Accepted by "Journal of Statistics: Advances in Theory and Applications" on 8/4/2021

Consistency Of The Modified Semi-parametric MLE Under The Linear Regression Model With Right-Censored Data

By Qiqing Yu*

Department of Mathematical Sciences, Binghamton University, NY 13902, USA

email address: qyu@math.binghamton.edu

Current version: 7/7/2021

First version: 6/1/2020

Short title: Consistency of the MSMLE.

MSC 2010 subject classification: Primary 62 N 01; Secondary 62 G08.

Key Words: Linear regression model, Modified Semi-parametric Maximum likelihood estimator, consistency, right-censored data.

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 (2005). The MSMLE $\hat{\beta}$ of β satisfying $P(\hat{\beta} \neq \beta$ infinitely often$) = 0$ if W is discontinuous, and the simulation study suggests that it is also consistent and efficient under certain regularity 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 (2019)).

1. Introduction. We shall establish the consistency of the modified semi-parametric maximum likelihood estimator (MSMLE) proposed in Yu and Wong (2005) under the linear regression model with right-censored data. We shall make the following assumptions.

(A1) Let $(M_i, \delta_i, \mathbf{X}_i)$, $i = 1, ..., n$, be i.i.d. observations from the random vector $(M, \delta, \mathbf{X})$, 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 Eucleadian 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), $i = 1, ..., n$, 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 (1950) and Sen (1968)),

(3) L-estimators and R-estimators (see, e.g., Montgomery and Peck (1992)),

(4) adaptive estimators (Bickel (1982)),

(5) various M-estimators (Huber (1964)),

(6) the quantile (or median) regression estimator (see *e.g.*, He and Zhu (2003)),

(7) the empirical likelihood estimator (Owen, Art B. (2001)),

(8) The semi-parametric maximum likelihood estimators (SMLE) (Yu and Wong (2003)), and the modified SMLE (MSMLE) (Yu and Wong (2005)).

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

$$Ł = Ł(f, b) = \prod_{i=1}^{n} f(Y_i - bX_i), \text{ where } f \text{ belongs to a class of density functions.} \quad (1.1)$$

If $f(t) = S(t-) - S(t)$, where $S(\cdot)$ is a survival function, it leads to the SMLE. If f in Eq. (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 $\frac{\partial \ln \mathbf{L}(\hat{f},b)}{\partial b}$, where $\hat{f}$ is a kernel estimate of f. Zhang and Li (1996) 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

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

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 (1979) estimator (BJE) is a modification of normal equations of the sum of least squares. Chatterjee and Mcleish (1986) and Leurgans (1987) propose several parametric and semi-parametric extensions of the LSE. Hillis (1991), Ritov (1990) and Zhang and Li (1996) consider M-estimators and their modifications. Ireson and Rao (1985) and Akritas, Murphy and Lavalley (1995) 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 (2005) 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 $\frac{d \ln \mathbf{L}(f,b)}{db} = 0$ a.e..

Yu and Wong (2005) 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) $\mathrm{P}(A) > 0$, where $A = \big\{(\mathbf{x}_1, ..., \mathbf{x}_{p+1}):$
$rank\begin{pmatrix} 1 & \cdots & 1 \\ \mathbf{x}_1 & \cdots & \mathbf{x}_{p+1}\end{pmatrix} = p+1,\ \mathbf{x}_i$'s are i.i.d. copies of $\mathbf{X}|\delta_1 = \cdots = \delta_{p+1} = 1\big\}$.

(A3) $F_o(t) = \int_{x \le t} f_o(x) d\mu(x)$, where μ is the sum of the Lebesgue measure on the real line and a counting measure on a countable set $\mathcal{M}$, 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

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

It is conjectured according to simulation results (see Yu and Wong (2005)) that under certain regularity conditions, $\hat{\beta}_n$ is efficient.

