& \text{(by (1) of Corollary 1 and (3.7))} \\
&\geq \int_{k \geq 1} \int_{B_k} \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x}) d\nu(k) && \text{(by (3.6))} \\
&= \sum_{k \geq 1} \int_{B_k} \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x}) = \int \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x}). && \text{(3.8)}
\end{aligned}$$

Thus (3.2) holds if $\alpha = \sup_{t \in \mathcal{A}} S_o(t) > 0$.

Now suppose that case (2) holds, that is, $\alpha = \sup_{t \in \mathcal{A}} S_o(t) = 0$. Let \mathcal{K} be the set of all positive integers, and $\mathcal{U}_m = \{(t, \mathbf{x}) : S(t - \mathbf{x}'\beta) \in (1/(m+1), 1/m]\}$, where $m \in \mathcal{K}$.

For each $m \in \mathcal{K}$, $\exists n_m$ such that $\hat{S}_\beta(t|\mathbf{x}) > 1/(2m)$, $\forall S_o(t - \mathbf{x}'\beta) \in (1/(m+1), 1/m]$. Denote $G(t, \mathbf{x}, n) = \frac{\hat{S}(t|\mathbf{x})}{\hat{S}_\beta(t|\mathbf{x})}$ as before, but re-define

$$A_k = \{(t, \mathbf{x}) : S_o(t - \mathbf{x}'\beta) \in (1/(m+1), 1/m], G(t, \mathbf{x}, n) \leq k, \forall n \geq n_m\}, \tag{3.9}$$

and $B_k = A_k \setminus A_{k-1}$, $k \in \mathcal{K}$, we have $\bigcup_{k \geq 1} B_k = \mathcal{U}_m$. Then by a similar arguments as in proving (3.6), (3.7), and (3.8) (i.e., replacing \mathcal{U} by \mathcal{U}_m), we can show that for $m \in \mathcal{K}$,

$$\liminf_{n \rightarrow \infty} \int_{(t, \mathbf{x}) \in \mathcal{U}_m} \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) \geq \int_{(t, \mathbf{x}) \in \mathcal{U}_m} \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x}), \quad (3.10)$$

$$\geq -1/e. \quad (3.11)$$

Their proofs are relegated to Appendix. Moreover, replacing B_k by \mathcal{U}_m in the proof of Inequality (3.8), we have

$$\begin{aligned} & \liminf_{n \rightarrow \infty} \int \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) \\ &= \liminf_{n \rightarrow \infty} \sum_{m=1}^{\infty} \int_{(t, \mathbf{x}) \in \mathcal{U}_m} \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) \\ &= \liminf_{n \rightarrow \infty} \sum_{m \geq 1} \int_{(t, \mathbf{x}) \in \mathcal{U}_m} \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) d\nu(m) \\ & \hspace{15em} (d\nu \text{ is a counting measure}) \\ &\geq \int_{m \geq 1} \liminf_{n \rightarrow \infty} \int_{(t, \mathbf{x}) \in \mathcal{U}_m} \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) d\nu(m) \\ & \hspace{15em} (\text{by (1) Corollary 1 and (3.11)}) \\ &\geq \sum_{m=1}^{\infty} \int_{(t, \mathbf{x}) \in \mathcal{U}_m} \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x}) \quad (\text{by (3.10)}) \\ &= \int \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x}), \end{aligned}$$

which is (3.2) in the case that $\sup_{t \in \mathcal{A}} S_o(t) = 0$. \square

Lemma 3. *Under the assumptions set in the proof of Theorem 2, inequality (3.3) holds, i.e.,*

$$\begin{aligned} \lim_{n \rightarrow \infty} \int \ln \frac{\hat{S}_\beta(t - \eta_n | \mathbf{x}) - \hat{S}_\beta(t + \eta_n | \mathbf{x})}{\hat{S}(t - \eta_n | \mathbf{x}) - \hat{S}(t + \eta_n | \mathbf{x})} d\hat{F}_n(t, 1, \mathbf{x}) \\ \geq \int \ln \frac{g(t | \mathbf{x}, \beta)}{g_*(t | \mathbf{x})} dF(t, 1, \mathbf{x}), \end{aligned}$$

where $g(t | \mathbf{x}, \beta) \equiv 1$ and $g_*(t | \mathbf{x})$ is as in (2.6).

