Acta Mathematica Sinica, 2002, 18, p.405-416.

ASYMPTOTIC PROPERTIES OF A MODIFIED SEMI-PARAMETRIC MLE IN LINEAR REGRESSION WITH RIGHT-CENSORED DATA

By Qiqing Yu and George Y. C. Wong
Department of Mathematical Sciences, State University of New York, Binghamton, NY 13902, USA
email address: qyu@math.binghamton.edu
and
Strang Cancer Prevention Center, 428 E 72nd Street, New York, NY 10021, USA

AMS 1991 subject classification: Primary 62 J05; Secondary 62 G05.
Chinese books classification: O212.1
Key Words: Consistency, super-efficiency, semi-parametric MLE.

Abstract: Consider a linear regression model, $Y = \beta'\mathbf{X} + \epsilon$, where Y may be right censored and the cdf F_o of ϵ is unknown. We show that a modified semi-parametric MLE, denoted by $\hat{\beta}$, is strongly consistent under certain regularity conditions. Moreover, if F_o is discontinuous, then P($\hat{\beta} \neq \beta$ i.o.)=0, which means that P($\hat{\beta} = \beta$ if the sample size is large)=1. The latter property has not been reported for the existing estimators. In contrast, most estimators, such as the Buckley-James estimator and M-estimators $\tilde{\beta}$, satisfy that P($\tilde{\beta} \neq \beta$ i.o.)=1.

1. Introduction.

We consider asymptotic properties of a modified semi-parametric maximum likelihood estimator (MSMLE) under the linear regression model with right-censored data.

Linear 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. In particular, one desires to estimate the relationship between one variable Y and one or more independent variables, say a vector $\mathbf{X}$. Under the linear regression model, we assume that they have the following relation:

$$Y = \beta'\mathbf{X} + \epsilon, \tag{1.1}$$

where β is a $p \times 1$ regression coefficient vector, β' is the transpose of β, and ϵ is a random variable with an unknown cdf F_o and pdf f_o. This is a semi-parametric set-up, as (β, F_o) is unknown.

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 estimators for β, for example, the least squares estimator (LSE), the Theil-Sen estimator (Theil [1] and Sen [2]), adaptive estimators (Bickel [3]), various M-estimators (Huber [4] and Ritov [5]), L-estimators and R-estimators (see, e.g., Montgomery and Peck [6]).

The LSE has a simple expression and has many nice properties. Thus it is a standard procedure. But it is not efficient unless ϵ is a normal random variable, and is not robust (see Draper and Smith [7]). The Theil-Sen estimator is the median of the set $\mathcal{B}_o = \{\frac{Y_i - Y_j}{X_i - X_j} : i < j \leq n, X_i \neq X_j\}$. The Theil-Sen estimator and Huber's M-estimator are robust but not efficient. Adaptive estimators are Newton-Raphson type corrections

Received April 23, 2001. Accepted November 13, 2001
Supported by the U.S. Army Grants DAMD17-99-1-9390 and DAMD17-0 1-0448.

of initial root-n consistent estimators. The R-estimators are based on ranks and the L-estimators are based on order statistics. They are not efficient.

Various M-estimators have been proposed to improve the LSE on efficiency or robustness. Zhang and Li [8] consider a class of M-estimators. Their M-estimator of β is a zero-crossing point of a function $\Phi(b,\hat{\phi})$, where

$$\Phi(b,\phi)=\sum_{i=1}^{n}\phi(Y_i-\overline{Y}-b(X_i-\overline{X}))(X_i-\overline{X}) \text{ with } \phi(t)=\frac{f_o'(t)}{f_o(t)}$$

(f_o' is the derivative of f_o and $\hat{\phi}$ is an estimator of ϕ). Zhang and Li [8] show that under certain regularity condition the M-estimator is efficient.

An MSMLE of (β, f_o) is a value of (b, f) that maximizes the likelihood function

$$Ł=\prod_{i=1}^{n} f(Y_i-bX_i), \text{ where } f \text{ belongs to a class of pdf's.} \tag{1.2}$$

The MSMLE has some appealing properties, as to be demonstrated in the paper. However, it has not been mentioned in the literature. In this paper, we study an MSMLE where f in (1.2) belongs to a special class to be specified in Section 2. The MSMLE is akin to the above M-estimators, in the sense that it is a zero point of $\Phi(b,\tilde{\phi})$ with a $\tilde{\phi}(t)=\frac{\frac{d\hat{f}(t)}{dt}}{\hat{f}(t)}$, where $\hat{f}$ is an estimate of f_o and $\Phi(b,\phi)$ is the derivative of lnL after centralizing (Y_i,X_i)'s. However, $\hat{f}$ is a step function and thus $\tilde{\phi}=0$ a.e. (see Remark 2.1). Consequently, the M-estimator approach is non-informative to the MSMLE considered in this paper.

It is also worth mentioning that the MSMLE approach allows $E(\Phi(b,\phi))=0$ a.e. (in b). We shall prove in this paper that the MSMLE is consistent even if $E(\Phi(b,\phi))=0$ a.e.. On the contrary, if $E(\Phi(b,\phi))=0$ a.e., then one expects that $\frac{1}{n}\Phi(b,\phi)\to 0$ for each values of b, thus each value of b is possible to be an M-estimate considered by Zhang and Li. Consequently, their M-estimation approach is meaningless. It is worth mentioning that the uniform distribution, among others, satisfies $E(\Phi(b,\phi))=0$ a.e..

Under right censoring, there are several extensions of the above estimators. Miller's ([9], p.146) estimator is an extension of the least squares principle. The Buckley-James [10] estimator (BJE) is a modification of normal equations of the sum of least squares. Chatterjee and Mcleish [11] and Leurgans [12] propose several parametric and semi-parametric extensions of the least squares estimator (LSE). Hillis [13] and Zhang and Li [8] consider extensions of M-estimators. Tsiatis [14] and Wei, *et al.* [15] consider extensions of R-estimators and L-estimators, respectively. Akritas *et al.* [16], among others, consider generalizations of the Theil-Sen estimator. For more references, we refer to Moon [17].

In this paper, we shall prove the consistency of the MSMLE $\hat{\beta}$ with right-censored data under a set of regularity assumptions specified in Section 3. The main assumption is that

$$F_o(t)=\int_{x\le t} f_o(x)d\mu(x), \text{ where } \mu \text{ is a measure.}$$

It allows F_o be continuous or discontinuous. We also show that if F_o has a discontinuity point, then

$$P\{\hat{\beta}\neq\beta \text{ infinitely often (i.o.)}\}=0. \tag{1.3}$$

(1.3) implies that P$\{\hat{\beta}=\beta$ if the sample is large enough$\}=1$. We say that an estimator is super efficient if (1.3) holds. Super efficiency was not reported for the existing estimators in the literature. In contrast, unless ϵ is a constant, most estimators $\tilde{\beta}$, *e.g.*, the LSE and the BJE, satisfy that

$$P\{\tilde{\beta}\neq\beta \text{ i.o. }\}=1.$$

The paper is organized as follows. In Section 2, we introduce the MSMLE $\hat{\beta}$. In Section 3, we prove the asymptotic properties of the MSMLE. In Section 4, we give the proofs of the lemmas needed in Section 3.