An MSMLE solution is a maximizer of $Ł(f, b)$, while an M-estimator solution is an approximation of a stationary point of $Ł(f, b)$. Even though Zhang and Li (1996) 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" $\frac{1}{n}\Phi(\hat{\phi}, b) \to E(\frac{1}{n}\Phi(\phi_o, b))$ 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 (1979) and Zhang and Li (1996)) 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 (2002, Example 4.2 and Figure 1) present such an example for the BJE, which is also an M-estimator. Because Yu and Wong (2002) 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 (2019)).

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 $\frac{d \ln \text{Ł}}{d\mathbf{b}}|_{\mathbf{b}=\hat{\beta}_n} = 0$. However, it does not work here, as $\frac{d \ln \text{Ł}}{d\mathbf{b}} = 0$ 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

$$\text{Ł}(S, \mathbf{b}, f) = \prod_{i=1}^{n} [(f(T_i(\mathbf{b})))^{\delta_i} (S(T_i(\mathbf{b})))^{1-\delta_i}], \text{ where } \ f(t) = S(t-) - S(t), \ S \in \mathcal{F}, \quad (2.1)$$

$\mathcal{F} = \{H : \ H$ is a non-increasing function on $[-\infty, \infty]$, $H(-\infty) = 1$ and $H(\infty) = 0\}$. Yu and Wong (2005) suggest to replace f by a kernel estimate,

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

(*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

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

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 $Y_1 - \mathbf{X}'_1\hat{\beta}$, ..., $Y_n - \mathbf{X}'_n\hat{\beta}$ 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 (2019) 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, $P(||T-t|| < c) > 0 \ \forall \ c > 0$ 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

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

Definition. Assume (A1) holds. (S_o, β) is said to be identifiable if

$$(S_*(t - \mathbf{x}'\beta_*) = S_o(t - \mathbf{x}'\beta) \ \forall \ t - \mathbf{x}'\beta \in \mathcal{A} \text{ implies that } (S_*(t), \beta_*) = (S_o(t), \beta) \ \forall \ t \in \mathcal{A}.$$

Theorem 1 (Yu and Dong (2019)). 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

$$\mathcal{B}_{\mathbf{x}_0} = \left\{ (w_1, \mathbf{x}_1, ..., w_p, \mathbf{x}_p) : \ \mathbf{x}_1 - \mathbf{x}_0, ..., \mathbf{x}_p - \mathbf{x}_0 \text{ are linearly independent}, \ \mathbf{x}_i \in \mathcal{D}_{\mathbf{X}}, \right.$$
$$\left. w_i \in \mathcal{D}_W \text{ and } w_i + \beta'\mathbf{x}_i \begin{cases} \leq \tau_C & \text{if } P(C = \tau_C) > 0 \\ < \tau_C & \text{otherwise} \end{cases} \right\}$$

Theorem 2 (Yu and Dong (2019)). 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

$$\mathcal{B}_{\mathbf{X}_0} = \big\{(\mathbf{x}_1, ..., \mathbf{x}_p)\text{:}\ \ \mathbf{x}_1 - \mathbf{x}_0, ..., \mathbf{x}_p - \mathbf{x}_0 \text{ are linearly independent, and } \mathbf{x}_i \in \mathcal{D}_{\mathbf{X}}\big\}.$$

Here $\mathbf{x}_0 = \mathbf{0}$ if $\mathbf{0} \in \mathcal{D}_{\mathbf{X}}$, otherwise $\mathbf{x}_0$, ..., $\mathbf{x}_p$ 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 (Yu (2020)). 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 $S(t|\mathbf{x}) = S_o(t - \mathbf{x}'\beta)$, $\hat{S}(t|\mathbf{x}) = \hat{S}_{\hat{\beta}}(t - \mathbf{x}'\hat{\beta})$, $\hat{S}_\beta(t|\mathbf{x}) = \hat{S}_\beta(t - \mathbf{x}'\beta)$ and $S_*(t|\mathbf{x}) = S_*(t - \mathbf{x}'\beta_*)$, where $S_* \in \mathcal{F}$ and $\beta_* \in \mathcal{R}^p$. In view of Eq. (2.3), one may write the measure w.r.t. the cdf's $F(m, s, \mathbf{x})$ ($\stackrel{def}{=} F_{M,\delta,\mathbf{X}}(m, s, \mathbf{x})$) and $F_*(m, s, \mathbf{x})$ as

