Technical Report to

"Consistency Of The MMGLE Under The Piecewise Proportional Hazards Models With Interval-Censored Data"

By Qiqing Yu and Qinggang Diao

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

Current version: 10/7/2018

AMS 1991 subject classification: Primary 62 G07; Secondary 62 J99.

Short Title: The PWPH model.

Key words and phrases: Cox's model, time-dependent covariates, semi-parametric MLE, identifiability, consistency, asymptotic normality.

Abstract: Wong *et al.* (2018) studied the piecewise proportional hazards (PWPH) model with interval-censored (IC) data under the distribution-free set-up. It is well known that the partial likelihood approach is not applicable for IC data, and Wong *et al.* (2018) showed that the standard generalized likelihood approach does not work neither. They proposed the maximum modified generalized likelihood estimator (MMGLE) and the simulation results suggest that the MMGLE is consistent. We establish the consistency and asymptotically normality of the MMGLE.

1. Introduction. We shall establish the asymptotic properties of the maximum modified generalized likelihood estimator (MMGLE) proposed by Wong *et al.* (2018) under the piece-wise proportional hazards (PWPH) model, with interval-censored (IC) continuous survival time Y. The proportional hazards (PH) model (Cox (1972)) is a common regression model. The PWPH model is a special PH model.

For a random variable Y, denote its survival function by $S_Y(t) = P(Y > t)$, its density function by $f_Y(t)$, and its hazard function by $h_Y(t) = \frac{f_Y(t)}{S_Y(t-)}$. Given a covariate (vector) $\mathbf{Z}$ which does not depend on time Y, $(\mathbf{Z}, Y)$ follows a time-independent covariate PH (TIPH) model if the conditional hazard function of $Y|\mathbf{Z}$ is

$$h(t|\mathbf{z}) = h_{Y|\mathbf{Z}}(t|\mathbf{z}) = h_o(t)e^{\boldsymbol{\beta}'\mathbf{Z}}, \text{ for } t < \tau, \tag{1.1}$$

where $\boldsymbol{\beta}'$ is the transpose of the $p \times 1$ vector $\boldsymbol{\beta}$, $\tau = \sup\{t : h_o(t) > 0\}$, and h_o is a hazard function.

IC data consist of n time intervals with the end-points $L_i \le R_i$, $i = 1, ..., n$, where the true survival time Y_i falls inside the interval. A realistic model for the IC data without exact observations is the mixed case interval censorship model (see Schick and Yu (2000)), which is specified as follows. Let K be the number of follow-up times for a patient. Conditional on $K = k$, Y and $(C_{k,1}, ..., C_{k,k})$ are independent, where $C_{k,1}$, ..., $C_{k,k}$ are the k follow-up times. Define $(L, R) = \sum_{i=0}^{K}(C_{K,i}, C_{K,i+1})\mathbf{1}(Y \in (C_{K,i}, C_{K,i+1}]$, where $\mathbf{1}(A)$ is the indicator function of an event A, $C_{k,0} = 0$ and $C_{k,k+1} = \infty$. Then (L_i, R_i) are i.i.d. from (L, R). For the PH model with IC data, it is assumed that $\mathbf{Z}$ and $(Y, K, \mathbf{C})$ are independent, where $\mathbf{C} = \{C_{ki} : \ i \in \{1, ..., k\}, k \ge 1\}$.

The PH model has been extended to the time-dependent covariates PH (TDPH) model (see, *e.g.*, Cox and Oak (1984, p. 115), Therneau and Grambsch (2000), Zhang and Huang (2006), or Wong *et al.* (2017)). A special case of the TDPH model is the PWPH model with k cut points formulated by Zhou (2001):

$$h(t|\mathbf{z}) = \sum_{i=0}^{k} h_o(t)e^{\beta_i z_i}\mathbf{1}(t \in [a_i, a_{i+1})), \text{ where } a_0 = 0 < a_1 < \cdots < a_{k+1} = \infty,$$

$\mathbf{z} = (z_0, z_1, ..., z_k)$ is a time-independent covariate vector. Wong *et al.* (2018) applied the PWPH model to analyze their interval-censored cancer research data.