2. The MSMLE.

We shall introduce notation and the definition of the MSMLE in this section. Suppose that Z is a random right censoring variable. The observable random vector is $(M,\delta,\mathbf{X})$, where $M=\min(Y,Z)$, $\delta=\mathbf{1}_{(Y\le Z)}$ and

$\mathbf{1}_A$ is the indicator function of an event A. Let $(M_i, \delta_i, \mathbf{x}_i)$, $i = 1, ..., n$, be i.i.d. copies of $(M, \delta, \mathbf{X})$. Denote $S_o = 1 - F_o$. For $\mathbf{b} \in \mathcal{R}^p$, denote $T(\mathbf{b}) = M - \mathbf{b}'\mathbf{X}$ and

$$T_i = T_i(\mathbf{b}) = M_i - \mathbf{b}'\mathbf{x}_i. \tag{2.1}$$

Then $(T_i(\beta), \delta_i)$'s are i.i.d. copies of $(T(\mathbf{b}), \delta)$. The likelihood function is

$$\text{Ł}(S, \mathbf{b}, f) = \prod_{i=1}^n [(f(T_i(\mathbf{b})))^{\delta_i}(S(T_i(\mathbf{b})))^{1-\delta_i}], \text{ where } S \in \mathcal{F}, \ f \text{ is a pdf of } S, \tag{2.2}$$

and $\mathcal{F} = \{H: \ H$ is a non-increasing function from $\mathcal{R}$ to $\mathcal{R}$, $H(-\infty) = 1$, $H(\infty) = 0\}$.

Both S and f in (2.2) are unknown. A standard approach is to set $f(t) = S(t-) - S(t)$, as is the case in the non-parametric model (with $\mathbf{b} = 0$). Then the MSMLE should be called the semi-parametric maximum likelihood estimator (SMLE). The SMLE is studied in an unpublished paper (see Yu and Wong [18]). The properties of the SMLE under a continuous distribution F_o is not quite attractive. If we allow ϵ be continuous, it is desirable to use a kernel estimator to estimate f_o. In this paper, we choose

$$f(x) = f_S(x) = \int \frac{1}{h_n} K(\frac{x-t}{h_n}) dF(t), \text{ where } K(x) = \tfrac{1}{2}\mathbf{1}_{(-1<x\le 1)}, \tag{2.3}$$

$F = 1 - S$, $\int = \int_{-\infty}^{\infty}$ and $h_n > 0$ with $\lim_{n\to\infty} h_n = 0$ (*e.g.*, $h_n = cn^{-1/5}$ with a constant c).
Then (2.2) becomes

$$\begin{aligned}\text{Ł}(S, \mathbf{b}, f_S) &= \prod_{i=1}^n [f_S(T_i(\mathbf{b}))]^{\delta_i}(S(T_i(\mathbf{b})))^{1-\delta_i}] \\ &= \prod_{i=1}^n [\frac{1}{2h_n}[S(T_i(\mathbf{b}) - h_n) - S(T_i(\mathbf{b}) + h_n)]^{\delta_i}(S(T_i(\mathbf{b})))^{1-\delta_i}], \ S \in \mathcal{F}. \end{aligned} \tag{2.4}$$

We call a value of $(S, \mathbf{b})$ that maximizes Ł in (2.4) a *modified* SMLE (MSMLE) of (S_o, β).

Now the likelihood function Ł in (2.4) can be viewed as one based on interval-censored data (L_i, R_i), $i = 1, ..., n$, where

$$(L_i, R_i) = (L_i(\mathbf{b}), R_i(\mathbf{b})) = \begin{cases} (T_i(\mathbf{b}) - h_n, T_i(\mathbf{b}) + h_n) & \text{if } \delta_i = 1, \\ (T_i(\mathbf{b}), \infty) & \text{if } \delta_i = 0, \end{cases} \ i = 1, ..., n. \tag{2.5}$$

It is well known that given $\mathbf{b}$, Ł is maximized by $\hat{S}_{\mathbf{b}}$, the generalized maximum likelihood estimator (GMLE) of S_o based on interval-censored data. Thus it suffices to maximizes $\text{Ł}(\hat{S}_{\mathbf{b}}, \mathbf{b}, f_{\hat{S}_{\mathbf{b}}})$ over $\mathbf{b} \in \mathcal{R}^p$. Denote $\hat{\beta}$ or $\hat{\beta}_n$ the MSMLE of β.

Remark 2.1. *It is well known that the GMLE $\hat{S}_{\mathbf{b}}(\cdot)$ is a step function (see, e.g., Yu et al. [19]). Since $f_{\hat{S}_{\mathbf{b}}}(t) = \frac{\hat{S}_{\mathbf{b}}(t-h_n) - \hat{S}_{\mathbf{b}}(t+h_n)}{2h_n}$, $f_{\hat{S}_{\mathbf{b}}}$ is also a step function of t.*

Theorem 2.1. *The range of the likelihood function $L(\hat{S}_{\mathbf{b}}, \mathbf{b}, f_{\hat{S}_{\mathbf{b}}})$, $\mathbf{b} \in \mathcal{R}^p$, is a set consists of finitely many elements. Thus the MSMLE of β exists.*

Proof. For each $\mathbf{b}$, there are $2n$ endpoints $L_i(\mathbf{b})$'s and $R_i(\mathbf{b})$'s. We can order these $2n$ numbers and their ranks depends on $\mathbf{b}$. It follows from (2.4) and (2.5) that

$$\text{Ł}(\hat{S}_{\mathbf{b}}, \mathbf{b}, f_{\hat{S}_{\mathbf{b}}}) = \prod_{i=1}^n [\hat{S}_{\mathbf{b}}(L_i(\mathbf{b})) - \hat{S}_{\mathbf{b}}(R_i(\mathbf{b}))](2h_n)^{-\delta_i}.$$

It is well known that the values $\hat{S}_{\mathbf{b}}(L_i(\mathbf{b})) - \hat{S}_{\mathbf{b}}(R_i(\mathbf{b}))$, $i = 1, ..., n$, only depend on the ranks of $L_i(\mathbf{b})$ and $R_i(\mathbf{b})$, $i = 1, ..., n$ (see Yu *el al.* [19]). Given n, the rank of $L_i(\mathbf{b})$ or $R_i(\mathbf{b})$, as a function of i and

$\mathbf{b}$, can only take on finitely many values, thus there are only finitely many distinct values of $\mathrm{L}(\hat{S}_{\mathbf{b}}, \mathbf{b}, f_{\hat{S}_{\mathbf{b}}})$. Consequently, the MSMLE of β exists. □.

The MSMLE of β may not be unique. Moreover, for each MSMLE $\hat{\beta}$, $\hat{S}_{\hat{\beta}}$ is an MSMLE of S_o. In view of Theorem 2.1, it is easy to construct an algorithm for obtaining an MSMLE's of β. It particular, an algorithm will contain the following steps.

1. Denote the set of all distinct values of $\mathrm{L}(\hat{S}_{\mathbf{b}}, \mathbf{b}, f_{\hat{S}_{\mathbf{b}}})$ by $\{v_1, ..., v_k\}$. Find, for each v_l, a representative $\mathbf{b}_l$;
2. For each $\mathbf{b}_l$, use self-consistent algorithm or other algorithms to compute the GMLE $\hat{S}_{\mathbf{b}_l}$ based on $(L_i(\mathbf{b}_l), R_i(\mathbf{b}_l))$'s (see, e.g., Groeneboom and Wellner [20]) and compute $v_l = \mathrm{L}(\hat{S}_{\mathbf{b}_l}, \mathbf{b}, f_{\hat{S}_{\mathbf{b}_l}})$;
3. Find a value among $\mathbf{b}_l$'s that maximizes these v_1, ..., v_k.