Proof. For the given $\omega \in \Omega_o$, $\hat{S}(t | \mathbf{x})(\omega)$ and (S_*, β_*) in the proof of Theorem 2, denote $G(t, \mathbf{x}, n) = \frac{\hat{S}(t - \eta_n | \mathbf{x}) - \hat{S}(t + \eta_n | \mathbf{x})}{\hat{S}_\beta(t - \eta_n | \mathbf{x}) - \hat{S}_\beta(t + \eta_n | \mathbf{x})}$. By (2.6),

$$\overline{\lim}_{n \rightarrow \infty} G(t, \mathbf{x}, n) = g_*(t | \mathbf{x}) = \frac{g_*(t | \mathbf{x})}{g(t | \mathbf{x}, \beta)} \text{ if } t - \mathbf{b}'\beta \in \mathcal{A} \cap \mathcal{D}_W.$$

Notice that the denominator $\hat{S}_\beta(t - \eta_n | \mathbf{x}) - \hat{S}_\beta(t + \eta_n | \mathbf{x})$ of $G(t, \mathbf{x}, n)$ can be zero if n is small. By the definition of \mathcal{D}_W , if $w \in \mathcal{D}_W \cap \mathcal{A}$, then $\hat{S}_\beta(w - \eta_n) - \hat{S}_\beta(w + \eta_n) > 0$ for n large enough. Thus $G(t, \mathbf{x}, n)$ is finite for n large enough, provided that $t - \mathbf{x}'\beta \in \mathcal{D}_W \cap \mathcal{A}$. Hence we can partition $\mathcal{D}_W \cap \mathcal{A}$ as follows. For each $m \in \mathcal{K}$, let

$$\mathcal{U}_m = \{(t, \mathbf{x}) : G(t, \mathbf{x}, n) \text{ is finite for } n > m, \text{ but not for } n = m\}.$$

We shall first prove inequalities similar to (3.6) and (3.7), and then prove (3.3).

Given $m \in \mathcal{K}$, let $A_k = \{(t, \mathbf{x}) \in \mathcal{U}_m : G(t, \mathbf{x}, n) \leq k, \forall n \geq m\}$ and $B_k = A_k \setminus A_{k-1}$, $k \geq 1$. Then $\cup_{k \geq 1} B_k = \mathcal{U}_m$.

For each $k \geq 1$, let $a_k \stackrel{\text{def}}{=} \ln \frac{\hat{S}_\beta(t - \eta_n | \mathbf{x}) - \hat{S}_\beta(t + \eta_n | \mathbf{x})}{\hat{S}(t - \eta_n | \mathbf{x}) - \hat{S}(t + \eta_n | \mathbf{x})} \mathbf{1}((t, \mathbf{x}) \in B_k)$.

$$\begin{aligned}
 & \underline{\lim}_{n \rightarrow \infty} \int_{B_k} \ln \frac{\hat{S}_\beta(t - \eta_n | \mathbf{x}) - \hat{S}_\beta(t + \eta_n | \mathbf{x})}{\hat{S}(t - \eta_n | \mathbf{x}) - \hat{S}(t + \eta_n | \mathbf{x})} d\hat{F}_n(t, 1, \mathbf{x}) \\
 & \geq \int_{B_k} \underline{\lim}_{n \rightarrow \infty} \ln \left(\frac{S_\beta(t - \eta_n | \mathbf{x}) - S_\beta(t + \eta_n | \mathbf{x})}{(\hat{S}(t - \eta_n | \mathbf{x}) - \hat{S}(t + \eta_n | \mathbf{x}))} \right) dF(t, 1, \mathbf{x}) \\
 & \hspace{25em} \text{(by (2) of Corollary 1)} \\
 & \hspace{10em} \text{as } a_1 \in [0, \infty), a_k \in [\ln(1/k), \ln(1/(k-1))], k \geq 2) \\
 & = \int_{B_k} \ln \left(\underline{\lim}_{n \rightarrow \infty} \frac{(\hat{S}_\beta(t - \eta_n | \mathbf{x}) - \hat{S}_\beta(t + \eta_n | \mathbf{x}))}{(\hat{S}(t - \eta_n | \mathbf{x}) - \hat{S}(t + \eta_n | \mathbf{x}))} \right) dF(t, 1, \mathbf{x}) \\
 & \hspace{25em} \text{(as } \ln(x) \text{ is continuous)} \\
 & = \int_{B_k} \ln \frac{g(t | \mathbf{x})}{g_*(t | \mathbf{x})} dF(t, 1, \mathbf{x}) \hspace{10em} \text{(see (2.6))} \\
 & = \int_{B_k} H \left(\frac{g(t | \mathbf{x})}{g_*(t | \mathbf{x})} \right) \frac{g_*(t | \mathbf{x})}{g(t | \mathbf{x})} dF(t, 1, \mathbf{x}) \quad \text{(where } H(t) \stackrel{\text{def}}{=} t \ln t \geq -1/e) \\
 & = \int_{B_k} H \left(\frac{g(t | \mathbf{x})}{g_*(t | \mathbf{x})} \right) dF_*(t, 1, \mathbf{x}) \hspace{10em} \text{(see (2.5))} \\
 & \geq \int_{B_k} (-1/e) dF_*(t, 1, \mathbf{x}) \geq -1/e.
 \end{aligned}$$

