Journal of Econometrics and Statistics
Vol. 2, Issue 2, 2022, pp. 137-147
https://doi.org/10.47509/JES.2022.v02i02.01

Consistency of the Semi-parametric MLE of the Lehmann Family with Right-Censored Data

Qiqing Yu*

Article History
Received : 25 May 2022; Revised : 19 June 2022; Accepted : 29 June 2022; Published : 15 December 2022

Abstract

We study the consistency of the semi-parametric maximum likelihood estimator (SMLE) of the Lehmann family (Lehmann (1959)) with right-censored data. The Lehmann family is a class of survival functions of the form $(S_o(t))^{\exp(\mathbf{x}'\beta)}$, where $S_o(\cdot)$ is a survival function and $\mathbf{x}$ is a p-dimensional vector. It is the same as Cox's regression model iff S_o is absolutely continuous. The Lehmann family and Cox's model are both popular semi-parametric regression models in survival analysis. Consistency proofs of an estimator under various semi-parametric set-ups are often based on additional regularity conditions. We establish the consistency of the SMLE of the Lehmann family without any additional assumption.

Keywords: Lehmann family, Cox's model, Semi-parametric Maximum likelihood estimator, consistency, Kullback-Leibler Inequality.

JEL Codes: N01 and G08

1 Introduction

We study the consistency of the semi-parametric maximum likelihood estimator (SMLE) of the Lehmann family (Lehmann (1959) or Cox and Oakes (1984,p.24)) with right censored (RC) data.

Let Y be a random variable, $\mathbf{X}$ a p-dimensional random vector and $g(\mathbf{x}) \geq 0$. Let F_T $(= 1 - S_T)$ be the cumulative distribution function (cdf) of T $(T = \mathbf{X}$

*Professor, Department of Mathematics and Statistics, Binghamton University (SUNY), NY 10902, USA. Email id: qyu@math.binghamton.edu, Tel: 1-607-777-4634, ORCID: 0000-0002-2376-4789.

To cite this paper
Qiqing Yu (2022). Consistency of the Semi-parametric MLE of the Lehmann Family with Right-Censored Data. *Journal of Econometrics and Statistics*. 2(2), 137-147.

or Y, *etc.*), f_T the density function (df) of T, $S_o(\cdot)$ a survival function and h_o a hazard function. Let $S(\cdot|\cdot)$ ($h(\cdot|\cdot)$, $F(\cdot|\cdot)$ and $f(\cdot|\cdot)$) be the conditional survival function (hazard function, cdf and df) of Y given $\mathbf{X}=\mathbf{x}$. The Lehmann family is the collection of all distributions of the form $S(t|\mathbf{x}) = (S_o(t))^{g(\mathbf{X})}$. We set $g(\mathbf{x}) = e^{\mathbf{X}'\beta}$ as usual. It is a regression model commonly used in the survival analysis and is also called the proportional integrated hazards model, as $\ln S(y|\mathbf{x}) = e^{\mathbf{X}'\beta}\ln S_o(y)$. Notice that the proportional hazards model or Cox's model (Cox (1972)) is defined by $h(y|\mathbf{x}) = e^{\beta'\mathbf{X}}h_o(y)\ \forall\ y < \tau_Y$, where $\tau_Y = \sup\{t:\ S_Y(t) > 0\}$.

The Lehmann family is often mistaken for Cox's model (see *e.g.*, Sun (2006) p.18), but they are the same iff S_o is absolutely continuous (see *e.g.*, Yu *et al.* (2008)). Even though Cox's model was proposed in 1972, later than the Lehmann family (in 1959), it becomes more popular due to the maximum partial likelihood estimator (mple) approach. The mple is much simpler and faster to compute than the SMLE. Finkelstein (1986) points out that the partial likelihood approach does not work if the data are interval-censored and she proposes to use the SMLE of Cox's model. When people specify Cox's model by $S(t|\mathbf{x}) = (S_o(t))^{e^{\mathbf{X}'\beta}}$ without assuming that S_o' exists (see *e.g.* Sun (2006,p.18) or Finkelstein (1986)), they are actually make use of the Lehmann family rather than Cox's model.

We make the following assumption in this paper.

(A1) Let $(M_i,\delta_i,\mathbf{X}_i)$, $i = 1, ..., n$, be i.i.d. observations from $(M,\delta,\mathbf{X})$, where $M = Y \wedge C$, C is a random censoring variable, $\delta = \mathbf{1}(Y \le C)$, $\mathbf{1}(A)$ is the indicator function of the event A, and $(Y,\mathbf{X})$ and C are independent. $S(t|\mathbf{x}) = (S_o(t))^{\exp(\beta'\mathbf{X})}$. $S_{\mathbf{X}}$ and S_C do not depend on the unknown (β, S_o).

