Ann. Inst. Statist. Math.
Vol. 53, No. 3, 469–486 (2001)

ASYMPTOTIC PROPERTIES OF SELF-CONSISTENT ESTIMATORS WITH MIXED INTERVAL-CENSORED DATA

QIQING YU[1*], GEORGE Y. C. WONG[2*] AND LINXIONG LI[3**]

[1]*Department of Mathematical Sciences, SUNY, Binghamton, NY 13902-6000, U.S.A.*
[2]*Strang Cancer Prevention Center, Cornell University Medical College, 428 E 72nd Street, NY 10021, U.S.A.*
[3]*Department of Mathematics, University of New Orleans, Lake Front, New Orleans, LA 70148, U.S.A.*

(Received September 6, 1999; revised April 24, 2000)

Abstract. Mixed interval-censored (MIC) data consist of n intervals with endpoints L_i and R_i, $i = 1, \dots, n$. At least one of them is a singleton set and one is a finite non-singleton interval. The survival time X_i is only known to lie between L_i and R_i, $i = 1, 2, \dots, n$. Peto (1973, *Applied Statistics*, **22**, 86–91) and Turnbull (1976, *J. Roy. Statist. Soc. Ser. B*, **38**, 290–295) obtained, respectively, the generalized MLE (GMLE) and the self-consistent estimator (SCE) of the distribution function of X with MIC data. In this paper, we introduce a model for MIC data and establish strong consistency, asymptotic normality and asymptotic efficiency of the SCE and GMLE with MIC data under this model with mild conditions.

Key words and phrases: Asymptotic normality, generalized maximum likelihood estimator, mixture distribution, strong consistency.

1. Introduction

Interval censoring refers to a situation in which, X, the time to occurrence of an event of interest is only known to lie in a half-open and half-closed time interval $(L, R]$, where the pair (L, R) is an extended random vector such that $-\infty \le L < X \le R \le \infty$. Interval-censored (IC) data may occur in medical follow-up studies when each patient had several visits and the event of interest was only known to take place either before the first visit, between two consecutive visits, or after the last one. Thus an IC data set may consist of strictly interval-censored (SIC) observations (i.e.,$0 < L < R < \infty$), and right-censored ($R = \infty$) and/or left-censored ($L = -\infty$) observations. Examples of IC data can be found in cancer research and AIDS studies (see, e.g., Finkelstein and Wolfe, (1985)).

Case 1 data (or current status data, see Ayer *et al.* (1955)) is a special case of IC data when each patient had only one visit. Observations in a case 1 data set are either left-censored or right-censored. Doubly-censored data (see Chang and Yang (1987)) consist of case 1 data and uncensored observations. It is clear that neither case 1 data nor IC data contain uncensored observations. Furthermore, doubly-censored data do not contain SIC observations. A data set may be a mixture of uncensored observations

*Partially supported by Army Grant DAMD17-99-1-9390.
**Partially supported by BoRSF Grant RD-A-31.

and IC data which contain SIC observations. We call such data *mixed interval-censored* (MIC) *data*.

MIC data arise in clinical follow-up studies. In a cancer follow-up study, a patient whose tumor marker value (for instance, CA 125 in ovarian cancer) is consistently on the high (or low) end of the normal range in repeated testing is usually monitored very closely for possible relapse. If such a patient should relapse, then time to clinical relapse can often be accurately determined, and an uncensored observation is obtained. However, if a patient is not under close surveillance, and would seek help only after some tangible symptoms of the disease have appeared, then time to relapse most likely has to be specified to be within the dates of two successive clinical visits.

Another situation in which MIC data can occur is in the usual right-censored survival analysis where actual dates of events are not recorded, or missing, for a subset of the study population, and can be established only to within specified intervals. An example from the Framingham Heart Study was presented by Odell *et al.* (1992). In this large-scale longitudinal heart disease study, times of occurrence of coronary heart disease were recorded for almost every participant. However, time of first occurrence of the coronary heart disease subcategory angina pectoris was only recorded for about 20% of the participants who suffered from angina pectoris, and may be specified only as between two clinical visits, several years apart, for the other participants.

For censored data, Peto (1973) proposed a Newton-Raphson algorithm to obtain the generalized MLE (GMLE) of the distribution function (cdf), F. Turnbull (1976) obtained a self-consistent estimator (SCE) of the cdf via an EM-algorithm. A detailed discussion of more efficient algorithms for obtaining the GMLE is given in Wellner and Zhan (1997).