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

In view of Eq. (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*

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

Then (1) $\int \ln\frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})}dF(t, 0, \mathbf{x}) + \int \ln\frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})}dF(t, 1, \mathbf{x}) \geq 0, \quad \forall\ S_*(t|\mathbf{x})$*; 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, $0 \le \int h(t,s|\mathbf{x})\ln\frac{h(t,s|\mathbf{X})}{h_*(t,s|\mathbf{X})}dF(t,s|\mathbf{x})$, and the equality holds iff $S_*(t|\mathbf{x}) = S(t|\mathbf{x})\ \forall\ t \in \mathcal{A}$. Thus $0 \le \int\int h(t,s|\mathbf{x})\ln\frac{h(t,s|\mathbf{X})}{h_*(t,s|\mathbf{X})}dF(t,s|\mathbf{x})dF_{\mathbf{X}}(\mathbf{x})$ and the equality holds iff $S_*(t|\mathbf{x}) = S(t|\mathbf{x})\ \forall\ t \in \mathcal{A}$. □

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 $(M, \delta, \mathbf{X})$. 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 $\hat{\beta} \to \beta_*$ and $\hat{S} \to S_* \in \mathcal{F}$. It is well known that $\sup_{t\in\mathcal{A}} |\hat{S}_\beta(t) - S_o(t)| \to 0$ a.s..

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

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

The last inequality yields

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

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

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

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

$$\begin{aligned} \text{Then } 0 \geq & \int \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t,0,\mathbf{x}) + \int \ln \frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})} dF(t,1,\mathbf{x}) \text{ (by Eq.s (3.1), (3.2) and (3.3))} \\ \geq & 0 \qquad \text{(by Proposition 2)} \end{aligned}$$

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.*, $S_*(t-\beta_*'\mathbf{x}) = S_o(t-\beta'\mathbf{x})$) $\forall$ $(t,\mathbf{x}) \in \mathcal{D}_{M,\mathbf{x}}$ 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 $t \in \mathcal{A}$. □

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 (1968), page 231). *Suppose that μ_n is a sequence of measures on the measurable space $(\mathcal{S},\mathcal{B})$ such that $\mu_n(B) \to \mu(B)$, $\forall$ $B \in \mathcal{B}$, 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) $\int f\,d\mu \leq \varliminf_{n\to\infty} \int f_n\,d\mu_n$;

(2) if $g_n \geq f_n$ (≥ 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 $(\mathcal{S},\mathcal{B})$ such that $\lim_{n\to\infty} \mu_n(B) \to \mu(B)$, $\forall$ $B \in \mathcal{B}$, f and f_n are integrable functions, $n \geq 1$.*

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

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

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

$$\varliminf_{n\to\infty} \int \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t,0,\mathbf{x}) \geq \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t,0,\mathbf{x}) \text{ (i.e. inequality (3.2) holds).}$$

Proof. For the given $\omega \in \Omega_o$ and (S_*, β_*) in Eq. (3.2), as assumed, $\hat{S}(t) \to S_*(t)$, $\hat{\beta}(\omega) \to \beta_*$, and $S_*(t)$ is continuous a.e.. $S_*(t|\mathbf{x}) = \overline{\lim}_{n\to\infty} \hat{S}(t - \mathbf{x}'\hat{\beta})$, 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 Eq. (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 $\sup_{t\in\mathcal{A}} |\hat{S}_\beta(t) - S_W(t)| \to 0$ a.s.. Denote $G(t, \mathbf{x}, n) = \frac{\hat{S}(t|\mathbf{x})}{\hat{S}_\beta(t|\mathbf{x})}$, we have

$$\overline{\lim_{n\to\infty}} G(t, \mathbf{x}, n) = \frac{S_*(t|\mathbf{x})}{S(t|\mathbf{x})} \text{ if } S(t|\mathbf{x}) > 0. \tag{3.4}$$