That is, for $k \geq 1$,

$$\begin{aligned}
 & \underline{\lim}_{n \rightarrow \infty} \int_{B_k} \ln \frac{\hat{S}_\beta(t - \eta_n | \mathbf{x}) - \hat{S}_\beta(t + \eta_n | \mathbf{x})}{\hat{S}(t - \eta_n | \mathbf{x}) - \hat{S}(t + \eta_n | \mathbf{x})} d\hat{F}_n(t, 1, \mathbf{x}) \\
 & \geq \int_{B_k} \ln \frac{g(t | \mathbf{x})}{g_*(t | \mathbf{x})} dF(t, 1, \mathbf{x}) \hspace{10em} (3.12)
 \end{aligned}$$

$$\geq -1/e. \hspace{25em} (3.13)$$

$$\begin{aligned}
&= > \lim_{n \rightarrow \infty} \int_{\mathcal{U}_m} \ln \frac{\hat{S}_\beta(t - \eta_n | \mathbf{x}) - \hat{S}_\beta(t + \eta_n | \mathbf{x})}{\hat{S}(t - \eta_n | \mathbf{x}) - \hat{S}(t + \eta_n | \mathbf{x})} d\hat{F}_n(t, 1, \mathbf{x}) \\
&= \lim_{n \rightarrow \infty} \sum_{k \geq 1} \int_{B_k} \ln \frac{\hat{S}_\beta(t - \eta_n | \mathbf{x}) - \hat{S}_\beta(t + \eta_n | \mathbf{x})}{\hat{S}(t - \eta_n | \mathbf{x}) - \hat{S}(t + \eta_n | \mathbf{x})} d\hat{F}_n(t, 1, \mathbf{x}) \\
&\hspace{25em} (\text{as } \cup_{k \geq 1} B_k = \mathcal{U}_m) \\
&= \lim_{n \rightarrow \infty} \int_{k \geq 1} \int_{B_k} \ln \frac{\hat{S}_\beta(t - \eta_n | \mathbf{x}) - \hat{S}_\beta(t + \eta_n | \mathbf{x})}{\hat{S}(t - \eta_n | \mathbf{x}) - \hat{S}(t + \eta_n | \mathbf{x})} d\hat{F}_n(t, 1, \mathbf{x}) d\nu(k) \\
&\hspace{25em} (d\nu \text{ is a counting measure}) \\
&\geq \int_{k \geq 1} \lim_{n \rightarrow \infty} \int_{B_k} \ln \frac{\hat{S}_\beta(t - \eta_n | \mathbf{x}) - \hat{S}_\beta(t + \eta_n | \mathbf{x})}{\hat{S}(t - \eta_n | \mathbf{x}) - \hat{S}(t + \eta_n | \mathbf{x})} d\hat{F}_n(t, 1, \mathbf{x}) d\nu(k) \\
&\hspace{25em} (\text{by (1) of Corollary 1 and (3.13)}) \\
&\geq \int_{k \geq 1} \int_{B_k} \ln \frac{g(t | \mathbf{x})}{g_*(t | \mathbf{x})} dF(t, 1, \mathbf{x}) d\nu(k) \hspace{5em} (\text{by (3.12)}) \\
&= \sum_{k \geq 1} \int_{B_k} \ln \frac{g(t | \mathbf{x})}{g_*(t | \mathbf{x})} dF(t, 1, \mathbf{x}) = \int_{\mathcal{U}_m} \ln \frac{g(t | \mathbf{x})}{g_*(t | \mathbf{x})} dF(t, 1, \mathbf{x}) \quad (\geq -1/e).
\end{aligned}$$