The common approach in the semi-parametric set-up under the PH model is the partial likelihood approach. For the standard PH model, Finkelstein (1986) showed that this approach does not work if the data are interval censored and she proposed an approach based on the generalized likelihood:

$$\mathscr{L} = \mathscr{L}(\boldsymbol{\beta}, S_o) = \prod_{i=1}^{n} [(S(L_i|\mathbf{z}_i) - S(R_i|\mathbf{z}_i))^{1-\delta_i} (S(L_i - |\mathbf{z}_i) - S(R_i|\mathbf{z}_i))^{\delta_i}], \tag{1.2}$$

where $(L_1, R_1, \mathbf{z}_1)$, ..., $(L_n, R_n, \mathbf{z}_n)$ are IC observations, $\delta_i = \mathbf{1}(L_i = R_i)$, $S_o(\cdot) = S(\cdot|0)$ and $S(t|\mathbf{z})$ is the conditional survival function corresponding to $h(t|\mathbf{z})$ in (1.1). The semi-parametric maximum likelihood estimator (SMLE) of $(\boldsymbol{\beta}, S_o)$ maximizes $\mathscr{L}$ over all survival functions S_o and all possible values of $\boldsymbol{\beta}$.

Moreover, under the PWPH model, it is shown (see Example 2.1 in Wong *et al.* (2018)) that $\boldsymbol{\beta}$ can be non-identifiable if the following assumption is violated:

$$\exists\ a, b \in (\mathscr{S}_{F_L} \cup \mathscr{S}_{F_R}) \cap [c, \infty) \text{ such that } S_o(b) > S_o(a) > 0, \tag{1.3}$$

where given a random variable, say Y, $\mathscr{S}_{F_Y}$ is the support set of F_Y, in the sense that if $x \in \mathscr{S}_{F_Y}$ then $F_Y(x+\epsilon) - F_Y(x-\epsilon) > 0\ \forall\ \epsilon > 0$ and $\mathscr{S}_{F_L}$ and $\mathscr{S}_{F_R}$ are defined in a similar manner.

Furthermore, in general, the SMLE of β under the likelihood function (1.2) may not be unique (see Example 2.3 in Wong *et al.* (2018)). Both phenomena do not occur if $\mathbf{Z}$ is time-independent (see Wong and Yu (2012)). They further established the identifiability condition:

Lemma 1. (Wong *et al.* (2018)). *Assume $h(t|\mathbf{z}) = h_o(t)e^{\boldsymbol{\beta}'\mathbf{z}\mathbf{1}(t\geq c)}$. Under the mixed case IC model and assuming that S_o is absolutely continuous, the parameter $\boldsymbol{\beta}$ is identifiable if assumption (1.3) holds. The parameter $S_o(c)$ is identifiable if $\boldsymbol{\beta} \neq 0$ in addition to (1.3).*

Moreover, they proposed a modification to the generalized likelihood and proposed an algorithm to find the estimator of $(\boldsymbol{\beta}, S_o)$ that maximizes the modified generalized likelihood. Thus we call the estimator the MMGLE.

We shall give the proof of the consistency and asymptotic normality of the MMGLE in this short note.

2. The Main Results. We study consistency of the MMGLE under the PWPH model with one cut point assuming Y is continuous in this paper. In particular, we consider the model

$$h_{Y|\mathbf{Z}}(t|\mathbf{z}) = h_o(t)\exp(\mathbf{z}'\boldsymbol{\beta}\mathbf{1}(t \geq c)), \text{ where } \mathbf{Z} \text{ is a time-independent covariate vector.} \tag{2.1}$$

Y is subject to interval censoring under the mixed case IC model with the following up times C_{ki}'s and the random number of follow-up times K. It is easy to show (see *e.g.*, (Wong *et al.* (2018)) that under model (2.1), if Y is continuous, then $S_{Y|\mathbf{Z}}(t|\mathbf{z}) = \begin{cases} S_o(t) & \text{if } t \le c \\ (S_o(c))^{1-e^{\beta'\mathbf{Z}}}(S_o(t))^{e^{\beta'\mathbf{Z}}} & \text{if } t > c. \end{cases}$ Abusing notations, write $h(t|\mathbf{z}) = h_{Y|\mathbf{Z}}(t|\mathbf{z})$, $S(t|\mathbf{z}) = S_{Y|\mathbf{Z}}(t|\mathbf{z})$ and $f(t|\mathbf{z}) = f_{Y|\mathbf{Z}}(t|\mathbf{z})$. We assume that $\mathbf{Z}$ is a $p \times 1$ random vector and $\mathbf{Z}$ takes on p linearly independent values.