In order to find a $\mathbf{b}_l$, it is important to notice that the $L_i(\mathbf{b})$'s and $R_i(\mathbf{b})$'s change their ranks at a value of $\mathbf{b}$ that is the solution of $L_i(\mathbf{b}) = L_j(\mathbf{b})$, or $R_i(\mathbf{b}) = R_j(\mathbf{b})$, or $L_i(\mathbf{b}) = R_j(\mathbf{b})$. Or equivalently,

$$M_i - \mathbf{b}'\mathbf{x}_i \pm h_n = M_j - \mathbf{b}'\mathbf{x}_j \pm h_n, \text{ or } M_i - \mathbf{b}'\mathbf{x}_i \pm h_n = M_j - \mathbf{b}'\mathbf{x}_j. \tag{2.6}$$

Thus each solution to a set of p linearly independent equations of form (2.6) is a candidate of $\mathbf{b}_l$. There are only finitely many such sets of linearly independent equations. Yu and Wong [21] propose a practical algorithm for an MSMLE and carry out simulation study to investigate properties of the MSMLE. Thus we refer the details on the algorithm for deriving an MSMLE and for deriving its asymptotic variance to their paper. Their simulation results suggests that the MSMLE is efficient under the normal distribution and the exponential distribution, and is consistent under uniform distributions, while Zhang and Li's M-estimator can be inconsistent under uniform distributions.

3. The Main Result.

We shall prove the consistency and the super efficiency of the MSMLE in this section. In order to estimate (S_o, β), we need the following identifiable conditions:

A1. ϵ and $(\mathbf{X}, Z)$ are independent.

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

A1 is an assumption used by several authors under regression models (see for example, Akritas *et al.* [16]). In addition, some authors assume that the censoring variable Z also follows a regression model, that is, the distribution of $Z - \beta'\mathbf{X}$ does not depend on β (see Sen [2]). It is easy to understand A2 when $p = 1$ with discrete complete data. Then we need at least two distinct values of $\mathbf{X}$, say z_1 and z_2, to identify β. Thus $\mathrm{rank}(\begin{pmatrix} 1 & z_1 \\ 1 & z_2 \end{pmatrix}) = 2$.

Our consistency proof relies on the following assumptions:

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

A4. $|E(\ln f_o(\epsilon))| < \infty$,
$F_o(t+x) - F_o(t-x) = f_o(t)(1 + O(x))$ uniformly for all $t \in \mathcal{C}$,
$\frac{F_o(t+x) - F_o(t-x)}{2x} = f_o(t)(1 + O(x))$ uniformly for all $t \notin \mathcal{C}$.

Discrete or continuous distributions in the exponential family and the uniform distributions satisfy A3 and A4, as $\mathcal{C}$ can be empty.

Note that the MSMLE of β is a value of $\mathbf{b}$ that maximizes

$$\mathcal{L}(\hat{S}_{\mathbf{b}}, \mathbf{b}) = \prod_{i=1}^{n} [(\eta(\hat{S}_{\mathbf{b}}, M_i, \mathbf{x}_i, \mathbf{b}))^{\delta_i} (\hat{S}_{\mathbf{b}}(T_i(\mathbf{b})))^{1-\delta_i}], \tag{3.1}$$

where

$$\begin{aligned} &\eta(S, y, \mathbf{x}, \mathbf{b}) = \eta_d(S, y, \mathbf{x}, \mathbf{b}) + \eta_c(S, y, \mathbf{x}, \mathbf{b}), \\ &\eta_d(S, y, \mathbf{x}, \mathbf{b}) = [S(y - \mathbf{b}'\mathbf{x} - h_n) - S(y - \mathbf{b}'\mathbf{x} + h_n)]\mathbf{1}_{(y - \beta'\mathbf{X} \in \mathcal{C})}, \text{ and} \\ &\eta_c(S, y, \mathbf{x}, \mathbf{b}) = \frac{1}{2h_n}[S(y - \mathbf{b}'\mathbf{x} - h_n) - S(y - \mathbf{b}'\mathbf{x} + h_n)]\mathbf{1}_{(y - \beta'\mathbf{X} \notin \mathcal{C})}, \end{aligned}$$

as $\mathcal{L}(S,\mathbf{b})/\mathrm{L}(S,\mathbf{b},f_S)=\prod_{i=1}^n(2h_n)^{\delta_i\mathbf{1}_{(T_i(\beta)\in\mathcal{C})}}$ is a constant which does not depend on $\mathbf{b}$. Note that η, η_c and η_d are functions of h_n. Hereafter, denote $\mu_{M,\mathbf{X}}$ the measure induced by the cdf of the random vector$(M,\mathbf{X})$, etc..

Theorem 3.1. *Under assumptions A1- A4, the MSMLE satisfies the following: (1) $\hat\beta \to \beta$ a.s., (2) $P\Big(\lim_{n\to\infty}\eta(\hat S_{\hat\beta},y,\mathbf{x},\hat\beta)=f_o(y-\beta'\mathbf{x})\ a.s.\mu_{M,\mathbf{X}}\Big)=1$ and (3) $P\Big(\lim_{n\to\infty}\hat S_{\hat\beta}(y-\hat\beta'\mathbf{x})=S_o(y-\beta'\mathbf{x})\ a.s.\mu_{M,\mathbf{X}}\Big)=1$.*

Remark 3.1. *Define $\hat f_n(t)=\eta(\hat S_{\hat\beta},y,\mathbf{x},\hat\beta)$ if $t=y-\hat\beta_n'\mathbf{x}$, then statement (2) implies that $\hat f_n(t)\to f_o(t)$ on the range of $T(\beta)$, denoted by $\mathcal{R}_{T(\beta)}$*

Proof of Theorem 3.1. Denote $h_i(S,\mathbf{b})=\delta_i\ln\eta(S,M_i,\mathbf{x}_i,\mathbf{b})+(1-\delta_i)\ln S(M_i-\mathbf{b}'\mathbf{x}_i)$. The normalized log likelihood function is

$$\mathcal{N}(S,\mathbf{b})=\frac{\ln\mathcal{L}(S,\mathbf{b})}{n}=\frac{1}{n}\sum_{i=1}^n h_i(S,\mathbf{b})=\sum_{i=1}^n\frac{\delta_i\ln\eta(S,M_i,\mathbf{x}_i,\mathbf{b})}{n}+\sum_{i=1}^n\frac{(1-\delta_i)\ln S(M_i-\mathbf{b}'\mathbf{x}_i)}{n}.$$

Given $S\in\mathcal{F}$, define a function $f=D(S)$ by

$$f(t)=D(S)(t)=\begin{cases}\sup_{y\to 0+}\frac{S(t-y)-S(t+y)}{2y} & \text{if } t\notin\mathcal{C},\\ S(t-)-S(t+) & \text{if } t\in\mathcal{C}.\end{cases}$$

Note that $D(S)(t)$ is the usual derivative of F if F is differentiable at t and $t\notin\mathcal{C}$. Denote $\mathcal{G}=\{(g,S):\ 0\le g\le D(S), S\in\mathcal{F}\}$. Let

$$\nu(S,\mathbf{b},f)=\int\int\ln f(y-\mathbf{b}'\mathbf{x})dQ(y,\mathbf{x})+\int\int\ln S(y-\mathbf{b}'\mathbf{x})dG(y,\mathbf{x}),\ \text{where} \tag{3.2}$$