Under assumption (A1), the density function $f_{M,\delta,\mathbf{X}}$ may not exist. If it does, then

$$f_{M,\delta,\mathbf{X}}(m,\delta,\mathbf{x}) = (S(m|\mathbf{x}))^{1-\delta}(f(m|\mathbf{x}))^{\delta}(f_C(m))^{1-\delta}(S_C(m))^{\delta}f_{\mathbf{X}}(\mathbf{x}), \text{ where} \quad (1.1)$$

$$m \in \mathcal{D}, \text{ and } \mathcal{D} = \begin{cases} (-\infty,\tau_M] & \text{if } P(Y=\tau_M|\mathbf{X}=0)=0 \text{ or } P(C \ge \tau_M) > 0 \\ (-\infty,\tau_M) & otherwise. \end{cases} \quad (1.2)$$

Definition. Under assumption (A1), $S(t|\mathbf{x})$ or $(S_o(t),\beta)$ is said to be identifiable if $(S_o(t))^{\exp(\mathbf{X}'\beta)} = (S_*(t))^{\exp(\mathbf{X}'\beta_*)}\ \forall\ (t,\mathbf{x}) \in \mathcal{D}_{M,\mathbf{X}} => (\beta, S_o(t)) = (\beta_*, S_*(t))\ \forall\ t \in \mathcal{D}$, where

$$\mathcal{D}_T = \{t:\ P(||T-t|| < \varepsilon) > 0\ \forall\ \varepsilon > 0\}, \text{ and } T = (M,\mathbf{X}), \text{ or } \mathbf{X}, \text{ or } Y \quad (1.3)$$

etc. and $||\cdot||$ is a norm.

The generalized likelihood function defined by Kiefer and Wolfowitz (1956) is $\mathcal{L}_0(S_o,\beta) = \prod_{i=1}^n[(S(M_i|\mathbf{X}_i))^{1-\delta_i}(S(M_i-|\mathbf{X}_i) - S(M_i|\mathbf{X}_i))^{\delta_i}]$. Since $S(t|\mathbf{x}) =$

$(S_o(t))^{e^{\mathbf{X}'\beta}}$ is the true distribution of $Y|(\mathbf{X}=\mathbf{x})$, $S(t-|\mathbf{x})-S(t|\mathbf{x})=0=\mathcal{L}_0(S_o,\beta)$ if S_o is continuous. The SMLE of (S_o,β), denoted by $(\hat{S}_o,\hat{\beta})$, maximizes $\mathcal{R}_0(S_*,\beta_*)$ over all possible discrete survival function S_* and over all $\beta_* \in \mathcal{R}^p$ ($\mathcal{R}=(-\infty,\infty)$). Denote $\hat{S}(t|\mathbf{x}) = (\hat{S}_o(t))^{\exp(\hat{\beta}'\mathbf{x})}$. Wong and Yu (2012) show that the Newton-Raphson method does not always work for the SMLE with censored data, and propose a feasible algorithm. For given RC data, the algorithm yields an estimate which is the SMLE of (β,S_o) under Cox's model assuming S_o is absolutely continuous, and is also the SMLE of (β,S_o) of the Lehmann family without any restriction on S_o, in particular, if there exist ties in the data. For technical reason in the proof, we modify $\mathcal{L}_0(S_o,\beta)$ as follows.

$$\mathcal{L}(S_o,\beta)=\prod_{i=1}^{n}[(S(M_i|\mathbf{X}_i))^{1-\delta_i}(S(M_i-\eta_n|\mathbf{X}_i)-S(M_i|\mathbf{X}_i))^{\delta_i}], \text{ where} \tag{1.4}$$

$\eta_n = \frac{1}{n} \wedge \min\{|M_i-M_j| : \ M_i \neq M_j, \ i,j \in \{1,2,...,n\}\}$. If S_o is discrete then $\mathcal{R}(S_o,\beta)=\mathcal{R}_0(S_o,\beta)$.

Several classical textbooks (see *e.g.*, Ferguson (1996) and Casella and Berger (2001)) provide typical sufficient conditions for the consistency of the MLE, such as

(C1) X_1, ..., X_n are i.i.d. observations from $f(\cdot;\theta)$, $\theta \in \Theta$ and θ_o is the true value of θ;

(C2) $\int |f(x;\theta)-f(x;\theta_o)|d\mu(x)=0$ implies that $\theta=\theta_o$ (identifiability);

(C3) The densities $f(x;\theta)$ have common support, and $f(x;\theta)$ is differentiable in θ;

C4) $\ln f(t)/f_o(t) \leq K(x)$, where $\int K(t)f_o(t)d\mu(t) < \infty$.

We succeed in establishing the consistency of the SMLE of the Lehmann family (specified in (A1)) without imposing any additional regularity conditions, provided that the parameter is identifiable (see Theorem 2). In particular, we allow the baseline survival function S_o can be any arbitrary one.

2 The Main Results

We shall first introduce some preliminary results. It is clear that an estimator of (S_o,β) is consistency only if (S_o,β) is identifiable.