Let $A_1, ..., A_m$ be all the innermost intervals (II) induced by I_i's. If the covariates are time-independent, it is well known that in order to maximize $\mathscr{L}$, it suffices to put the weights of S_o to the right-end points of the IIs. Let t_j's be the right-end point of the II's, or c, or $\pm\infty$, and $t_0 = -\infty < t_1 < \cdots < t_{i_c} = c < t_{i_c+1} < \cdots < t_m = \infty$. Write $S_j = S_o(t_j)$. For each i, let (l_i, r_i) satisfy $\begin{cases} t_{r_i} \le R_i < t_{r_i+1} \text{ and } t_{l_i} \le L_i < t_{l_i+1} & \text{if } L_i < R_i < \infty \\ t_{r_i} = t_m \text{ and } t_{l_i} \le L_i < t_{l_i+1} & \text{if } L_i < R_i = \infty \\ t_{r_i} = R_i \text{ and } t_{l_i} = t_{r_i-1} & \text{if } R_i = L_i. \end{cases}$

Theorem 1. *Suppose that $h(t|\mathbf{z}) = h_o(t)e^{\boldsymbol{\beta}'\mathbf{z}\mathbf{1}(t\ge c)}$, Y is continuous and subject to the mixed case IC model, $E(K) < \infty$, and the identifiable condition in Lemma 1 is satisfied. Then the MMGLE of $(S_o, \boldsymbol{\beta})$ is consistent.*

Proof. We shall give the proof in 4 steps. Abusing notation, write $S_o^{(u)}(t) = S(t|u)$ and $S_o^{(0)}(t) = S_o(t)$. Let Ω be the sample space.

Step 1 (preliminary). Under the mixed interval censorship model, by (1.2), the normalized generalized log-likelihood becomes $Ł_n(S, \mathbf{b})$

$$= \frac{1}{n}\sum_{j=1}^{n} \log((S(c))^{1-e^{\mathbf{b}'\mathbf{u}_j\mathbf{1}_{(L_j\ge c)}}}(S(L_j))^{e^{\mathbf{b}'\mathbf{u}_j\mathbf{1}_{(L_j\ge c)}}} - (S(c))^{1-e^{\mathbf{b}'\mathbf{u}_j\mathbf{1}_{(R_j\ge c)}}}(S(R_j))^{e^{\mathbf{b}'\mathbf{u}_j\mathbf{1}_{(R_j\ge c)}}})$$
$$= \frac{1}{n}\sum_{j=1}^{n} \log(S^{(\mathbf{u}_j)}(L_j) - S^{(\mathbf{u}_j)}(R_j)),\ \{S^{(\mathbf{u}_j)}\} \in \mathscr{C}.$$

where $\mathscr{C}$ is the collection of all nonincreasing functions S from $[0,\infty)$ into $[0,1]$ with $S(0) = 1$ and $S(\infty) = 0$. By the strong law of large numbers (SLLN), $Ł_n(S, b)$ converges almost surely to its mean

$$Ł(S, \mathbf{b}) = E(\log(S^{(\mathbf{Z})}(L) - S^{(\mathbf{Z})}(R))) = E(E(E(w_{S^{(\mathbf{Z})}}(\mathbf{C}, K))|\mathbf{Z})|K), \text{ where}$$

$$w_{S^{(\mathbf{u})}}(\mathbf{C}, k) = (1 - S_o^{(\mathbf{u})}(C_{k1}))\log(1 - S^{(\mathbf{u})}(C_{k1})) + S_o^{(\mathbf{u})}(C_{kk})\log S^{(\mathbf{u})}(C_{kk})$$
$$+ \sum_{i=2}^{k}(S_o^{(\mathbf{u})}(C_{k,i-1}) - S_o^{(\mathbf{u})}(C_{ki}))\log(S^{(\mathbf{u})}(C_{k,i-1}) - S^{(\mathbf{u})}(C_{ki})).$$