$dG(y,\mathbf{x})=S_o(y-\beta'\mathbf{x})dF_{Z,\mathbf{X}}(y,\mathbf{x})$, $dQ(y,\mathbf{x})=f_o(y-\beta'\mathbf{x})dQ_o(y,\mathbf{x})$, $dQ_o(y,\mathbf{x})=d\mu(y-\beta'\mathbf{x})\int_{y\le c}dF_{Z,\mathbf{X}}(c,\mathbf{x})$, $(f,S)\in\mathcal{G}$, and $F_{Z,\mathbf{X}}$ is the cdf of $(Z,\mathbf{X})$. Here and after, we define $\ln 0=-\infty$, $0\ln 0=0$ and $\ln\infty=\infty$. Then

$$\begin{aligned}&\lim_{n\to\infty}\mathcal{N}(S_o,\beta)\\ =&\lim_{n\to\infty}\left\{\frac{1}{n}\sum_{i=1}^n\delta_i\ln[f_o(T_i(\beta))(1+O(h_n))]+\frac{1}{n}\sum_{i=1}^n(1-\delta_i)\ln S_o(T_i(\beta))\right\}\ \text{(by A4)}\\ =&\lim_{n\to\infty}\left\{\frac{1}{n}\sum_{i=1}^n\delta_i\ln f_o(T_i(\beta))+O(h_n)+\frac{1}{n}\sum_{i=1}^n(1-\delta_i)\ln S_o(T_i(\beta))\right\}\\ =&\nu(S_o,\beta,f_o)\ \text{with probability one (by the strong law of large number).}\end{aligned}\tag{3.3}$$

Let Ω be the event that $\lim_{n\to\infty}\mathcal{N}(S_o,\beta)=\nu(S_o,\beta,f_o)$. Then $\mathrm{P}(\Omega)=1$ by (3.3). For each $\omega\in\Omega$, let $\mathbf{b}^*=\mathbf{b}^*(\omega)$ be a limiting point of $\hat\beta$, and let $S^*=S^*(\cdot,\omega)$ be a point-wise limit of $\hat S_{\hat\beta}$ in the sense that there exist a subsequence, say $\{n_k\}_{k\ge1}$ such that

$$\lim_{n_k\to\infty}\hat\beta_{n_k}=\mathbf{b}^*,\quad \lim_{n_k\to\infty}\hat S_{\hat\beta_{n_k}}(y-\hat\beta_{n_k}'\mathbf{x})=S^*(y-\mathbf{b}^{*\prime}\mathbf{x}),\ \forall\ (y,\mathbf{x})\in\mathcal{R}^{p+1}, \tag{3.4}$$

$$(f^*,S^*)\in\mathcal{G},\ \text{where } f^*(y-\mathbf{b}^{*\prime}\mathbf{x})=\lim\sup_{n_k\to\infty}\eta(\hat S_{\hat\beta_{n_k}},y,\mathbf{x},\hat\beta_{n_k}), \tag{3.5}$$

and the coordinates of $\mathbf{b}^*$ may be $\pm\infty$. The existence of S^* and f^* are proved later in Lemma 4.1 (see Section 4) due to Helly's Lemma. Without loss of generality (WLOG), we can assume that $\{n_k\}=\{n\}$.

Note that $|\nu(S_o,\beta,f_o)|\le E(\ln f_o(\epsilon))+1<\infty$ by A4. It is proved in Lemma 4.2 that

$$\nu(S,\mathbf{b},g)<\nu(S_o,\beta,f_o),\ \forall\ (g,S)\in\mathcal{G}\ \text{ and }\forall\ \mathbf{b} \tag{3.6}$$

$$\text{unless } \mathbf{b}=\beta \ \text{ and } S(y-\mathbf{b}'\mathbf{x})=S_o(y-\beta'\mathbf{x})\ a.s.\mu_{M,\mathbf{X}}\ \text{with } g=D(S_o); \tag{3.7}$$

$$\nu(S^*,\mathbf{b}^*,f^*)\ge\nu(S_o,\beta,f_o),\ \text{ where } (f^*,S^*)\in\mathcal{G}\ \ \text{(by (3.5))}. \tag{3.8}$$

It follows from (3.6), (3.7) and (3.8) that $\mathbf{b}^* = \beta$, $S^*(y-\mathbf{b}^{*\prime}\mathbf{x}) = S_o(y-\beta'\mathbf{x})$ $a.s.$ $\mu_{M,\mathbf{X}}$, and $f^*(y-\mathbf{b}^{*\prime}\mathbf{x}) = f_o(y-\beta'\mathbf{x})$ $a.s.\mu_{M,\mathbf{X}}$, as $(f^*, S^*) \in \mathcal{G}$. Thus

$$\begin{aligned}
&\lim_{n\to\infty} \hat\beta_n = \beta, \\
&\lim_{n\to\infty} \hat S_{\hat\beta_n}(y-\hat\beta_n'\mathbf{x}) = S_o(y-\beta'\mathbf{x}) \ a.s.\mu_{M,\mathbf{X}}, \\
&\lim_{n\to\infty} \eta(\hat S_{\hat\beta_n}, y, \mathbf{x}, \hat\beta_n) = f_o(y-\beta'\mathbf{x}) \ a.s.\mu_{M,\mathbf{X}}.
\end{aligned} \tag{3.9}$$

The theorem then follows from the above three equations, and the arbitrariness of $\omega \in \Omega$ □

Before we present the second theorem, we define $\tau = \sup\{t: \ P\{Z - \beta'\mathbf{X} < t\} < 1\}$.

Theorem 3.2. *Suppose that assumptions A1, A2, A3 and A4 hold. If F_o has a discontinuity point $t_0 < \tau$, then $P\{\hat\beta_n \neq \beta \ i.o.\} = 0$.*

Proof. Let Ω_1 be the event that statements (1), (2) and (3) of Theorem 3.1 hold. Note that $\hat\beta_n$ and $\hat S_{\hat\beta_n}$ both depend on ω, an element in the sample space. Hereafter, fix an $\omega \in \Omega_1$ and abusing notation, write $\hat\beta_n = \hat\beta_n(\omega)$ and $\hat S_{\hat\beta_n}(\cdot) = \hat S_{\hat\beta_n(\omega)}(\cdot,\omega)$. If $\delta = 1$, then $Y - \mathbf{b}'\mathbf{X} = \epsilon + (\beta - \mathbf{b})'\mathbf{X}$. Letting $\mathcal{R}_{\mathbf{X}}$ be the range of $\mathbf{X}$. Since $\hat S_{\hat\beta_n}$ and S_o are both monotone and bounded, it follows from (3.9) that

$$\hat S_{\hat\beta_n}(t + (\beta - \hat\beta_n)'\mathbf{x}) \to S_o(t) \text{ for all } \mathbf{x} \in \mathcal{R}_{\mathbf{X}} \text{ and for all } t < \tau. \tag{3.10}$$

Now assume that for the fixed $\omega \in \Omega_1$, $\hat\beta_n \neq \beta$ i.o.. Let $\{n_k\}_{k\ge1}$ be the subsequence that $\hat\beta_{n_k} \neq \beta$. We shall show that it leads to a contradiction.