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

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

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

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

That is, for $k \geq 1$,

$$\underline{\lim}_{n\to\infty} \int_{B_k} \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) \geq \int_{B_k} \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x}) \tag{3.6}$$

$$\geq -1/e. \tag{3.7}$$

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

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 redefine

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

and $B_k = A_k\setminus A_{k-1}$, $k\in\mathcal{K}$, we have $\cup_{k\ge1}B_k = \mathcal{U}_m$. 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 $m\in\mathcal{K}$,

$$
\begin{aligned}
\varliminf_{n\to\infty}\int_{(t,\mathbf{x})\in\mathcal{U}_m}\ln\frac{\hat S_\beta(t|\mathbf{x})}{\hat S(t|\mathbf{x})}d\hat F_n(t,0,\mathbf{x}) &\ge \int_{(t,\mathbf{x})\in\mathcal{U}_m}\ln\frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})}dF(t,0,\mathbf{x}) && (3.10)\\
&\ge -1/e. && (3.11)
\end{aligned}
$$

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

$$
\begin{aligned}
& \varliminf_{n\to\infty}\int \ln\frac{\hat S_\beta(t|\mathbf{x})}{\hat S(t|\mathbf{x})}d\hat F_n(t,0,\mathbf{x}) \\
= & \varliminf_{n\to\infty}\sum_{m=1}\int_{(t,\mathbf{x})\in\mathcal{U}_m}\ln\frac{\hat S_\beta(t|\mathbf{x})}{\hat S(t|\mathbf{x})}d\hat F_n(t,0,\mathbf{x})
\end{aligned}
$$

$$= \varliminf_{n\to\infty} \int_{m\ge 1}^{\infty} \int_{(t,\mathbf{x})\in\mathcal{U}_m} \ln\frac{\hat S_\beta(t|\mathbf{x})}{\hat S(t|\mathbf{x})} d\hat F_n(t,0,\mathbf{x}) d\nu(m) \qquad (d\nu \text{ is a counting measure})$$

$$\ge \int_{m\ge 1} \varliminf_{n\to\infty} \int_{(t,\mathbf{x})\in\mathcal{U}_m} \ln\frac{\hat S_\beta(t|\mathbf{x})}{\hat S(t|\mathbf{x})} d\hat F_n(t,0,\mathbf{x}) d\nu(m) \qquad \text{(by (1) Coro. 1 and (3.11))}$$

$$\ge \sum_{m=1}^{\infty} \int_{(t,\mathbf{x})\in\mathcal{U}_m} \ln\frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t,0,\mathbf{x}) \qquad \text{(by (3.10)}$$

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

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

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

$$\varliminf_{n\to\infty} \int \ln\frac{\hat S_\beta(t-\eta_n|\mathbf{x})-\hat S_\beta(t+\eta_n|\mathbf{x})}{\hat S(t-\eta_n|\mathbf{x})-\hat S(t+\eta_n|\mathbf{x})} d\hat F_n(t,1,\mathbf{x}) \ge \int \ln\frac{g(t|\mathbf{x},\beta)}{g_*(t|\mathbf{x})} dF(t,1,\mathbf{x}),$$

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),

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

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

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

$$\varliminf_{n\to\infty} \int_{B_k} \ln\frac{\hat S_\beta(t-\eta_n|\mathbf{x})-\hat S_\beta(t+\eta_n|\mathbf{x})}{\hat S(t-\eta_n|\mathbf{x})-\hat S(t+\eta_n|\mathbf{x})} d\hat F_n(t,1,\mathbf{x})$$