Step 2. It can be verified that $w_{S^{(\mathbf{u})}}(\mathbf{c}, k)$ is maximized by a nonincreasing function $S^{(\mathbf{u})} \in \mathscr{C}$, iff $S^{(\mathbf{u})}(c_{ki}) =$

$S_o^{(\mathbf{u})}(c_{ki})$, $i \in \{1,...,k\}$. Since $\sup\{|p\log p| : 0 \le p \le 1\} \le 1$, $w_{S^{(\mathbf{u})}}(\mathbf{C}, K)$ is bounded by $K+1$, and thus Ł(S,b) is finite, as $E(K) < \infty$ by the assumption in the theorem. If the identifiable conditions hold, by Lemma 1 and the Shannon-Kolmogorov inequality, we can conclude that $(S^{(\mathbf{u})}(t), S^{(0)}(t)) = (S_o^{(\mathbf{u})}(t), S_o(t))$ $\forall\ t \in \mathscr{S}_{F_L} \cup \mathscr{S}_{F_R}$ and $\forall\ \mathbf{u} \in \mathscr{S}_{F_\mathbf{Z}}$. Recall that $\mathbf{Z}$ takes on p linearly independent values, say, $\mathbf{u}_1$, ..., $\mathbf{u}_p$. As a consequence, for $j \in \{1,...,p\}$, $\mathbf{b}'\mathbf{u}_j = \log(\frac{\log \frac{S^{(\mathbf{u}_j)}(t_2)}{S^{(\mathbf{u}_j)}(t_1)}}{\log \frac{S^{(0)}(t_2)}{S^{(0)}(t_1)}})$ and $\boldsymbol{\beta}'\mathbf{u}_j = \log(\frac{\log \frac{S_o^{(\mathbf{u}_j)}(t_2)}{S_o^{(\mathbf{u}_j)}(t_1)}}{\log \frac{S_o(t_2)}{S_o(t_1)}})$, where $c < t_1 < t_2 < \tau$ and $t_1, t_2 \in \mathscr{S}_{F_L} \cup \mathscr{S}_{F_R}$, $(S^{(\mathbf{u}_j)}(t_2), S^{(\mathbf{u}_j)}(t_1), S^{(0)}(t_2), S^{(0)}(t_1)) = (S_o^{(\mathbf{u}_j)}(t_2), S_o^{(\mathbf{u}_j)}(t_1), S_o(t_2), S_o(t_1))$. Thus $\mathbf{b} = \boldsymbol{\beta}$. Consequently, $(S_o, \boldsymbol{\beta})$ maximizes Ł$(S, \mathbf{b})$ and any other nonincreasing function $S \in \mathscr{C}$ and $\mathbf{b}$ satisfying Ł$(S,\mathbf{b}) =$ Ł$(S_o, \boldsymbol{\beta})$ satisfy $S = S_o$ a.s. μ (the measure induces by $dF_L + dF_R$) and $\mathbf{b} = \boldsymbol{\beta}$.

Step 3. $\vdash$: $\liminf_{n\to\infty}$ Ł$_n(\hat{S}_n, \hat{\mathbf{b}}_n) \ge \liminf_{n\to\infty}$Ł$_n(S_o, \beta) =$ Ł$(S_o, \boldsymbol{\beta})$ a.s..

Let $\Omega_0 = \{\omega \in \Omega :$ Ł$_n(S_o, \boldsymbol{\beta})(\omega) \to (S_o, \boldsymbol{\beta})\}$. Then $P(\Omega_0) = 1$ by the SLLN. Hereafter, we fix an $\omega \in \Omega_0$ and suppress it in the expressions of most random variables. For $n > 0$, let $B_n(\omega)$ be the collection of all the distinct points 0, L_i, R_i, c, where $1 \le i \le n$. Write $B_n = \{q_{n,j} :\ 1 \le j \le m_n\}$, where $0 = q_0 < q_{n,1} < ... < q_{n,m_n} = \infty$. Denote the intervals $A_{n,j} = (q_{n,j-1}, q_{n,j}]$ and let $p_{0,n,j} = S_o(q_{n,j-1}) - S_o(q_{n,j})$, $1 \le j \le m_n$. Then $\sum_{j=1}^{m_n} p_{0,n,j} = 1$ and $S_o(t) = \sum_{A_{n,j} \in (t,\infty)} p_{0,n,j}$ for each $t \in B_n$. Moreover,