By A1 and A2, we can assume that there exist $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{R}_{\mathbf{X}}$ such that

$$(\beta - \hat\beta_n)'\mathbf{x}_1 < (\beta - \hat\beta_n)'\mathbf{x}_2, \tag{3.11}$$

for infinitely many $n \in \{n_k\}_{k\ge1}$. By taking a subsequence, WLOG, we can assume that (3.11) holds for each n. Since $\hat S_{\hat\beta_n}(t)$ is decreasing in t and is bounded, (3.9) and (3.10) actually imply that

$$\hat S_{\hat\beta_n}(t + (\beta - \hat\beta_n)'\mathbf{x}_i) \to S_o(t) \text{ uniformly on } t \le \tfrac{t_0+\tau}{2} \text{ for } i = 1, 2. \tag{3.12}$$

By assumption, t_0 is a discontinuity point of S_o and $t_0 < \tau$. Then $S_o(t) \ge S_o(t_0) + f_o(t_0)$ if $t < t_0$. Since $t_0 \le \frac{t_0+\tau}{2}$, it follows from (3.12) that there exists an n_0 such that if $n \ge n_0$ then

$$\hat S_{\hat\beta_n}(t_0 + (\beta - \hat\beta_n)'\mathbf{x}_1) < S_o(t_0) + f_o(t_0)/2; \tag{3.13}$$

$$\begin{aligned}
\hat S_{\hat\beta_n}(t + (\beta - \hat\beta_n)'\mathbf{x}_2) &> S_o(t) - f_o(t_0)/3 \\
&> S_o(t_0) + f_o(t_0)/2 \text{ uniformly for all } t < t_0.
\end{aligned} \tag{3.14}$$

Let $t + (\beta - \hat\beta_n)'\mathbf{x}_2 = t_0 + (\beta - \hat\beta_n)'\mathbf{x}_1$ in (3.14). Thus $t < t_0$ by (3.11). Then (3.14) yields

$$\hat S_{\hat\beta_n}(t_0 + (\beta - \hat\beta_n)'\mathbf{x}_1) > S_o(t_0) + f_o(t_0)/2. \tag{3.15}$$

Inequality (3.15) contradicts (3.13). This concludes the proof of the theorem. □

4. Proofs Of Lemmas.

We shall prove the lemmas needed in the proof of Theorem 3.1. We assume that assumptions A1 through A4 hold in this section.

Lemma 4.1. *Equations (3.4) and (3.5) hold.*

Proof. Recall that $\hat{\beta} \to \mathbf{b}^*$ by assumption, where the coordinates of $\mathbf{b}^*$ maybe $\pm\infty$. Denote $H_n(y, \mathbf{x}) = \hat{S}_{\hat{\beta}}(y - \hat{\beta}'\mathbf{x})$. WLOG, we can assume that $\mathbf{b}^*$ is finite. Otherwise, we shall prove an easier statement that $H_n(y, \mathbf{x}) \to$ a constant for a subsequence of $\{n\}$.

Since H_n is a bounded function, by Helly's Lemma, given any arbitrary countable subset of $\mathcal{R}^{p+1}$, there exist a subsequence $\{n_k\}_{k\geq 1}$ such that $H_{n_k}(y, \mathbf{x})$ converges on the subset. Let D be a subset of $\mathcal{R}^{p+1}$ such that all the coordinates of $(y, \mathbf{x}) \in D$ are rational and assume, WLOG, that H_n converges to a function H on D. We shall extend H to $\mathcal{R}^{p+1}$. Note that $H_n(y, \mathbf{x})$ satisfies that H_n is a function of $y - \hat{\beta}'\mathbf{x}$ and $H_n(y, \mathbf{x}) \leq H_n(c, \mathbf{d})$ if $y - \hat{\beta}'\mathbf{x} \geq c - \hat{\beta}'\mathbf{d}$. Since $\hat{\beta} \to \mathbf{b}^*$ and $H_n \to H$ on D, $H(y, \mathbf{x}) \leq H(c, \mathbf{d})$ if $y - \mathbf{b}^{*\prime}\mathbf{x} \geq c - \mathbf{b}^{*\prime}\mathbf{d}$, where $(y, \mathbf{x}), (c, \mathbf{d}) \in D$, and H is a function of $y - \mathbf{b}^{*\prime}\mathbf{x}$. Now, define $S^*(t) = H(y, \mathbf{x})$ for $t = y - \mathbf{b}^{*\prime}\mathbf{x}$. Then S^* is defined on a subset of $\mathcal{R}$ that is dense everywhere. Moreover, S^* is non-increasing, thus we can extend it to be a function on $\mathcal{R}$ with at most countably many discontinuous points. Denote the set of the latter points by D_d. By taking a further subsequence of $\{n\}_{n\geq 1}$, we can assume, WLOG, that $H_n(y, \mathbf{x})$ converges to $S^*(t)$ at all those $(y, \mathbf{x})$ such that $t = y - \mathbf{b}^{*\prime}\mathbf{x} \in D_d$ by Helly's Lemma. Moreover, for each $(y, \mathbf{x})$ such that $t = y - \mathbf{b}^{*\prime}\mathbf{x} \notin D_d$, t is a continuity point of S^* and the boundedness and monotonicity of H_n (in $y - \hat{\beta}'\mathbf{x}$) ensure that $H_n(y, \mathbf{x})$ also converges to $S^*(t)$. This implies that $\hat{S}_{\hat{\beta}}(y - \hat{\beta}'\mathbf{x})$ converges to $S^*(y - \mathbf{b}^{*\prime}\mathbf{x})$ for each pair $(y, \mathbf{x}) \in \mathcal{R}^{p+1}$. This completes the proof of the second equation of (3.4).

We now prove (3.5). WLOG, we can assume that $\{n_k\} = \{n\}$ for the $\{n_k\}$ given in the above two paragraphs. It is easy to show that $\overline{\lim\limits_{n\to\infty}}\, \eta(\hat{S}_{\hat{\beta}}, y, \mathbf{x}, \hat{\beta})$ is a function of $y - \mathbf{b}^{*\prime}\mathbf{x}$ (see (3.1)) and thus can be denoted by $f^*(y - \mathbf{b}^{*\prime}\mathbf{x})$. We shall now show that

$$(f^*, S^*) \in \mathcal{G}, \textit{ i.e. }, \frac{\overline{\lim\limits_{n\to\infty}}\, \eta(\hat{S}_{\hat{\beta}}, y, \mathbf{x}, \hat{\beta})}{D(S^*)(t)} \leq 1 \text{ for each } t = y - \mathbf{b}^{*\prime}\mathbf{x}.$$

First consider the case that $y - \beta'\mathbf{x} \notin \mathcal{C}$. Define V_n and W_n by $V_n(y, \mathbf{x}) = H_n(y - h_n, \mathbf{x})$ and $W_n(y, \mathbf{x}) = H_n(y + h_n, \mathbf{x})$, respectively. The argument in the above two paragraphs can also be applied to show that for each subsequence of $\{n\}_{n\geq 1}$, there exist a further subsequence $\{n_j\}_{j\geq 1}$ such that H_{n_j}, V_{n_j} and W_{n_j} converge point-wisely for $(y, \mathbf{x}) \in \mathcal{R}^{p+1}$, say, to H, V and W, respectively. Denote $S^-(t) = V(y, \mathbf{x})$ and $S^+(t) = W(y, \mathbf{x})$ whenever $t = y - \mathbf{b}^{*\prime}\mathbf{x}$. Then verify that $S^- = S^* = S^+$ except possibly at the discontinuity points of S^*. Furthermore, since

$$H_n(y - u, \mathbf{x}) \geq H_n(y - h_n, \mathbf{x}) = V_n(y, \mathbf{x}) \geq H_n(y, \mathbf{x}) \geq W_n(y, \mathbf{x}) \geq H_n(y + u, \mathbf{x}) \ \forall\ u > h_n,$$

by taking limits, the above inequalities yield

$$S^*(t-) \geq S^-(t) \geq S^*(t) \geq S^+(t) \geq S^*(t+). \tag{L1.1}$$

Note that for $t = y - \mathbf{b}^{*\prime}\mathbf{x}$ where $y - \beta'\mathbf{x} \notin \mathcal{C}$,

$$\frac{\eta_c(\hat{S}_{\hat{\beta}}, y, \mathbf{x}, \hat{\beta})}{\frac{S^*(t-h_n) - S^*(t+h_n)}{2h_n}} = \frac{\frac{V_n(y,\mathbf{x}) - W_n(y,\mathbf{x})}{2h_n}}{\frac{S^*(t-h_n) - S^*(t+h_n)}{2h_n}} = \frac{V_n(y, \mathbf{x}) - W_n(y, \mathbf{x})}{S^*(t - h_n) - S^*(t + h_n)}. \tag{L1.2}$$

By taking a subsequence, WLOG, we assume that $\frac{S^*(t-h_n) - S^*(t+h_n)}{2h_n}$ and $\frac{V_n(y,\mathbf{x}) - W_n(y,\mathbf{x})}{S^*(t-h_n) - S^*(t+h_n)}$ converge. It follows from (L1.1) and (L1.2) that $\frac{\overline{\lim\limits_{n\to\infty}}\, \eta_c(\hat{S}_{\hat{\beta}}, y, \mathbf{x}, \hat{\beta})}{D(S^*)(t)} \leq 1$.