Finally,

$$\begin{aligned}
&\lim_{n \rightarrow \infty} \int \ln \frac{\hat{S}_\beta(t - \eta_n | \mathbf{x}) - \hat{S}_\beta(t + \eta_n | \mathbf{x})}{\hat{S}(t - \eta_n | \mathbf{x}) - \hat{S}(t + \eta_n | \mathbf{x})} d\hat{F}_n(t, 1, \mathbf{x}) \\
&= \lim_{n \rightarrow \infty} \sum_{m=1}^{\infty} \int_{\mathcal{U}_m} \ln \frac{\hat{S}_\beta(t - \eta_n | \mathbf{x}) - \hat{S}_\beta(t + \eta_n | \mathbf{x})}{\hat{S}(t - \eta_n | \mathbf{x}) - \hat{S}(t + \eta_n | \mathbf{x})} d\hat{F}_n(t, 1, \mathbf{x}) \\
&= \lim_{n \rightarrow \infty} \int_{m \geq 1} \int_{\mathcal{U}_m} \ln \frac{\hat{S}_\beta(t - \eta_n | \mathbf{x}) - \hat{S}_\beta(t + \eta_n | \mathbf{x})}{\hat{S}(t - \eta_n | \mathbf{x}) - \hat{S}(t + \eta_n | \mathbf{x})} d\hat{F}_n(t, 1, \mathbf{x}) d\nu(m)
\end{aligned}$$

$$\begin{aligned}
 &\geq \int_{m \geq 1} \lim_{n \rightarrow \infty} \int_{\mathcal{U}_m} \ln \frac{\hat{S}_\beta(t - \eta_n | \mathbf{x}) - \hat{S}_\beta(t + \eta_n | \mathbf{x})}{\hat{S}(t - \eta_n | \mathbf{x}) - \hat{S}(t + \eta_n | \mathbf{x})} d\hat{F}_n(t, 1, \mathbf{x}) d\nu(m) \\
 &\geq \sum_{m \geq 1} \int_{\mathcal{U}_m} \ln \frac{g(t | \mathbf{x})}{g_*(t | \mathbf{x})} dF(t, 1, \mathbf{x}) \\
 &= \int \ln \frac{g(t | \mathbf{x})}{g_*(t | \mathbf{x})} dF(t, 1, \mathbf{x}).
 \end{aligned}$$

Thus (3.3) holds. □

Remark. Even though the MSMLE is an extension of the SMLE, the SMLE is not always consistent under the identifiability condition. The reason is that $f_W(t) = P(W \in (t - \eta, t])$, where $\eta = \min\{|W_i - W_j| : i \neq j\}$. Under the continuity assumption, there is no observation within $[W - \eta, W + \eta]$. But the MSMLE is different in this regard.

References

- [1] M. G. Akritas, S. A. Murphy and M. P. Lavalley, The Theil-Sen estimator with doubly censored data and applications to astronomy, *Journal of the American Statistical Association* 90(429) (1995), 170-177.
DOI: <https://doi.org/10.2307/2291140>
- [2] P. J. Bickel, On adaptive estimation, *Annals of Statistics* 10(3) (1982), 647-671.
DOI: <https://doi.org/10.1214/aos/1176345863>
- [3] J. Buckley and I. James, Linear regression with censored data, *Biometrika* 66(3) (1979), 429-436.
DOI: <https://doi.org/10.2307/2335161>
- [4] S. Chatterjee and D. L. McLeish, Fitting linear regression models to censored data by least squares and maximum likelihood methods, *Communications in Statistics - Theory and Methods* 15(11) (1986). 3227-3243.
DOI: <https://doi.org/10.1080/03610928608829305>
- [5] M. G. Gu and T. L. Lai, Functional laws of the iterated logarithm for the product-limit estimator of a distribution function under random censorship or truncation, *Annals of Probability* 18(1) (1990), 160-189.
DOI: <https://doi.org/10.1214/aop/1176990943>