$$\begin{aligned}
\text{Ł}_n(S_o, \boldsymbol{\beta})(\omega) =& \frac{1}{n}\sum_{j=1}^{n} \log\{S_o(c)^{1-e^{\boldsymbol{\beta}'\mathbf{u}_j \mathbf{1}_{(L_j \ge c)}}} S_o(L_j)^{e^{\boldsymbol{\beta}'\mathbf{u}_j \mathbf{1}_{(L_j \ge c)}}} - S_o(c)^{1-e^{\boldsymbol{\beta}'\mathbf{u}_j \mathbf{1}_{(R_j \ge c)}}} S_o(R_j)^{e^{\boldsymbol{\beta}'\mathbf{u}_j \mathbf{1}_{(R_j \ge c)}}}\} \\
=& \frac{1}{n}\sum_{j=1}^{n} \log\{(\sum_{A_{n,i} \in (c,\infty)} p_{0,n,i})^{1-e^{\boldsymbol{\beta}'\mathbf{u}_j \mathbf{1}_{(L_j \ge c)}}} (\sum_{A_{n,i} \in (L_j,\infty)} p_{0,n,i})^{e^{\boldsymbol{\beta}'\mathbf{u}_j \mathbf{1}_{(L_j \ge c)}}} \\
& - (\sum_{A_{n,i} \in (c,\infty)} p_{0,n,i})^{1-e^{\boldsymbol{\beta}'\mathbf{u}_j \mathbf{1}_{(R_j \ge c)}}} (\sum_{A_{n,i} \in (R_j,\infty)} p_{0,n,i})^{e^{\boldsymbol{\beta}'\mathbf{u}_j \mathbf{1}_{(R_j \ge c)}}}\}.
\end{aligned}$$

Now we assign weight $p_{n,i}$ to each interval $A_{n,i}$ with $\sum_{i=1}^{m_n} p_{n,i} = 1$.Then

$$\begin{aligned}
\text{Ł}_n(S, b)(\omega) = & \frac{1}{n}\sum_{j=1}^{n} \log\{(\sum_{A_{n,i} \in (c,\infty)} p_{n,i})^{1-e^{\boldsymbol{\beta}'\mathbf{u}_j \mathbf{1}_{(L_j \ge c)}}} (\sum_{A_{n,i} \in (L_j,\infty)} p_{n,i})^{e^{\boldsymbol{\beta}'\mathbf{u}_j \mathbf{1}_{(L_j \ge c)}}} \\
& - (\sum_{A_{n,i} \in (c,\infty)} p_{n,i})^{1-e^{\boldsymbol{\beta}'\mathbf{u}_j \mathbf{1}_{(R_j \ge c)}}} (\sum_{A_{n,i} \in (R_j,\infty)} p_{n,i})^{e^{\boldsymbol{\beta}'\mathbf{u}_j \mathbf{1}_{(R_j \ge c)}}}\}.
\end{aligned}$$

Let $\hat{S}_n^{(\mathbf{u})}(t) = (\sum_{A_{n,i} \in (c,\infty)} \hat{p}_{n,i})^{1-e^{\hat{\mathbf{b}}_n'\mathbf{u}\mathbf{1}_{(t \ge c)}}} (\sum_{A_{n,i} \in (t,\infty)} \hat{p}_{n,i})^{e^{\hat{\mathbf{b}}_n'\mathbf{u}\mathbf{1}_{(t \ge c)}}}$ be the GMLE of $S^{(\mathbf{u})}(t)$ under Ł$_n$. In particular, $\hat{S}_n^{(0)}(t) = \hat{S}_n(t) = \sum_{A_{n,i} \in (t,\infty)} \hat{p}_{n,i}$.

Let $\{S_n(x)\}$ be a sequence in $\mathscr{C}$. By a pointwise limit of this sequence we mean $S^* \in \mathscr{C}$ such that $S_{n'}(x) \to S^*(x)$ for all x and some sequence $\{n'\}_{n' \ge 1}$. Let $S^{(0)*}(t)$ be the pointwise limit function of $\hat{S}_n^{(0)}(t)$

for all t and for some subsequence $\{n'\}_{n'\geq 1}$. Helly's selection theorem guarantees the existence of pointwise limits. Let $\mathbf{b}^*$ be the limiting point of $\{\hat{\mathbf{b}}_n\}$ for some subsequence $\{n''\}_{n''\geq 1}$ of $\{n'\}$.