Theorem 1. *The necessary and sufficient identifiable condition for* (S_o,β) *under (A1) is*

(A2) (A2.1) $S_o(\tau_M) < 1$*, where* $\tau_M = \sup\{t:\ S_M(t) > 0\}$*;*

(A2.2) $\exists$ $\mathbf{0}$*,* $\mathbf{x}_1$*, ...,* $\mathbf{x}_p \in \mathcal{D}_{\mathbf{X}}$ *(see (1.3)) such that* $\mathbf{x}_1$*, ...,* $\mathbf{x}_p$ *are linearly independent.*

Proof. (=>). Suppose (A2) holds. If

$$(S_o(t))^{\exp(\mathbf{X}'\beta)} = (S_*(t))^{\exp(\mathbf{X}'\beta_*)}\ \forall\ (t,\mathbf{x}) \in \mathcal{D}_{M,\mathbf{X}} \text{ (see (1.3))}, \tag{2.1}$$

then $S_o(t) = (S_o(t))^{\exp(\mathbf{0}'\beta)} = (S_*(t))^{\exp(\mathbf{0}'\beta_*)} = S_*(t)\ \forall\ t \in \mathcal{D} = \mathcal{D}_{M,\mathbf{0}}$, as $\mathbf{0} \in \mathcal{D}_{\mathbf{X}}$.
By (2.1), $(S_o(\tau_M))^{\exp(\mathbf{X}_i'\beta)} = (S_o(\tau_M))^{\exp(\mathbf{X}_i'\beta_*)}$, where $\mathbf{x}_i$ is as in (A2.2).
Thus $\mathbf{x}_i'\beta = \mathbf{x}_i'\beta_*$ for $i = 1, ..., p$, as $S_o(\tau_M) < 1$ by (A2.1).
=> $\beta = \beta_*$ as $\beta \in \mathcal{R}^p$ and $\mathbf{x}_1, ..., \mathbf{x}_p$ are linearly independent by (A2.2).
Thus (2.1) yields $(\beta_*, S_*(t)) = (\beta, S_o(t))\ \forall\ t \in \mathcal{D}$. That is, (A2) is sufficient.
(<=). If (A2) fails, then either (a) $S_o(\tau_M) = 1$, or (b) $\mathbf{0} \notin \mathcal{D}_{\mathbf{X}}$, or (c) $\mathbf{0} \in \mathcal{D}_{\mathbf{X}}$ but (A2.2) fails. Thus it suffices to show that $(\beta, S_o(t))$ is not identifiable in these 3 cases.
Case (a). Let $S_*(t) = S_o(\tau_M) = 1\ \forall\, t \le \tau_M$. Then $1 = (S_o(t))^{\exp(\mathbf{X}'\beta)} = (S_*(t))^{\exp(\mathbf{X}'\beta_*)}$ $\forall\ t \le \tau_M$ where $\beta_* = 2\beta$. Hence equation (2.1) holds with $\beta_* = 2\beta$. Thus β is not identifiable.
Case (b). Let $S_o(\tau_M) < 1$ and $S_*(t) = (S_o(t))^2$, then $\ln(-\ln(S_o(t))^2) = \ln 2 + \ln(-\ln S_o(t))$. Taking ln(-ln) on both sides of (2.1) yields $\mathbf{x}'(\beta - \beta_*) = \ln 2$, which yields β_* ($\neq \beta$). That is, $\exists\ \beta^* \neq \beta$ such that (2.1) holds if $\mathbf{0} \notin \mathcal{D}_{\mathbf{X}}$.
Case (c). Let $S_* = S_o$, then taking ln(-ln) on both sides of (2.1) yields $\mathbf{x}_i'\beta = \mathbf{x}_i'\beta_*$ for $i = 1, ..., p$. => $\beta = \beta_*$ iff $\mathbf{x}_1, ..., \mathbf{x}_p$ are linearly independent. But the latter condition is violated in Case (c). Thus $\exists\ \beta^* \neq \beta$ such that (2.1) holds in Case (c).
□

We shall make use of the Kullback-Leibler (KL) information inequality.
The KL inequality. $\int f_o(t)\ln(f_o/f)(t)d\mu(t)$ *always exists (though may be* ∞*), where* f_o *and* f *are two densities w.r.t. a measure* μ*. Moreover,*
$\int f_o(t)\ln(f_o/f)(t)d\mu(t) \ge 0$*; with equality iff* $f = f_o$ *a.e. w.r.t.* μ*.*

In the KL inequality, f and f_o are the densities of the same type of distributions, *e.g.*, (1) absolutely continuous ones or (2) discrete ones. Then μ is the Lebesgue measure in Case (1) and the counting measure in Case (2). Under our semi-parametric set-up, S_o and S may belong to different types of distributions, or even the Cantor distribution. Then the densities f_o and f, as well as the measure μ need to be properly defined.
Example 1. Let $\mu(t)$ be the cdf of the Cantor distribution on $[0,1]$, $f_o(t)$ be its df ($= \frac{d\mu(t)}{d\mu(t)} = \mathbf{1}(t \in \mathcal{A} \cap [0,1])$), where $\mathcal{A}$ is the Cantor ternary set, and let f be the df of the uniform distribution on the interval $[0,1]$ with the cdf $F(t)$, then $f(t) = \frac{dF(t)}{d\mu(t)} = +\infty\mathbf{1}(t \in [0,1] \setminus \mathcal{A})$. $\int f(t)d\mu(t) = 0 \neq 1 = \int f_o(t)d\mu(t)$.