$$\geq \int_{B_k} \underline{\lim}_{n\to\infty} \ln(\frac{(S_\beta(t-\eta_n|\mathbf{x}) - S_\beta(t+\eta_n|\mathbf{x}))}{(\hat S(t-\eta_n|\mathbf{x}) - \hat S(t+\eta_n|\mathbf{x}))}) dF(t,1,\mathbf{x}) \text{ (by (2) of Corollary 1}$$
$$\text{as } a_1 \in [0,\infty),\ a_k \in [\ln(1/k), \ln(1/(k-1))],\ k \geq 2)$$
$$= \int_{B_k} \ln(\underline{\lim}_{n\to\infty} \frac{(\hat S_\beta(t-\eta_n|\mathbf{x}) - \hat S_\beta(t+\eta_n|\mathbf{x}))}{(\hat S(t-\eta_n|\mathbf{x}) - \hat S(t+\eta_n|\mathbf{x}))}) dF(t,1,\mathbf{x}) \qquad \text{(as } \ln(x) \text{ is continuous)}$$
$$= \int_{B_k} \ln \frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})} dF(t,1,\mathbf{x}) \qquad \text{(see (2.6))}$$
$$= \int_{B_k} H(\frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})}) \frac{g_*(t|\mathbf{x})}{g(t|\mathbf{x})}) dF(t,1,\mathbf{x}) \qquad \text{(where } H(t) \overset{def}{=} t\ln t \geq -1/e)$$
$$= \int_{B_k} H(\frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})}) dF_*(t,1,\mathbf{x}) \qquad \text{(see (2.5))}$$
$$\geq \int_{B_k} (-1/e) dF_*(t,1,\mathbf{x}) \geq -1/e.$$

That is, for $k \geq 1$,

$$\underline{\lim}_{n\to\infty} \int_{B_k} \ln \frac{\hat S_\beta(t-\eta_n|\mathbf{x}) - \hat S_\beta(t+\eta_n|\mathbf{x})}{\hat S(t-\eta_n|\mathbf{x}) - \hat S(t+\eta_n|\mathbf{x})} d\hat F_n(t,1,\mathbf{x}) \geq \int_{B_k} \ln \frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})} dF(t,1,\mathbf{x}) \tag{3.12}$$
$$\geq -1/e. \tag{3.13}$$
$$=> \underline{\lim}_{n\to\infty} \int_{\mathcal{U}_m} \ln \frac{\hat S_\beta(t-\eta_n|\mathbf{x}) - \hat S_\beta(t+\eta_n|\mathbf{x})}{\hat S(t-\eta_n|\mathbf{x}) - \hat S(t+\eta_n|\mathbf{x})} d\hat F_n(t,1,\mathbf{x})$$
$$= \underline{\lim}_{n\to\infty} \sum_{k\geq 1} \int_{B_k} \ln \frac{\hat S_\beta(t-\eta_n|\mathbf{x}) - \hat S_\beta(t+\eta_n|\mathbf{x})}{\hat S(t-\eta_n|\mathbf{x}) - \hat S(t+\eta_n|\mathbf{x})} d\hat F_n(t,1,\mathbf{x}) \qquad \text{(as } \cup_{k\geq 1} B_k = \mathcal{U}_m)$$
$$= \underline{\lim}_{n\to\infty} \int_{k\geq 1} \int_{B_k} \ln \frac{\hat S_\beta(t-\eta_n|\mathbf{x}) - \hat S_\beta(t+\eta_n|\mathbf{x})}{\hat S(t-\eta_n|\mathbf{x}) - \hat S(t+\eta_n|\mathbf{x})} d\hat F_n(t,1,\mathbf{x}) d\nu(k) \quad (d\nu \text{ is a counting measure})$$
$$\geq \int_{k\geq 1} \underline{\lim}_{n\to\infty} \int_{B_k} \ln \frac{\hat S_\beta(t-\eta_n|\mathbf{x}) - \hat S_\beta(t+\eta_n|\mathbf{x})}{\hat S(t-\eta_n|\mathbf{x}) - \hat S(t+\eta_n|\mathbf{x})} d\hat F_n(t,1,\mathbf{x}) d\nu(k) \text{ (by (1) of Coro. 1 and (3.13))}$$
$$\geq \int_{k\geq 1} \int_{B_k} \ln \frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})} dF(t,1,\mathbf{x}) d\nu(k) \qquad \text{(by (3.12))}$$
$$= \sum_{k\geq 1} \int_{B_k} \ln \frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})} dF(t,1,\mathbf{x}) = \int_{\mathcal{U}_m} \ln \frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})} dF(t,1,\mathbf{x}) \qquad (\geq -1/e).$$