Since $Ł_n(\hat{S}_n,\hat{\mathbf{b}}_n)\geq Ł_n(S_o,\boldsymbol{\beta})$ by the definition of the GMLE, the claim in Step 3 is proved.

Step 4 (Conclusion). Let $\hat{Q}_n$ denote the empirical estimator of Q, the distribution of $(L,R,\mathbf{Z})$ and $\Omega' = \{\omega \in \Omega :\ \hat{Q}_n(l,r,\mathbf{z})(\omega) \to Q(l,r,\mathbf{z})\text{−pointwisely in } (l,r,\mathbf{z})\}$. By the SLLN, $P(\Omega') = 1$. $\Omega_U = \{\hat{Q}_n(U) \to Q(U)\}$ a.s. for every Borel subset U of $\Delta = \{(l,r,\mathbf{u}) : 0 \leq l < r \leq \infty, \mathbf{u} \in \mathscr{S}_{F_\mathbf{Z}}\}$. Let S_n denote the survival function defined by $S_n(x) = \hat{S}_n(x;\omega)$, $\mathbf{b}_n$ defined by $\mathbf{b}_n = \hat{\mathbf{b}}_n(\omega)$, and Q_n the measure defined by $Q_n(A) = \hat{Q}_n(A;\omega)$. For simplicity in notation we shall assume that $S_n(x) \to S^*(x)$ for all $x \in R$ and $\mathbf{b}_n \to \mathbf{b}^*$.

Let $\omega \in \Omega' \cap \Omega_0$ hereafter. $\liminf_{n\to\infty} Ł_n(\hat{S}_n,\hat{\mathbf{b}}_n) \geq Ł(S_o,\boldsymbol{\beta})$, $S_n(t) \to S^*(t)$, for all $t \in R$ and $\mathbf{b}_n \to \mathbf{b}^*$. We shall show that

$$Ł(S_o,\boldsymbol{\beta}) \leq \liminf_{n\to\infty} Ł_n(\hat{S}_n,\hat{\mathbf{b}}_n)(\omega) \leq \limsup_{n\to\infty} Ł_n(\hat{S}_n,\hat{\mathbf{b}}_n)(\omega) \leq Ł(S^*,\mathbf{b}^*). \tag{2.2}$$

By the previous discussion, it suffices to prove the last inequality.

Now let $S_n^{(\mathbf{u})}(t) = S_n(c)^{1-e^{\mathbf{b}_n'\mathbf{u}\mathbf{1}_{(t\geq c)}}} S_n(t)^{e^{\mathbf{b}_n'\mathbf{u}\mathbf{1}_{(t\geq c)}}}$. Since

$$Ł_n(\hat{S}_n,\hat{\mathbf{b}}_n)(\omega) = \int_\Delta \log(S_n^{(\mathbf{u})}(l) - S_n^{(\mathbf{u})}(r))\,dQ_n(l,r,\mathbf{u}),$$

the desired inequality is thus equivalent to

$$\limsup_{n\to\infty} \int_\Delta \log(S_n^{(\mathbf{u})}(l) - S_n^{(\mathbf{u})}(r))\,dQ_n(l,r,\mathbf{u}) \leq \int_\Delta \log(S^{(\mathbf{u})*}(l) - S^{(\mathbf{u})*}(r))\,dQ(l,r,\mathbf{u}). \tag{2.3}$$

The inequality is proved in Lemma 2. It follows from inequality (2.2) that $Ł(S^*,\mathbf{b}^*) \geq Ł(S_o,\boldsymbol{\beta})$. As (S_o,β) maximizes L, we can conclude that $Ł(S^*,\mathbf{b}^*) = Ł(S_o,\boldsymbol{\beta})$ and therefore $S^* = S_o$, a.s. μ. If the identifiable conditions (1.3) holds, we have $\mathbf{b}^* = \boldsymbol{\beta}$. □

In order to prove Inequality (2.3), we will introduce the Fatou's Lemma with varying measures.