Example 2. Let $\mu(t)$ be the cdf of a binomial distribution and f_o its df, f as in Example 1, then $f(t) = \frac{dF(t)}{d\mu(t)} = \infty$ if $t \in (0,1)$. Again,

$\int f(t)d\mu(t) = 0 \neq 1 = \int f_o(t)d\mu(t)$.

In both Examples 1 and 2, we expect $\int f_o(t)\ln\frac{f_o(t)}{f(t)}d\mu(t) \geq 0$, thus the KL inequality is not directly applicable, as $\int f(t)d\mu(t) \neq 1$. We shall make use of a modified KL inequality, which allows $\int f(t)d\mu(t) \in [0,1]$.

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

Proof. Without loss of generality (WLOG), we can assume that $\int f_2(t)d\mu_1(t) = c \in [0,1)$ and $\int_{\{t_o\}} f_1(t)d\mu_1(t) = 0$. Define $f_3(t) = \begin{cases} f_2(t) & \text{if } t \neq t_o \\ 1-c & \text{if } t = t_o \end{cases}$, $f_4(t) = f_1(t)\mathbf{1}(t \neq t_o)$, $\mu_2(\{t\}) = \mathbf{1}(t = t_o)$ and $\mu = \mu_1 + \mu_2$. Then $\int f_4(t)d\mu(t) = \int f_3(t)d\mu(t) = 1$. Moreover,

$$
\begin{aligned}
0 &\leq \int f_4(t)\ln\frac{f_4(t)}{f_3(t)}d\mu(t) \qquad\qquad \text{(by the KL inequality)}\\
&= \int_{t\neq t_o} f_4(t)\ln\frac{f_4(t)}{f_3(t)}d\mu_1(t) + \int_{t=t_o} f_4(t)\ln\frac{f_4(t)}{f_3(t)}d\mu_1(t) + \int f_4(t)\ln\frac{f_4(t)}{f_3(t)}d\mu_2(t)\\
&= \int_{t\neq t_o} f_4(t)\ln\frac{f_4(t)}{f_3(t)}d\mu_1(t) \qquad (as\ f_4(t_o)\ln\frac{f_4(t_o)}{f_3(t_o)}(\mu_1(\{t_o\})+1) = 0\ln 0 \stackrel{def}{=} 0)\\
&= \int_{t\neq t_o} f_1(t)\ln\frac{f_1(t)}{f_2(t)}d\mu_1(t) = \int f_1(t)\ln\frac{f_1(t)}{f_2(t)}d\mu_1(t)\ (as \int_{t=t_o} f_1(t)d\mu_1(t) = 0). \qquad \square
\end{aligned}
$$

In view of Eq. (1.1), one may write the measure w.r.t. $F_{M,\delta,\mathbf{X}}$ as

$$
\begin{aligned}
dF(m,s,\mathbf{x}) =& \mathbf{1}(s=0)dF(m,0,\mathbf{x}) + \mathbf{1}(s=1)dF(m,1,\mathbf{x}),\\
dF(m,0,\mathbf{x}) =& S(m|\mathbf{x})dF_C(m)dF_{\mathbf{X}}(\mathbf{x}), \quad dF(m,1,\mathbf{x}) = S_C(m)dF(m|\mathbf{x})dF_{\mathbf{X}}(\mathbf{x}), \quad (2.2)\\
dF_*(m,s,\mathbf{x}) =& \mathbf{1}(s=0)S_*(m|\mathbf{x})dF_C(m)dF_{\mathbf{X}}(\mathbf{x}) + \mathbf{1}(s=1)S_C(m)dF_*(m|\mathbf{x})dF_{\mathbf{X}}(\mathbf{x}).
\end{aligned}
$$

In view of Eq. (2.2), the KL inequality under the Lehmann family is modified as follows.

Proposition 2. *Let $S_*(\cdot|\cdot)$ and $S(\cdot|\cdot)$ be two conditional survival functions. Let $g(t|\mathbf{x}) = 1$ and*

$$
g_*(t|\mathbf{x}) = \begin{cases} \frac{S_*(t-|\mathbf{x})-S_*(t|\mathbf{x})}{S(t-|\mathbf{x})-S(t|\mathbf{x})} & \textit{if } S(t-|\mathbf{x})-S(t|\mathbf{x}) > 0,\\ \frac{S'_*(t|\mathbf{x})}{S'(t|\mathbf{x})} & \textit{if } S'(t|\mathbf{x}) < 0 \textit{ and } S'_*(t|\mathbf{x}) \textit{ exist,} \\ \limsup_{s\downarrow 0} \frac{S_*(t-s|\mathbf{x})-S_*(t|\mathbf{x})}{S(t-s|\mathbf{x})-S(t|\mathbf{x})} & \textit{otherwise.} \end{cases} \quad (2.3)
$$