Finally, $$\underline{\lim}_{n\to\infty} \int \ln \frac{\hat S_\beta(t-\eta_n|\mathbf{x}) - \hat S_\beta(t+\eta_n|\mathbf{x})}{\hat S(t-\eta_n|\mathbf{x}) - \hat S(t+\eta_n|\mathbf{x})} d\hat F_n(t,1,\mathbf{x})$$
$$= \underline{\lim}_{n\to\infty} \sum_{m=1}^{\infty} \int_{\mathcal{U}_m} \ln \frac{\hat S_\beta(t-\eta_n|\mathbf{x}) - \hat S_\beta(t+\eta_n|\mathbf{x})}{\hat S(t-\eta_n|\mathbf{x}) - \hat S(t+\eta_n|\mathbf{x})} d\hat F_n(t,1,\mathbf{x})$$

$$= \varliminf_{n\to\infty} \int_{m\geq 1} \int_{\mathcal{U}_m} \ln \frac{\hat{S}_\beta(t-\eta_n|\mathbf{x}) - \hat{S}_\beta(t+\eta_n|\mathbf{x})}{\hat{S}(t-\eta_n|\mathbf{x}) - \hat{S}(t+\eta_n|\mathbf{x})} d\hat{F}_n(t,1,\mathbf{x}) d\nu(m)$$
$$\geq \int_{m\geq 1} \varliminf_{n\to\infty} \int_{\mathcal{U}_m} \ln \frac{\hat{S}_\beta(t-\eta_n|\mathbf{x}) - \hat{S}_\beta(t+\eta_n|\mathbf{x})}{\hat{S}(t-\eta_n|\mathbf{x}) - \hat{S}(t+\eta_n|\mathbf{x})} d\hat{F}_n(t,1,\mathbf{x}) d\nu(m)$$
$$\geq \sum_{m\geq 1} \int_{\mathcal{U}_m} \ln \frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})} dF(t,1,\mathbf{x})$$
$$= \int \ln \frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})} dF(t,1,\mathbf{x}). \text{Thus (3.3) holds.} \ \square$$

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.

* Akritas, M.G., Murphy S.A. and Lavalley M.P. (1995). The Theil-Sen estimator with doubly censored-data and applications to astronomy. *JASA* 90 170-177.
* Bickel, P.J. (1982). On adaptive estimation. *Ann. Statist.* 10 647-671.
* Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* 66 429-436.
* Chatterjee, S. and Mcleish, D. L. (1986). Fitting linear regression models to censored data by least squares and maximum likelihood methods. *Comm. Stat.* A 15 3227-3243.
* Gu, M.G. and Lai, T.L. (1990). Functional laws of the iterative logarithm for the product-limit estimator of a distribution function under random censorship or truncation. *Ann. Prob.* 18 160-189.
* He, X.M. and Zhu, L.X. (2003). A lack-of-fit test for quantile regression. *JASA* Vol. 98, No. 464, 1013-1022.
* Huber, P.J. (1964). Robust estimation of a location parameter". *Ann. Math. Statist.* 35 73-101.
* Hillis, S.L. (1991). Extending M-estimation to include censored data via James's method. *Comm. Stat.* B 20 121-128.