Theorem 2. Suppose that μ_n is a sequence of measures on the measurable space (S,Σ) such that $\mu_n(B) \to \mu(B)$, $\forall B \in \Sigma$. If f_n non-negative integrable functions and $f = \liminf_{n\to\infty} f_n$, then

$$\int_S f\,d\mu \leq \liminf_{n\to\infty} \int_S f_n\,d\mu_n.$$

Theorem 2 is almost the same as Proposition 17 in Royden (1968), page 231, and so is the proof of Theorem 2. The proof of Theorem 2 can also be found in the Appendix.

Lemma 2. Inequality (2.3) holds.

Proof of Lemma 2 Since $\liminf_{n\to\infty} -\log(S_n^{(u)}(l) - S_n^{(u)}(r)) = -\log(S^{(u)*}(l) - S^{(u)*}(r))$, and $-\log(S_n^{(u)}(l) - S_n^{(u)}(r)) \geq 0$. $Q_n(U) \to Q(U)$ for every Borel subset U of Δ, where $\Delta = \{(l, r, u) : -\infty \leq l < r \leq \infty, u \in \mathscr{D}_{\mathbf{Z}}\}$.

Thus an application of Theorem 2 yields

$$\begin{aligned}
\limsup_{n\to\infty} \int_\Delta \log(S_n^{(u)}(l) - S_n^{(u)}(r))dQ_n(l,r,u) &= -\liminf_{n\to\infty}\int_\Delta -\log(S_n^{(u)}(l) - S_n^{(u)}(r))dQ_n(l,r,u) \\
&\leq -\int_\Delta \liminf_{n\to\infty} -\log(S_n^{(u)}(l) - S_n^{(u)}(r))dQ_n(l,r,u) \\
&= \int_\Delta \log(S^{(u)*}(l) - S^{(u)*}(r))dQ(l,r,u). \ \square
\end{aligned}$$

Theorem 3. *Suppose that the assumptions in Theorem 1 hold and the support set $\mathscr{S}_{F_L} \cup \mathscr{S}_{F_R} \cup \mathscr{S}_{F_{\mathbf{Z}}}$ contains finitely many elements. Then the MMGLE of (S_o, β) is asymptotically normally distributed.*

Proof. By assumption $\mathscr{S}_{F_L} \cup \mathscr{S}_{F_R} = \{t_j\}_{j=0}^m$ and m is finite. Then the parameter (S_o, β) can be represented by $(S_o(t_0), ..., S_(t_m), \beta)$, and the problem becomes an estimation problem of a multinomial distribution subject to certain constraints. Thus the asymptotic normality follows and the asymptotic covariance matrix can be estimated by the inverse of the empirical Fisher information matrix. $\square$

Acknowledgement. The authors thank the editor, an associate editor and two referees for their invaluable comments.

References.

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

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

Groeneboom, P. and Wellner, J.A. (1992), *Information bounds and nonparametric maximum likelihood estimation*, Birkhäuser Verlag, Basel.

Royden, H.L. (1968), *Real Analysis*, Second edition, New York: Macmillan.

Schick, A. and Yu, Q.Q. (2000), 'Consistency of the GMLE with mixed case interval-censored data', *Scandinavian Journal of Statistics*, 27, 45-55.

Therneau, T., and Grambsch, P. (2000), *Modeling survival data : extending the Cox model*, New York: Springer.

Wong, G. Y. C., Osborne, M. P., Diao, Q.G., and Yu, Q.Q. (2017), 'Piece-wise Cox Models with right-censored data', *communication in statistics simulation and computation*, 46, 7894-7908.

Wong, G. Y. C., Diao, Q.G., and Yu, Q.Q. (2018), 'Piece-wise Proportional Hazards Models with interval-censored data', *Journal of Statistical Computation and Simulation*, 88, 140-155.

Zhang H.M. and Huang C.F. (2006), 'Nonparametric survival analysis on time-dependent covariate effects in case-cohort sampling design', *Statistica Sinica*, 16, 267-285.

Zhou, M. (2001), 'Understanding the Cox regression models with time-change covariates', *American Statistian*, 55, 153-155

Appendix

Proof of Theorem 2. We will prove something a bit stronger here. Namely, we will allow f_n to converge μ-almost everywhere on a subset B of S. We seek to show that $\int_B f\,d\mu \le \liminf_{n\to\infty} \int_B f_n\,d\mu_n$.