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

(2) the equality holds iff $S_*(t|\mathbf{x}) = S(t|\mathbf{x}) \ \forall\, t \in \mathcal{D}$ *and* $\forall\, \mathbf{x}$ *(see Eq. (1.2)).*

Proof. Treat $h_*(t,s|\mathbf{x}) = \mathbf{1}(s=0)\frac{S_*(t|\mathbf{x})}{S(t|\mathbf{x})} + \mathbf{1}(s=1)\frac{g_*(t|\mathbf{x})}{g(t|\mathbf{x})}$ as the df of $F_*(t,s|\mathbf{x})$ w.r.t. the measure $dF(t,s|\mathbf{x})$, where $F(t,s,\mathbf{x}) = F(t,s|\mathbf{x})F_{\mathbf{X}}(\mathbf{x})$ induced by $(M,\delta,\mathbf{x})$. Then the df of $F(t|\mathbf{x})$ 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 $S_*(t|\mathbf{x}) = S(t|\mathbf{x}) \ \forall\, t \in \mathcal{D}$. Thus

$0 \leq \int\int h(t,s|\mathbf{x})\ln\frac{h(t,s|\mathbf{x})}{h_*(t,s|\mathbf{x})}dF(t,s|\mathbf{x})dF_{\mathbf{X}}(\mathbf{x})$

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

Theorem 2. *Given assume (A1), the SMLE* $(\hat{S}_o,\hat{\beta})$ *is consistent if (A2) holds.*

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 SMLE $(\hat{S}_o,\hat{\beta})$ is a function of (ω,n), say $(\hat{S}_o(\cdot)(\omega,n),\hat{\beta}_n(\omega))$, 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, $\hat{S}_o$ is a sequence of bounded non-increasing functions, Helly's selection theorem ensures that given any subsequence of $\hat{S}_o$, there exists a further subsequence which is convergent. By taking convergent subsequence, WLOG, we can assume that $\hat{S}_o \to S_*$ and $\hat{\beta} \to \beta_*$, where S_* is nonincreasing and $S_*(0-) = 1$ and $S_*(\infty) = 0$. Moreover, $S_*(t|\mathbf{x}) = (S_*(t))^{\exp(\mathbf{x}'\beta_*)}$.

Since $(\hat{S}_o,\hat{\beta})$ is the SMLE, $\frac{1}{n}\ln\mathcal{L}(\hat{S}_o,\hat{\beta}) \geq \frac{1}{n}\ln\mathcal{L}(S_o,\beta)$, that is,

$$\begin{aligned}&\int \ln\hat{S}(t|\mathbf{x})d\hat{F}_n(t,0,\mathbf{x}) + \int \ln(\hat{S}(t-\eta_n|\mathbf{x}) - \hat{S}(t|\mathbf{x}))d\hat{F}_n(t,1,\mathbf{x}) \qquad \text{(by (1.4))}\\ \geq &\int \ln S(t|\mathbf{x})d\hat{F}_n(t,0,\mathbf{x}) + \int \ln(S(t-\eta_n|\mathbf{x}) - S(t|\mathbf{x}))d\hat{F}_n(t,1,\mathbf{x}),\end{aligned}$$

where $\eta_n = \frac{1}{n} \wedge min\{|M_i - M_j| : \ M_i \neq M_j, \ i,j \in \{1,...,n\}\}$. The last inequality yields

$$0 \geq \int \ln\frac{S(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})}d\hat{F}_n(t,0,\mathbf{x}) + \int \ln\frac{(S(t-\eta_n|\mathbf{x}) - S(t|\mathbf{x}))}{(\hat{S}(t-\eta_n|\mathbf{x}) - \hat{S}(t|\mathbf{x}))}d\hat{F}_n(t,1,\mathbf{x}). \tag{2.4}$$

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

$$\varliminf_{n\to\infty}\int \ln\frac{S(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})}d\hat{F}_n(t,0,\mathbf{x}) \geq \int \ln\frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})}dF(t,0,\mathbf{x}) \text{ (in Lemma 2)}, \tag{2.5}$$

$$\underline{\lim}_{n\to\infty}\int \ln\frac{(S(t-\eta_n|\mathbf{x})-S(t|\mathbf{x}))}{(\hat{S}(t-\eta_n|\mathbf{x})-\hat{S}(t|\mathbf{x}))}d\hat{F}_n(t,1,\mathbf{x}) \geq \int \ln\frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})}dF(t,1,\mathbf{x}) \quad (2.6)$$

(in Lemma 3), where $g(t|\mathbf{x}) \equiv 1$ (see Proposition 2) and $g_*(t|\mathbf{x})$ is as in (2.3). Then

$$0 \geq \int \ln\frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})}dF(t,0,\mathbf{x}) + \int \ln\frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})}dF(t,1,\mathbf{x}) \text{ (by Eq.s (2.4), (2.5) and (2.6))}$$

$\geq$0 (by Proposition 2).