In the case that $y - \beta'\mathbf{x} \in \mathcal{C}$, it is much easier to show that (3.5) holds. In fact, $\overline{\lim\limits_{n\to\infty}}\, \eta_d(\hat{S}_{\hat{\beta}}, y, \mathbf{x}, \hat{\beta}) = S^-(t) - S^+(t) \leq D(S^*)(t)$. This completes the proof of (3.5). □

Lemma 4.2. *Equations (3.6), (3.7) and (3.8) hold.*

Proof. We now prove (3.6) and (3.7) as follows. If $(g,S) \in \mathcal{G}$ then $g \le f$ $(= D(S))$. Verify that $f_o(t)\ln g(t) \le f_o(t)\ln f(t)$. Then $\nu(S,\mathbf{b},f) \le \nu(S_o,\beta,f_o)$ implies (3.6). Thus WLOG, we can assume that g in (3.6), (3.7) and (3.8) satisfies that $g=f$. By definition,

$$\nu(S,\mathbf{b},f) = \int\int f_o(y-\beta'\mathbf{x})\ln f(y-\mathbf{b}'\mathbf{x})dQ_o(y,\mathbf{x}) + E(S_o(Z-\beta'\mathbf{X})\ln S(Z-\mathbf{b}'\mathbf{X})). \tag{L2.1}$$

Since $|x\ln x| \le 1$ for $x \in [0,1]$,

$$\nu(S_o,\beta,f_o) < |E(\ln f_o(\epsilon))| + 1 < \infty, \tag{L2.2}$$

by (L2.1) and A4. $\nu(S,\mathbf{b},f)$ can also be written as

$$\nu(S,\mathbf{b},f) = \int \psi_o(y,\mathbf{x},z,\beta)\ln\psi(y,\mathbf{x},z,\mathbf{b})d\mu(y,\mathbf{x},z),$$

where $\psi(y,\mathbf{x},z,\beta) = \begin{cases} f(y-\beta'\mathbf{x}) & \text{if } z=1,\\ S(y-\beta'\mathbf{x}) & \text{if } z=0,\end{cases}$ $\psi_o(y,\mathbf{x},z,\beta) = \begin{cases} f_o(y-\beta'\mathbf{x}) & \text{if } z=1,\\ S_o(y-\beta'\mathbf{x}) & \text{if } z=0,\end{cases}$
$d\mu(y,x,z) = \begin{cases} dQ_o(y,\mathbf{x}) & \text{if } z=1,\\ dF_{Z,\mathbf{X}}(y,\mathbf{x}) & \text{if } z=0.\end{cases}$ Then it follows from the Shannon-Kolmogorov inequality and (L2.2) that, under the restriction $(f,S) \in \mathcal{G}$,

$$\nu(S,\mathbf{b},f) \text{ is maximized iff } P(S(M-\mathbf{b}'\mathbf{X}) = S_o(M-\beta'\mathbf{X})) = 1 \text{ and } f = D(S). \tag{L2.3}$$

It follows from Lemma 4.4, A2 and (L2.3) that $\mathbf{b}=\beta$. Thus (3.6) and (3.7) hold.

We now prove (3.8). Since $\hat\beta$ is an MSMLE, $\mathcal{N}(\hat S_{\hat\beta},\hat\beta) \ge \mathcal{N}(S_o,\beta)$. Then by (3.3),

$$\underline{\lim}_{n\to\infty} \mathcal{N}(\hat S_{\hat\beta},\hat\beta) \ge \lim_{n\to\infty} \mathcal{N}(S_o,\beta) = \nu(S_o,\beta,f_o) \text{ with probability one.}$$

Thus in order to prove (3.8), it suffices to show that

$$\nu(S^*,\mathbf{b}^*,f^*) \ge \underline{\lim}_{n\to\infty} \mathcal{N}(\hat S_{\hat\beta},\hat\beta).$$

The latter inequality follows from Lemma 4.3 below and equation (3.2), *i.e.*,

$$\nu(S^*,\mathbf{b}^*,f^*) = \int\int \ln f^*(y-\mathbf{b}^{*\prime}\mathbf{x})dQ(y,\mathbf{x}) + \int\int \ln S^*(y-\mathbf{b}^{*\prime}\mathbf{x})dG(y,\mathbf{x}). \quad \square$$

Lemma 4.3. *With probability one,*

(1) $$\underline{\lim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^n (1-\delta_i)\ln S_{\hat\beta}(M_i-\hat\beta'\mathbf{x}_i) \le \int\int \ln S^*(y-\mathbf{b}^{*\prime}\mathbf{x})dG(y,\mathbf{x}),$$

(2) $$\underline{\lim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \delta_i\ln\eta(\hat S_{\hat\beta},M_i,\mathbf{x}_i,\hat\beta) \le \int\int \ln f^*(y-\mathbf{b}^{*\prime}\mathbf{x})dQ(y,\mathbf{x}).$$

Proof. Denote $d\hat G$ and $d\hat Q$ the empirical versions of dG and dQ, respectively. Statement (1) is the same as

$$\underline{\lim}_{n\to\infty} \int \ln\hat S_{\hat\beta}(y-\hat\beta'\mathbf{x})d\hat G \le \int \ln S^*(y-\mathbf{b}^{*\prime}\mathbf{x})dG.$$

Let $\mu_s = (\mu_Z + \mu_{X_1} + \mu_{X_2} + \cdots + \mu_{X_p} + \mu_\epsilon)$, where μ_{X_i} is the measure induced by the cdf of X_i, etc. and $\mathbf{X}' = (X_1, ..., X_p)$. Then μ_s is a finite measure (bounded by 0 and $p+2$). For each positive integer v, there are finitely many real numbers (quantiles of μ_s) $q_{1,v} < \cdots < q_{u,v}$ such that $E(\mathbf{1}_{(\epsilon \in (q_{i,v}, q_{i+1,v}))}) < 2^{-v}$, $E(\mathbf{1}_{(Z \in (q_{j,v}, q_{j+1,v}))}) < 2^{-v}$ and $E(\mathbf{1}_{(X_k \in (q_{h,v}, q_{h+1,v}))}) < 2^{-v}$, $k = 1, ..., p$, for all possible $i, j, h = 0, ..., u$, where $q_v = -\infty$ and $q_{u+1,v} = \infty$.

Fixed a positive integer v and form the sets $U_0, ..., U_{2u+2}$ by setting $U_{2i-1} = (q_{i-1,v}, q_{i,v})$ for $i = 1, ..., u+1$, and $U_{2i} = [q_{i,v}, q_{i,v}]$ for $i = 0, ..., u+1$. Let $\mathcal{U}_v$ denote the collection of all nonempty set of the form $U_{j_1,...,j_{p+1}} = U_{j_1} \times U_{j_2} \times \cdots \times U_{j_{p+1}}$. Let $\mathcal{U} = \cup_v \mathcal{U}_v$. Let μ_G be the measure on $\mathcal{R}^{p+1}$ induced by G and denote $\mu_{\hat{G}}$ in an obvious way. Let Ω_0 be the subspace of Ω such that $\mu_{\hat{G}}(U, \omega) \to \mu_G(U)$ for each $U \in \mathcal{U}$. Since $\mathcal{U}$ is countable, $\mathrm{P}(\Omega_0) = 1$. WLOG, we can assume that $\Omega = \Omega_0$. Hereafter, we fixed $\omega \in \Omega_0$.