Let $\mathscr{K} = \{x \in B | f_n(x) \to f(x)\}$. Then $\mu(B \backslash \mathscr{K}) = 0$ and $\int_B f\,d\mu = \int_{B\backslash\mathscr{K}} f\,d\mu$, and $\int_B f_n\,d\mu = \int_{B\backslash\mathscr{K}} f_n\,d\mu\ \forall\ n \in N$. Thus, replacing B by $B \setminus \mathscr{K}$ we may assume that f_n converge to f pointwise on B.

Recall that a simple function ϕ is of the form that $\phi(x) = \sum_{i=1}^{k} \alpha_k \mathbf{1}(x \in A_i)$, where A_i's are disjoint measurable sets. Given a simple function ϕ we have $\int_B \phi\,d\mu = \lim_{n\to\infty} \int_B \phi\,d\mu_n$. Hence, by the definition of the Lebesgue Integral, it is enough to show that if ϕ is any non-negative simple function less than or equal to f, then $\int_B \phi\,d\mu \le \liminf_{n\to\infty} \int_B f_n\,d\mu_n$

Let a be the minimum non-negative value of ϕ. Define $A = \{x \in B :\ \phi(x) > a\}$.

We first consider the case when $\int_B \phi\,d\mu = \infty$. We must have that $\mu(A)$ is infinite since $\int_B \phi\,d\mu \le M\mu(A)$, where M is the (necessarily finite) maximum value of that ϕ attains.

Next, we define $A_n = \{x \in B :\ f_k(x) > a\ \forall\ k \ge n\}$. We have that $A \subseteq \bigcup_n A_n \Rightarrow \mu(\bigcup_n A_n) = \infty$. But A_n is a nested increasing sequence of functions, $\lim_{n\to\infty} \mu(A_n) = \mu(\lim_{n\to\infty} A_n) = \infty$. Thus, $\lim_{n\to\infty} \mu_n(A_n) = \mu(A) = \infty$.

At the same time, $\int_B f_n\, d\mu_n \geq a\mu_n(A_n) \Rightarrow \liminf_{n\to\infty} \int_B f_n\, d\mu_n = \infty = \int_B \phi\, d\mu$, proving the claim in this case.

It suffices to prove the theorem in the case $\int_B \phi\, d\mu < \infty$. We must have that $\mu(A)$ is finite. Denote, as above, by M the maximum value of ϕ and fix $\epsilon > 0$. Define $A_n = \{x \in B | f_k(x) > (1-\epsilon)\phi(x)\ \forall\ k \geq n\}$. Then A_n is a nested increasing sequence of sets whose union contains A. Thus, $A - A_n$ is a decreasing sequence of sets with empty intersection. Since A has finite measure (this is why we needed to consider the two separate cases), $\lim_{n\to\infty} \mu(A - A_n) = 0$. Thus, there exists n such that $\mu(A - A_k) < \epsilon,\ \forall\ k \geq n$. Since $\lim_{n\to\infty} \mu_n(A - A_k) = \mu(A - A_k)$, there exists N such that $\mu_k(A - A_k) < \epsilon,\ \forall\ k \geq N$. Hence, for $k \geq N$,

$$\int_B f_k\, d\mu_k \geq \int_{A_k} f_k\, d\mu_k \geq (1-\epsilon) \int_{A_k} \phi\, d\mu_k.$$

At the same time, $\int_B \phi\, d\mu_k = \int_A \phi\, d\mu_k = \int_{A_k} \phi\, d\mu_k + \int_{A-A_k} \phi\, d\mu_k$. Hence,

$$(1-\epsilon) \int_{A_k} \phi\, d\mu_k \geq (1-\epsilon) \int_B \phi\, d\mu_k - \int_{A-A_k} \phi\, d\mu_k.$$

These inequalities yields that

$$\int_B f_k\, d\mu_k \geq (1-\epsilon) \int_B \phi\, d\mu_k - \int_{A-A_k} \phi\, d\mu_k \geq \int_B \phi\, d\mu_k - \epsilon \left(\int_B \phi\, d\mu_k + M \right).$$

Hence, letting $\epsilon \to 0$ and taking the $\liminf$ in n, we get that

$$\liminf_{n\to\infty} \int_B f_n\, d\mu_k \geq \int_B \phi\, d\mu. \ \square$$