In other words, $\int \ln\frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})}dF(t,0,\mathbf{x}) + \int \ln\frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})}dF(t,1,\mathbf{x}) = 0$. Thus $S_*(t|\mathbf{x}) = S(t|\mathbf{x})$ (*i.e.*, $(S_*(t))^{\exp(\beta_*'\mathbf{x})} = (S_o(t))^{\exp(\beta'\mathbf{x})}$) $\forall$ $(t,\mathbf{x}) \in \mathcal{D}_{M,\mathbf{X}}$ by Statement (2) of Proposition 2. As a consequence, $(S_*(t),\beta_*) = (S_o(t),\beta)$ $\forall$ $t \in \mathcal{D}$ by Theorem 1. Recall $P(\Omega_o) = 1$, thus the SMLE $(\hat{S}_o(t),\hat{\beta})$ is consistent for $t \in \mathcal{D}$. ▫

Hereafter, we prove Lemmas 2 and 3 needed in the proof of Theorem 2. We shall make use of Fatou's Lemma with varying measures (see Lemma 1 below).

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

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

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

(3) If f_n are bounded below then $\int \ln \underline{\lim}_{n\to\infty} f_n\,d\mu \leq \underline{\lim}_{n\to\infty}\int \ln f_n\,d\mu_n$.

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

The proof is relegated to a technical report (see Yu (2021)) for simplicity.

Lemma 2. *Inequality (2.5) in the proof of Theorem 2 holds, that is,*

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

Proof. For the given $\omega \in \Omega_o$ and (S_*,β_*) in Eq. (2.5), as assumed, $(\hat{S}_o,\hat{\beta})(\omega) \to (S_*,\beta_*)$. $S_*(t|\mathbf{x}) = (S_*(t))^{\exp(\beta_*'\mathbf{x})}$, which is a continuous function of S_* and β_*.

Consequently, $\hat{S}(\cdot|\cdot) \to S_*(\cdot|\cdot)$. Let $\mu_n(B) = \int_B \frac{\hat{S}(t|\mathbf{x})}{S(t|\mathbf{x})} d\hat{F}_n(t,0,\mathbf{x})$, where B is a measurable set in $\mathcal{R}^{p+1}$. To apply Lemma 1, let

$$K(t,0,\mathbf{x}) \stackrel{def}{=} \frac{1}{S(t|\mathbf{x})} \quad (\geq \frac{\hat{S}(t|\mathbf{x})}{S(t|\mathbf{x})}), \text{ and } K(t,0,\mathbf{x}) \text{ is integrable, as} \tag{2.7}$$

$$\begin{aligned}
\int K(t,0,\mathbf{x}) d\hat{F}_n(t,0,\mathbf{x}) &= \int \frac{1}{S(t|\mathbf{x})} d\hat{F}_n(t,0,\mathbf{x}) \\
&\to \int \frac{1}{S(t|\mathbf{x})} dF(t,0,\mathbf{x}) \text{ (as } \omega \in \Omega_o) \qquad (2.8)\\
&= \int \frac{1}{S(t|\mathbf{x})} S(t|\mathbf{x}) dF_C(t) dF_{\mathbf{X}}(\mathbf{x}) \text{ (by (1.1)).}
\end{aligned}$$

$$\begin{aligned}
&\lim_{n\to\infty} \mu_n(B) = \lim_{n\to\infty} \int_B \frac{\hat{S}(t|\mathbf{x})}{S(t|\mathbf{x})} d\hat{F}_n(t,0,\mathbf{x}) \qquad (2.9)\\
&= \int_B \lim_{n\to\infty} \frac{\hat{S}(t|\mathbf{x})}{S(t|\mathbf{x})} dF(t,0,\mathbf{x}) \text{ (by statement (2) of Lemma 1, (2.7) and (2.8))}\\
&= \int_B \frac{S_*(t|\mathbf{x})}{S(t|\mathbf{x})} dF(t,0,\mathbf{x}) \quad (= \int_B \frac{S_*(t|\mathbf{x})}{S(t|\mathbf{x})} S(t|\mathbf{x}) dF_C(t) dF_{\mathbf{X}}(\mathbf{x}) \quad (see\ Eq.(2.2)))\\
&= \int_B dF_*(t,0,\mathbf{x}) \stackrel{def}{=} \mu(B) \quad \text{(see Eq. (2.2)).} \qquad (2.10)
\end{aligned}$$