Now fix a positive integer v and a negative number ψ. Denote $\hat{S} = \hat{S}_{\hat{\beta}}$. Then

$$\int \ln\hat{S}(y - \hat{\beta}'\mathbf{x})d\hat{G} \le \int \psi \vee \ln\hat{S}(y - \hat{\beta}'\mathbf{x})d\hat{G} \le \sum_{U \in \mathcal{U}_v} \mathcal{M}(U, \hat{S})\mu_{\hat{G}}(U), \qquad (L3.1)$$

where $\mathcal{M}(U, \hat{S}) = \sup\{\psi \vee \ln\hat{S}(y - \hat{\beta}'\mathbf{x}) : (y, \mathbf{x}) \in \overline{U}\}$ and $\overline{U}$ is the closure of the set U. Since $\mathcal{M}$ is bounded, $\overline{U}$ is a closed set and $\mathcal{M}(U, \hat{S}) \to \mathcal{M}(U, S^*)$,

$$\sum_{U \in \mathcal{U}_v} \mathcal{M}(U, \hat{S})\mu_{\hat{G}}(U) \to \sum_{U \in \mathcal{U}_v} \mathcal{M}(U, S^*)\mu_G(U). \qquad (L3.2)$$

Since $\psi \vee \ln S^*(y - \mathbf{b}^{*\prime}\mathbf{x})$ is integrable by (3.6), and dG is absolutely continuous w.r.t. μ_s^{p+1},

$$\sum_{U \in \mathcal{U}_v} \mathcal{M}(U, S^*)\mu_G(U) \to \int \psi \vee \ln S^*(y - \mathbf{b}^{*\prime}\mathbf{x})dG \text{ as } v \to \infty, \qquad (L3.3)$$

by the definition of integral. As a consequence of (L3.1), (L3.2) and (L3.3),

$$\begin{aligned} \varlimsup_{n\to\infty} \int \ln\hat{S}_{\hat{\beta}}(y - \hat{\beta}'\mathbf{x})d\hat{G} \le & \varlimsup_{n\to\infty} \int \psi \vee \ln\hat{S}(y - \hat{\beta}'\mathbf{x})d\hat{G} \\ & \to \int \ln S^*(y - \mathbf{b}^{*\prime}\mathbf{x})dG \text{ as } \psi \downarrow -\infty, \end{aligned}$$

by the monotone convergence theorem and Lemma 4.1. Thus statement (1) holds.

Statement (2) is equivalent to $\varlimsup_{n\to\infty} \int \ln\eta(\hat{S}_{\hat{\beta}}, y, \mathbf{x}, \hat{\beta})d\hat{Q} \le \int \ln f^*(y - \mathbf{b}^{*\prime}\mathbf{x})dQ$. It suffices to show that

$$\begin{aligned} & \varlimsup_{n\to\infty} \int \ln\eta_c(\hat{S}_{\hat{\beta}}, y, \mathbf{x}, \hat{\beta})d\hat{Q} \le \int \ln f_c^*(y - \mathbf{b}^{*\prime}\mathbf{x})dQ, && (L3.4) \\ & \text{where } f_c^*(y - \mathbf{b}^{*\prime}\mathbf{x}) = \varlimsup_{n\to\infty} \eta_c(\hat{S}_{\hat{\beta}}, y, \mathbf{x}, \hat{\beta})\mathbf{1}_{(y-\beta'\mathbf{X}=t\notin\mathcal{C})}, \\ & \varlimsup_{n\to\infty} \int \ln\eta_d(\hat{S}_{\hat{\beta}}, y, \mathbf{x}, \hat{\beta})d\hat{Q} \le \int \ln f_d^*(y - \mathbf{b}^{*\prime}\mathbf{x})dQ, && (L3.5) \\ & \text{where } f_d^*(y - \mathbf{b}^{*\prime}\mathbf{x}) = \varlimsup_{n\to\infty} \eta_d(\hat{S}_{\hat{\beta}}, y, \mathbf{x}, \hat{\beta})\mathbf{1}_{(y-\beta'\mathbf{X}=t\in\mathcal{C})}. \end{aligned}$$

Recall that

$$\begin{aligned} \eta_c(\hat{S}_{\hat{\beta}}, y, \mathbf{x}, \hat{\beta}) = & \frac{h_n}{2}[\hat{S}_{\hat{\beta}}(y - \hat{\beta}'\mathbf{x} - h_n) - \hat{S}_{\hat{\beta}}(y - \hat{\beta}'\mathbf{x} + h_n)] \\ = & \frac{h_n}{2}[V_n(y, \mathbf{x}) - W_n(y, \mathbf{x})] \text{ (see (L1.2))}, \end{aligned}$$

where $h_n = n^{1/5}$. Verify that replacing η_c (or η_d) and G for $\hat{S}_{\hat{\beta}}$ and Q, in the proof of statement (1), respectively, the arguments are still valid.

In particular, fix a positive integer v and a negative number ψ. Let $\hat f(y,\mathbf{x}) = \eta_c(\hat S_{\hat\beta}, y, \mathbf{x}, \hat\beta)$. Then

$$\int \ln\hat f(y,\mathbf{x})d\hat Q \le \int \psi \vee \ln\hat f(y,\mathbf{x})d\hat Q \le \sum_{U\in\mathcal{U}_v} \mathcal{M}(U,\hat f)\mu_{\hat Q}(U), \tag{L3.6}$$

where $\mathcal{M}(U,\hat f) = \sup\{\psi\vee\ln\hat f(y,\mathbf{x}) : (y,\mathbf{x}) \in \overline{U}\}$ and $\overline{U}$ is the closure of the set U. Verify that $\overline{\lim}_{n\to\infty} \hat f(y,\mathbf{x})$ is a function of $y-\mathbf{b}^{*\prime}\mathbf{x}$, denoted by $f_c^*(y-\mathbf{b}^{*\prime}\mathbf{x})$. Since $\mathcal{M}$ is bounded, $\overline{U}$ is a closed set and $\mathcal{M}(U,\hat f) \to \mathcal{M}(U,f_c^*)$,

$$\sum_{U\in\mathcal{U}_v} \mathcal{M}(U,\hat f)\mu_{\hat Q}(U) \to \sum_{U\in\mathcal{U}_v} \mathcal{M}(U,f_c^*)\mu_{Q}(U). \tag{L3.7}$$

Since $\psi\vee\ln f_c^*(y-\mathbf{b}^{*\prime}\mathbf{x})$ is integrable by (3.6), and dQ is absolutely continuous w.r.t μ_s^{p+1},

$$\sum_{U\in\mathcal{U}_v} \mathcal{M}(U,f_c^*)\mu_{Q}(U) \to \int \psi\vee\ln f_c^*(y-\mathbf{b}^{*\prime}\mathbf{x})dQ \text{ as } v\to\infty, \tag{L3.8}$$

by the definition of integral. As a consequence of (L3.6), (L3.7) and (L3.8),

$$\begin{aligned} \underline{\lim}_{n\to\infty} \int \ln\hat f(y,\mathbf{x})d\hat Q &\le \underline{\lim}_{n\to\infty} \int \psi\vee\ln\hat f(y,\mathbf{x})d\hat Q \\ &\to \int \ln f_c^*(y-\mathbf{b}^{*\prime}\mathbf{x})dQ \text{ as } \psi\downarrow -\infty, \end{aligned}$$

by the monotone convergence theorem and Lemma 4.1. This completes the proof of (L3.4). (L3.5) can be proved in a similar way by replacing η_c by η_d. Thus statement (2) holds. This completes the proof of the lemma. □

Lemma 4.4. *$S\in\mathcal{F}$ and $P(S(M-\mathbf{b}'\mathbf{X}) = S_o(M-\beta'\mathbf{X})) = 1$ imply that $\mathbf{b}=\beta$.*

Proof. Suppose that $\mathbf{b}\ne\beta$, $S\in\mathcal{F}$ and $\mathrm{P}(S(M-\mathbf{b}'\mathbf{X}) = S_o(M-\beta'\mathbf{X})) = 1$. We shall show that they lead to a contradiction.