* Ireson, M. J. and Rao, P. V. (1985). Interval Estimation of Slope with Right-Censored Data. *Biometrika* 72 601-608.
* Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika* 74 301-309.
* Montegomery, D.C. and Peck, E. A. (1992). Introduction to linear regression analysis. *Wiley*, N.Y.,
* Owen, Art B. (2001). Empirical Likelihood, Chapman & Hall.
* Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* 18 303-328.
* Royden, H.L. (1968). Real analysis. Macmillan. N.Y.
* Sen, P.K. (1968). Estimates of the regression coefficient based on Kendall' tau. *JASA* 63 1379-1389.
* Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis, I. Proc. Kon. Ned. Akad. v. Wetensch. A 53 386-392.
* Yu, Q.Q. and Wong, G.Y.C. (2005). A Modified semi-parametric MLE in linear regression with complete data or right-censored data. *Technometrics* 47 34-42.
* Yu, Q.Q. and Wong, G.Y.C. (2002). How to find all Buckley-James estimates instead of just one ? *Journal of Statistical Computation and Simulation* 72 451-460.
* Yu, Q.Q. and Wong, G.Y.C. (2003). Semi-parametric MLE in simple linear regression with interval-censored data. *Comm. Statistics—Simulation Comput.*, 32 147-164.
* Yu, Q.Q. and Dong, J.Y. (2019). Identifiability Conditions For The Linear Regression Model Under Right Censoring *Comm. Statist. A—Theory and Methods* https://www.tandfonline.com/doi/full/10.1080/03610926.2020.1743315
* Yu, Q.Q. (2020). Consistency Of The Semi-parametric MLE Under The Cox Model With Right-Censored Data. *The Open Mathematics, Statistics and Probability Journal* 10 21-27.
* Zhang, C.H. and Li, X. (1996). Linear regression with doubly censored data. *Ann.*

Statist. 24 2720-2743.

4. 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, 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 $f_n^- \to f^-$ pointwisely, as $f_n \to f$ in Case (1). Then

$$\begin{aligned}
&\varliminf_{n\to\infty} \int f_n d\mu_n = \varliminf_{n\to\infty} \int (f_n^+ + f_n^-) d\mu_n = \varliminf_{n\to\infty} [\int f_n^+ d\mu_n + \int f_n^- d\mu_n] \\
&\geq \varliminf_{n\to\infty} \int f_n^+ d\mu + \varliminf_{n\to\infty} \int f_n^- d\mu_n \\
&= \varliminf_{n\to\infty} \int f_n^+ d\mu_n + \int \lim_{n\to\infty} f_n^- d\mu_n \qquad \text{(by statement (2) of Lemma 1, as } |f_n^-(x)| \leq k) \\
&\geq \int \lim_{n\to\infty} f_n^+ d\mu + \int f^- d\mu \qquad \text{(by statement (1) of Lemma 1, as } f_n^+(x) \text{ is nonnegative)} \\
&= \int f^+ d\mu + \int f^- d\mu = \int (f^+ + f^-) d\mu = \int f d\mu \quad \textit{i.e.}, \text{ statement (1) holds.}
\end{aligned}$$

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

$$\begin{aligned}
\int \varliminf_{n\to\infty} f_n \, d\mu &= \int \lim_{n\to\infty} g_n \, d\mu \leq \varliminf_{n\to\infty} \int g_n \, d\mu_n \text{ (by statement (1)), as } g_n \text{ is bounded below)} \\
&= \varliminf_{n\to\infty} \int \inf\{f_k : \ k \geq n\} \, d\mu_n \\
&\leq \varliminf_{n\to\infty} \int f_n \, d\mu_n \qquad \text{(which is statement (2)).} \ \square
\end{aligned}$$

Proof of Eq. (3.10) and Eq. (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 $n \geq n_o$, we have $\cup_{k\geq 1} B_k = \mathcal{U}_m$. For each $k \geq 1$, let $a_k \overset{def}{=} \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} \mathbf{1}((t, \mathbf{x}) \in B_k)$.

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

$$= \int_{B_k} H(\frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})}) dF_*(t, 0, \mathbf{x}) \qquad \text{(see (2.5))}$$
$$\geq \int_{B_k} (-1/e) dF_*(t, 0, \mathbf{x}).$$

That is, for $k \geq 1$,
$$\underline{\lim}_{n\to\infty} \int_{B_k} \ln \frac{\hat{S}_\beta(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t, 0, \mathbf{x}) \geq \int_{B_k} \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF(t, 0, \mathbf{x}) \qquad (4.1)$$
$$\geq -1/e \int_{B_k} 1 dF(t, 0, \mathbf{x}). \qquad (4.2)$$

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

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

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