Verify that

$$\int \ln \frac{S(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t,0,\mathbf{x}) = \int H(\frac{S(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})}) \frac{\hat{S}(t|\mathbf{x})}{S(t|\mathbf{x})} d\hat{F}_n(t,0,\mathbf{x}), \text{ where} \tag{2.11}$$

$$H(t) = t\log t \geq -1/e \text{ for } t > 0 \text{ and } H(S(t|\mathbf{x})/\hat{S}(t|\mathbf{x})) \geq -1/e. \tag{2.12}$$

$$\begin{aligned}
&\varliminf_{n\to\infty} \int \ln \frac{S(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})} d\hat{F}_n(t,0,\mathbf{x}) = \varliminf_{n\to\infty} \int H(\frac{S(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})}) \frac{\hat{S}(t|\mathbf{x})}{S(t|\mathbf{x})} d\hat{F}_n(t,0,\mathbf{x}) && \text{(by (2.11))}\\
&= \varliminf_{n\to\infty} \int H(\frac{S(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})}) d\mu_n(t,\mathbf{x}) && (see\ (2.9))\\
&\geq \int \lim_{n\to\infty} H(\frac{S(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})}) d\mu(t,\mathbf{x}) \text{ (by (2.10), (2.12) and Corollary 1)}\\
&= \int \lim_{n\to\infty} H(\frac{S(t|\mathbf{x})}{\hat{S}(t|\mathbf{x})}) dF_*(t,0,\mathbf{x}) && \text{(see (2.10))}\\
&= \int \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} \ln \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})} dF_*(t,0,\mathbf{x}) && \text{(see (2.11))}
\end{aligned}$$

$$=\int\int \frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})}\ln\frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})}S_*(t|\mathbf{x})dF_C(t)dF_{\mathbf{X}}(\mathbf{x}) \qquad \text{(by Eq. (2.2))}$$
$$=\int\int \ln\frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})}S(t|\mathbf{x})dF_C(t)dF_{\mathbf{X}}(\mathbf{x})$$
$$=\int \ln\frac{S(t|\mathbf{x})}{S_*(t|\mathbf{x})}dF(t,0,\mathbf{x}), \text{ which is (2.5)). } \square$$

Lemma 3. *Inequality (2.6) in the proof of Theorem 2 holds with* $g(t|\mathbf{x})\equiv 1$ *and* $g_*(t|\mathbf{x})$ *as in (2.3). That is,* $\int \ln\frac{g(t|\mathbf{X})}{g_*(t|\mathbf{X})}dF(t,1,\mathbf{x}) \le \underline{\lim}_{n\to\infty}\int \ln\frac{S(t-\eta_n|\mathbf{X})-S(t|\mathbf{X})}{\hat S(t-\eta_n|\mathbf{X})-\hat S(t|\mathbf{X})}d\hat F_n(t,1,\mathbf{x})$.

Proof. For the given $\omega\in\Omega_o$, $\hat S(t|\mathbf{x})(\omega)$ and (S_*,β_*) in the proof of Theorem 2, denote $G(t,\mathbf{x},n)=\frac{\hat S(t-\eta_n|\mathbf{X})-\hat S(t|\mathbf{X})}{S(t-\eta_n|\mathbf{X})-S(t|\mathbf{X})}$, $A_k=\{(t,\mathbf{x}):\ G(t,\mathbf{x},n)\le k,\ \forall\ n\}$ and $B_k=A_k\setminus A_{k-1}$, $k\ge 1$. Then $\overline{\lim}_{n\to\infty}G(t,\mathbf{x},n)=\overline{\lim}_{n\to\infty}\frac{\hat S(t-\eta_n|\mathbf{X})-\hat S(t|\mathbf{X})}{S(t-\eta_n|\mathbf{X})-S(t|\mathbf{X})}=g_*(t|\mathbf{x})=\frac{g_*(t|\mathbf{X})}{g(t|\mathbf{X})}$ as $g(t|\mathbf{x})\equiv 1$. Since $G(t,\mathbf{x},n)$ is finite for each n, provided that $t\in\mathcal{D}_Y$ (see (1.3)), we have
$$\int \mathbf{1}(\cup_{k\ge 1}B_k)dF(t,s,\mathbf{x})=1. \tag{2.13}$$
For each $k\ge 1$, let $a_k\overset{def}{=}\ln\frac{S(t-\eta_n|\mathbf{X})-S(t|\mathbf{X})}{\hat S(t-\eta_n|\mathbf{X})-\hat S(t|\mathbf{X})}\mathbf{1}((t,\mathbf{x})\in B_k)$.

$$\underline{\lim}_{n\to\infty}\int_{B_k}\ln\frac{S(t-\eta_n|\mathbf{x})-S(t|\mathbf{x})}{\hat S(t-\eta_n|\mathbf{x})-\hat S(t|\mathbf{x})}d\hat F_n(t,1,\mathbf{x})$$
$$\ge\int_{B_k}\underline{\lim}_{n\to\infty}\ln\left(\frac{(S(t-\eta_n|\mathbf{x})-S(t|\mathbf{x}))}{(\hat S(t-\eta_n|\mathbf{x})-\hat S(t|\mathbf{x}))}\right)dF(t,1,\mathbf{x}) \text{ (by (2) of Corollary 1,}$$
$$\text{(as } a_1\in[0,\infty),\ a_k\in[\ln(1/k),\ln(1/(k-1))],\ k\ge 2))$$
$$=\int_{B_k}\ln\left(\underline{\lim}_{n\to\infty}\frac{(S(t-\eta_n|\mathbf{x})-S(t|\mathbf{x}))}{(\hat S(t-\eta_n|\mathbf{x})-\hat S(t|\mathbf{x}))}\right)dF(t,1,\mathbf{x}) \qquad \text{(as } \ln(x) \text{ is continuous)}$$
$$=\int_{B_k}\ln\frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})}dF(t,1,\mathbf{x}) \qquad \text{(see (2.3))}$$
$$=\int_{B_k}H\left(\frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})}\right)\frac{g_*(t|\mathbf{x})}{g(t|\mathbf{x})}dF(t,1,\mathbf{x}) \qquad \text{(see (2.11) and (2.12))}$$
$$=\int_{B_k}H\left(\frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})}\right)dF_*(t,1,\mathbf{x}) \qquad \text{(see (2.2))}$$
$$\ge\int_{B_k}(-1/e)dF_*(t,1,\mathbf{x})=(-1/e)\int_{B_k}1dF_*(t,1,\mathbf{x})\ge(-1/e)\int 1dF_*(t,s,\mathbf{x})\ge -1/e;$$