It is obvious that S is right continuous, as S_o is right continuous. Define $S^-(t) = S(t-)$ and $S_o^-(t) = S_o(t-)$. If $\delta = 1$, then $M-\mathbf{b}'\mathbf{X} = \epsilon + (\beta-\mathbf{b})'\mathbf{X}$. Thus,

$$P\{S(\epsilon-\mathbf{b}'\mathbf{X}) = S_o(\epsilon-\beta'\mathbf{X}),\ S^-(\epsilon-\mathbf{b}'\mathbf{X}) = S_o^-(\epsilon-\beta'\mathbf{X})|\delta=1\} = 1 \tag{L4.1}$$

by A1, A2 and the assumption at the beginning of the proof. By A2, there are two points $\mathbf{x}_1$ and $\mathbf{x}_2$ in the range of $\mathbf{X}$ such that

$$(\beta-\mathbf{b})'\mathbf{x}_1 < (\beta-\mathbf{b})'\mathbf{x}_2. \tag{L4.2}$$

By A1 and A2, there exists a $t_0\in\mathcal{R}_{T(\beta)}$ such that one of the following cases is true:

(1) t_0 is a discontinuity point of S_o;

(2) t_0 is a continuity of point of S_o, $t_o<\tau$, and $S_o(t_0+c) < S_o(t_0)$ for all $c>0$.

Suppose Case (1) is true. Then (L4.1) implies that

$$\begin{aligned} S(t_0+(\beta-\mathbf{b})'\mathbf{x}_1) = S(t_0+(\beta-\mathbf{b})'\mathbf{x}_2) = S_o(t_0) \\ <S_o^-(t_0) = S^-(t_0+(\beta-\mathbf{b})'\mathbf{x}_1) = S^-(t_0+(\beta-\mathbf{b})'\mathbf{x}_2) \end{aligned} \tag{L4.3}$$

as $S_o^-(t_0) = S_o(t_0-) > S_o(t_0)$. Thus

$$S(t_0+(\beta-\mathbf{b})'\mathbf{x}_1) < S^-(t_0+(\beta-\mathbf{b})'\mathbf{x}_2). \tag{L4.4}$$

However, by (L4.2), $S(t_0+(\beta-\mathbf{b})'\mathbf{x}_1) \ge S^-(t_0+(\beta-\mathbf{b})'\mathbf{x}_2)$, as $S\in\mathcal{F}$, which contradicts (L4.4). The contradiction implies that $\mathbf{b}=\beta$.

Suppose Case (2) is true. Then there is a point t_1 such that

$$\begin{aligned} 0 < t_1 - t_0 < (\beta-\mathbf{b})'\mathbf{x}_2 - (\beta-\mathbf{b})'\mathbf{x}_1 \text{ and} \\ S(t_1+(\beta-\mathbf{b})'\mathbf{x}_1) = S(t_1+(\beta-\mathbf{b})'\mathbf{x}_2) = S_o(t_1). \end{aligned} \tag{L4.5}$$

Moreover, (L4.3) holds. It follows that

$$S(t_1+(\beta-\mathbf{b})'\mathbf{x}_1) = S_o(t_1) < S_o(t_0) = S(t_0+(\beta-\mathbf{b})'\mathbf{x}_2). \tag{L4.6}$$

However, since $t_1+(\beta-\mathbf{b})'\mathbf{x}_1 < t_0+(\beta-\mathbf{b})'\mathbf{x}_2$ by (L4.5), and $S\in\mathcal{F}$,

$$S(t_1+(\beta-\mathbf{b})'\mathbf{x}_1) \ge S(t_0+(\beta-\mathbf{b})'\mathbf{x}_2),$$

contradicting (L4.6). The contradiction implies that $\mathbf{b}=\beta$. Thus (C.1) holds. □

References.

[1] Theil, H. A rank-invariant method of linear and polynomial regression analysis, I. Proc. Kon. Ned. Akad. v. Wetensch. A, 1950, 53: 386-392.

[2] Sen, P.K. Estimates of the regression coefficient based on Kendall' tau. *JASA*, 1968, 63: 1379-1389.

[3] Bickel, P.J. On adaptive estimation. *Ann. Statist.*, 1982, 10: 647-671.

[4] Huber, P.J. Robust estimation of a location parameter. *Ann. Math. Statist.*, 1964, 35: 73-101.

[5] Ritov, Y. Estimation in a linear regression model with censored data. *Ann. Statist.*, 1990, 18: 303-328.

[6] Montegomery, D.C. and Peck, E. A. Introduction to linear regression analysis. *Wiley*, N.Y., 1992.

[7] Draper, N.R. and Smith, H. Applied Regression analysis. *Wiley* NY., 1981.

[8] Zhang, C.H. and Li, X. Linear regression with doubly censored data. *Ann. Statist.*, 1996, 24: 2720-2743.

[9] Miller, R.G. Least squares regression with censored data. *Biometrika*, 1976, 63: 449-464.

[10] Buckley, J. and James, I. Linear regression with censored data. *Biometrika*, 1979, 66: 429-436.

[11] Chatterjee, S. and Mcleish, D. L. Fitting linear regression models to censored data by least squares and maximum likelihood methods. *Comm. Stat.*, A, 1986, 15: 3227-3243.

[12] Leurgans, S. Linear models, random censoring and synthetic data. *Biometrika*, 1987, 74: 301-309.

[13] Hillis, S.L. Extending M-estimation to include censored data via James's method. *Comm. Stat.* B, 1991, 20: 121-128.

[14] Tsiatis, A.A Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.*, 1990, 18: 354-372.

[15] Wei, L.J., Ying, Z.L. and Lin, D.Y. Linear regression analysis of censored survival data based on rank tests. *Biometrika*, 1990, 77: 845-851.

[16] Akritas, M.G., Murphy S.A. and Lavalley M.P. The Theil-Sen estimator with doubly censored -data and applications to astronomy. *JASA*, 1995, 90: 170-177.

[17] Moon, C.G. A Monte Carlo comparison of semiparametric Tobit estimators. *Journal Of Applied Econometics*, 1989, 4: 361-382.

[18] Yu, Q.Q. and Wong, G.Y.C. The semi-parametric MLE in linear regression analysis with right-censored data. (Submitted for publication).

[19] Yu, Q. Q., Schick, A., Li, L. X. and Wong, G.Y.C. Asymptotic properties of the GMLE of a survival function with case 2 interval-censored data. *Statist. & Prob. Let.*, 1998, 37: 223-228.

[20] Groeneboom, P. and Wellner, J.A. Information bounds and nonparametric maximum likelihood estimation. *Birkhäuser Verlag, Basel*, 1992

[21] Yu, Q.Q. and Wong, G.Y.C. A modified semi-parametric MLE in linear regression analysis with complete data and right-censored data. *Technometrics* (under revision).