- [6] X. M. He and L. X. Zhu, A lack-of-fit test for quantile regression, *Journal of the American Statistical Association* 98(464) (2003), 1013-1022.
DOI: <https://doi.org/10.1198/016214503000000963>
- [7] P. J. Huber, Robust estimation of a location parameter, *Annals of Mathematical Statistics* 35(1) (1964), 73-101.
DOI: <https://doi.org/10.1214/aoms/1177703732>
- [8] S. L. Hillis, Extending M-estimation to include censored data via James's method, *Communications in Statistics - Simulation and Computation* 20(1) (1991), 121-128.
DOI: <https://doi.org/10.1080/03610919108812943>
- [9] M. J. Ireson and P. V. Rao, Interval estimation of slope with right-censored data, *Biometrika* 72(3) (1985), 601-608.
DOI: <https://doi.org/10.1093/biomet/72.3.601>
- [10] S. Leurgans, Linear models, random censoring and synthetic data, *Biometrika* 74(2) (1987), 301-309.
DOI: <https://doi.org/10.1093/biomet/74.2.301>
- [11] D. C. Montgomery and E. A. Peck, *Introduction to Linear Regression Analysis*, Wiley, N.Y., 1992.
- [12] Art B. Owen, *Empirical Likelihood*, Chapman & Hall, 2001.
- [13] Y. Ritov, Estimation in a linear regression model with censored data, *Annals of Statistics* 18(1) (1990), 303-328.
DOI: <https://doi.org/10.1214/aos/1176347502>
- [14] H. L. Royden, *Real Analysis*, Macmillan, N.Y., 1968.
- [15] P. K. Sen, Estimates of the regression coefficient based on Kendall's Tau, *Journal of the American Statistical Association* 63(324) (1968), 1379-1389.
- [16] H. Theil, A rank-invariant method of linear and polynomial regression analysis, *Proceedings of the Royal Netherlands Academy of Sciences* 53 (1950), 386-392.
DOI: https://doi.org/10.1007/978-94-011-2546-8_20
- [17] Q. Q. Yu and G. Y. C. Wong, Modified semiparametric maximum likelihood estimator in linear regression with complete data or right-censored data, *Technometrics* 47(1) (2005), 34-42.
DOI: <https://doi.org/10.1198/004017004000000554>
- [18] Q. Q. Yu and G. Y. C. Wong, How to find all Buckley-James estimates instead of just one?, *Journal of Statistical Computation and Simulation* 72(6) (2002), 451-460.
DOI: <https://doi.org/10.1080/00949650213701>

- [19] Q. Q. Yu and G. Y. C. Wong, Semi-parametric MLE in simple linear regression analysis with interval-censored data, *Communications in Statistics - Simulation and Computation* 32(1) (2003), 147-163.

DOI: <https://doi.org/10.1081/SAC-120013118>

- [20] Q. Q. Yu and J. Y. Dong, Identifiability conditions for the linear regression model under right censoring, *Communications in Statistics - Theory and Methods* (2019).

DOI: <https://doi.org/10.1080/03610926.2020.1743315>

- [21] Q. Q. Yu, Consistency of the semi-parametric MLE under the Cox model with right-censored data, *The Open Mathematics, Statistics and Probability Journal* 10 (2020), 21-27.

DOI: <https://doi.org/10.2174/2666148902010010021>

- [22] C. H. Zhang and X. Li, Linear regression with doubly censored data, *Annals of Statistics* 24(6) (1996), 2720-2743.

DOI: <https://doi.org/10.1214/aos/1032181177>



Appendix

Proof of Corollary 1. Let $k = \inf_n \inf_x f_n(x)$. If $k \geq 0$ then the corollary follows from Lemma 1. Otherwise, $f_n^-(x) = 0 \wedge f_n(x)$, $f_n^+(x) = 0 \vee f_n(x)$, $f^-(x) = 0 \wedge f(x)$ and $f^+(x) = 0 \vee f(x)$. Then $f_n^+ \rightarrow f^+$ and $f_n^- \rightarrow f^-$ point-wisely, as $f_n \rightarrow f$ in Case (1). Then

$$\begin{aligned} \underline{\lim}_{n \rightarrow \infty} \int f_n d\mu_n &= \underline{\lim}_{n \rightarrow \infty} \int (f_n^+ + f_n^-) d\mu_n = \underline{\lim}_{n \rightarrow \infty} \left[\int f_n^+ d\mu_n + \int f_n^- d\mu_n \right] \\ &\geq \underline{\lim}_{n \rightarrow \infty} \int f_n^+ d\mu + \underline{\lim}_{n \rightarrow \infty} \int f_n^- d\mu_n \\ &= \underline{\lim}_{n \rightarrow \infty} \int f_n^+ d\mu_n + \int \underline{\lim}_{n \rightarrow \infty} f_n^- d\mu_n \end{aligned}$$

(by statement (2) of Lemma 1, as $|f_n^-(x)| \leq k$)

$\geq \int \underline{\lim}_{n \rightarrow \infty} f_n^+ d\mu + \int f^- d\mu$ (by statement (1) of Lemma 1, as $f_n^+(x)$ is nonnegative) $= \int f^+ d\mu + \int f^- d\mu = \int (f^+ + f^-) d\mu = \int f d\mu$, i.e., statement (1) holds.