i.e.,
$$\underline{\lim}_{n\to\infty}\int_{B_k}\ln\frac{S(t-\eta_n|\mathbf{x})-S(t|\mathbf{x})}{\hat S(t-\eta_n|\mathbf{x})-\hat S(t|\mathbf{x})}d\hat F_n(t,1,\mathbf{x})\ge\int_{B_k}\ln\frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})}dF(t,1,\mathbf{x}) \tag{2.14}$$
$$\ge -1/e \text{ for } k\ge 1. \tag{2.15}$$

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

$$=\lim_{n\to\infty}\sum_{k\geq 1}\int_{B_k}\frac{S(t-\eta_n|\mathbf{x})-S(t|\mathbf{x})}{\hat{S}(t-\eta_n|\mathbf{x})-\hat{S}(t|\mathbf{x})}d\hat{F}_n(t,1,\mathbf{x}) \qquad \text{(by (2.13))}$$

$$=\lim_{n\to\infty}\int_{k\geq 1}\int_{B_k}\ln\frac{S(t-\eta_n|\mathbf{x})-S(t|\mathbf{x})}{\hat{S}(t-\eta_n|\mathbf{x})-\hat{S}(t|\mathbf{x})}d\hat{F}_n(t,1,\mathbf{x})d\nu(k) \quad (d\nu \text{ is a counting measure})$$

$$\geq\int_{k\geq 1}\lim_{n\to\infty}\int_{B_k}\ln\frac{S(t-\eta_n|\mathbf{x})-S(t|\mathbf{x})}{\hat{S}(t-\eta_n|\mathbf{x})-\hat{S}(t|\mathbf{x})}d\hat{F}_n(t,1,\mathbf{x})d\nu(k) \text{ (by (1) of Corollary 1 and (2.15))}$$

$$\geq\int_{k\geq 1}\int_{B_k}\ln\frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})}dF(t,1,\mathbf{x})d\nu(k) \qquad \text{(by (2.14))}$$

$$=\sum_{k\geq 1}\int_{B_k}\ln\frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})}dF(t,1,\mathbf{x})=\int\ln\frac{g(t|\mathbf{x})}{g_*(t|\mathbf{x})}dF(t,1,\mathbf{x}).$$ Thus (2.6) holds. □

Concluding Remark. The SMLE of β is a consistent estimator of β of the Lehmann family. If Y is continuous, then the mple is a consistent estimator of β of the Lehmann family too, but otherwise it is not.

No conflict of interest statement: The author states that there is no conflict of interest.

Acknowledgments. We thank the editor and two referees for their invaluable comments.

References

Casella, G. and Berger, R. (2001). Statistical Inference. 2nd Ed. Duxbury, NY.

Cox, D.R. (1972). "Regression Models and Life-Tables". *Journal of the Royal Statistical Society*, Series B. 34 (2): 187220.

Cox, D.R. and Oakes, D. (1984). Analysis of Survival Data. *Chapman & Hall*, NY.

Ferguson, T.S. (1996). A Course In Large Sample Theory. Chapman & Hall, NY.

Kiefer, J and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* 27 887-906.

Finkelstein, D.M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* 42 845-854.

Lehmann, E. L. (1959). Testing statistical hypotheses. 1st ed. *Wiley*, New York.

Royden, H.L. (1968). Real analysis. Macmillan. N.Y.

Sun, Jianguo (2006). The statistical analysis of interval-censored failure time data. *Spring*, NY.

Wong, G.Y.C and Yu, Q.Q (2012). Estimation under the Lehmann regression model with interval-censored data. *Comm. Statist. — Comput. Simul.* 41 1489-1500.

Yu, Q.Q., Chappell, R., Wong, G.Y.C., Hsu, Y.T., and Mazur, M. (2008). Relationship between the Cox, Lehmann, Weibull and accelerated lifetime models. *Comm. Statist. A—Theory and methods*. 37, 1458-1470.

Yu, Q.Q. (2021). Technical Report for "Consistency of the semi-parametric MLE of the Lehmann family with right-censored data".
http://people.math.binghamton.edu/qyu/ftp/tech.cleh.pdf