Let $g_n(x) = \inf\{f_k(x) : k \geq n\}$, then $g_n(x) \rightarrow g(x) = \underline{\lim}_{n \rightarrow \infty} f_n(x)$. We have

$$\int \underline{\lim}_{n \rightarrow \infty} f_n d\mu = \int \underline{\lim}_{n \rightarrow \infty} g_n d\mu \leq \underline{\lim}_{n \rightarrow \infty} \int g_n d\mu_n$$

(by statement (1)), as g_n is bounded below)

$$= \underline{\lim}_{n \rightarrow \infty} \int \inf\{f_k : k \geq n\} d\mu_n$$

$$\leq \underline{\lim}_{n \rightarrow \infty} \int f_n d\mu_n. \quad (\text{which is statement (2)}).$$

□

Proof of Equation (3.10) and Equation (3.11). Given $m \in \mathcal{K}$, since $G(t, \mathbf{x}, n)$ is finite $\forall (t, \mathbf{x}) \in \mathcal{U}_m$ and for each $n \geq n_0$, we have

$\cup_{k \geq 1} B_k = \mathcal{U}_m$. For each $k \geq 1$, let $a_k \stackrel{\text{def}}{=} \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} \mathbf{1}((t, \mathbf{x}) \in B_k)$.

$$\begin{aligned}
 & \lim_{n \rightarrow \infty} \int_{B_k} \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) \\
 & \geq \int_{B_k} \lim_{n \rightarrow \infty} \ln \left(\frac{S(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} \right) dF(t, 0, \mathbf{x}) \quad (\text{by (2) of Corollary 1}) \\
 & \quad \text{as } a_1 \in [0, \infty), a_k \in [\ln(1/k), \ln(1/(k-1))], k \geq 2 \\
 & = \int_{B_k} \ln \left(\lim_{n \rightarrow \infty} \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} \right) dF(t, 0, \mathbf{x}) \quad (\text{as } \ln(x) \text{ is continuous}) \\
 & = \int_{B_k} \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x}) \quad (\text{see (3.4)}) \\
 & = \int_{B_k} H \left(\frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} \right) \frac{S_*(t|\mathbf{x})}{S(t|\mathbf{x})} dF(t, 0, \mathbf{x}) \\
 & \quad \text{(where } H(t) \stackrel{\text{def}}{=} t \ln t \geq -1/e) \\
 & = \int_{B_k} H \left(\frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} \right) dF_*(t, 0, \mathbf{x}) \quad (\text{see (2.5)}) \\
 & \geq \int_{B_k} (-1/e) dF_*(t, 0, \mathbf{x}).
 \end{aligned}$$

That is, for $k \geq 1$,

$$\lim_{n \rightarrow \infty} \int_{B_k} \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) \geq \int_{B_k} \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x}) \quad (4.1)$$

$$\geq -1/e \int_{B_k} 1 dF(t, 0, \mathbf{x}). \quad (4.2)$$

Notice that the proof so far till Equation (4.1) and Equation (4.2) is identical to the proof of Equation (3.6) and Equation (3.7).

Then

$$\begin{aligned}
& \underline{\lim}_{n \rightarrow \infty} \int_{(t, \mathbf{x}) \in \mathcal{U}_m} \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) \\
&= \underline{\lim}_{n \rightarrow \infty} \sum_{k \geq 1} \int_{B_k} \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) \quad (\text{as } \cup_k B_k = \mathcal{U}_m) \\
&= \underline{\lim}_{n \rightarrow \infty} \int_{k \geq 1} \int_{B_k} \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) d\nu(k) \\
& \hspace{15em} (d\nu \text{ is a counting measure}) \\
&\geq \int_{k \geq 1} \underline{\lim}_{n \rightarrow \infty} \int_{B_k} \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) d\nu(k) \\
& \hspace{15em} (\text{by (1) of Corollary 1 and (4.2)}) \\
&\geq \int_{k \geq 1} \int_{B_k} \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x}) d\nu(k) \quad (\text{by (4.1)}) \\
&= \sum_{k \geq 1} \int_{B_k} \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x}) = \int_{(t, \mathbf{x}) \in \mathcal{U}_m} \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x}). \tag{3.6} \\
&\geq -1/e. \tag{3.7}
\end{aligned}$$

□

It is worth mentioning that the proof upto Equation (3.6) is almost the same as the proof in deriving Equation (3.8), except that \int in Equation (3.8) is replaced by $\int_{\mathcal{U}_m}$ in Equation (3.6).