For IC data, Groeneboom and Wellner (1992) formulated the case 2 model; Wellner (1995) formulated a case k model, where $k \geq 1$; Schick and Yu (2000) modified Wellner's case k model by further assuming that k, the number of visits by a patient in a follow-up study, is a random integer and the observation (L, R) is a mixture of various case k models.

Various asymptotic distribution results of the GMLE have been obtained for censored data. For case 1 model the GMLE is asymptotically normally distributed (a.n.) and the convergence rate is $n^{1/2}$ if the underlying censoring distribution is discrete (Yu *et al.*, (1998*b*)), but the GMLE is not a.n. and the convergence rate is $n^{1/3}$ if cdfs have positive derivatives (Groeneboom and Wellner (1992)). For case 2 model the GMLE is a.n. with rate $n^{1/2}$ if the censoring vector takes on finitely many values (Yu *et al.*, (1998*c*)), and Groeneboom and Wellner's (1992) conjecture that under certain smoothness conditions the GMLE has a pointwise convergence rate of $(n\ln n)^{1/3}$. For more recent development on the latter conjecture, we refer to Groeneboom (1996) and Van De Geer (1996) .

For MIC data, several models have been proposed, and the asymptotic properties of the GMLE have been investigated under the assumptions that either the censoring vector takes on finitely many values (see Petroni and Wolfe (1994), and Yu *et al.* (1998*a*, 2000)), or the censoring and survival distributions are strictly increasing and continuous, and they have "positive separation" (see Huang (1999)).

In this paper, we shall use the model in Yu *et al.* (1998*a*) to establish asymptotic properties of the GMLE based on MIC data under the assumption that all underlying distributions are arbitrary with some mild conditions. Since a GMLE is also an SCE (but an SCE may not be a GMLE; see Yu *et al.* (1998*a*)), and our proofs basically use the properties of SCEs, we shall focus on the asymptotic properties of SCEs for MIC data. The main results are given in Section 2. The consistency result is proved in Section 3 and the asymptotic normality result is proved in Section 4. Some detailed proofs of lemmas in Sections 3 and 4 are relegated to Appendices A and B.

2. Main results

We introduce a *mixture interval censorship* model to characterize MIC data. Let (T, U, V) be a random censoring vector and $\mathcal{K}$ a random integer taking values 0 and 2. Assume that X and $(\mathcal{K}, T, U, V)$ are independent. The observable extended random vector (L, R) is generated by a two-stage experiment. In the first stage, a value of $\mathcal{K}$ is selected, then the random variable (L, R) corresponds to the observation from a right censorship model if $\mathcal{K} = 0$ and from a case 2 model if $\mathcal{K} = 2$, i.e.,

$$(2.1)\quad (L,R) = \begin{cases} (X,X)\mathbf{1}_{(X\le T)} + (T,\infty)\mathbf{1}_{(X>T)} & \text{if } \mathcal{K}=0, \\ (-\infty,U)\mathbf{1}_{(X\le U)} + (U,V)\mathbf{1}_{(U<X\le V)} + (V,\infty)\mathbf{1}_{(X>V)} & \text{if } \mathcal{K}=2, \end{cases}$$

where $\mathbf{1}_{(A)}$ is the indicator function of the set A and $(\cdot,\cdot)$ stands for a vector. The notation $(\cdot,\cdot)$ may stand for an open interval later, but it would be clear from the context or otherwise we would point out. MIC data in (2.1) can be considered as a "mixture" of right-censored data and interval-censored data. It is known that in order to estimate F, we only need to observe (L, R) (see Peto (1973)). Thus, in our model, X, $\mathcal{K}$, U, V and T may not be observed. Let $\pi_k = P(\mathcal{K} = k) > 0$, $k = 0, 2$, and $\pi_0 + \pi_2 = 1$. Denote $(L_1, R_1), \ldots, (L_n, R_n)$ a random sample from the random vector (L, R).

Denote the cdfs of X, (L, R), (U, V), L, R, T, U and V by F, Q, G, Q_L, Q_R, G_T, G_U and G_V, respectively. Define $\tau_o = \sup\{x : F(x) = 0\}$, $\tau_v = \sup\{x : G_V(x) < 1\}$, $\tau_t = \sup\{x : G_T(x) < 1\}$ and $\tau = \inf\{x : F(x) = 1 \text{ or } G_T(x) = 1\}$. An SCE of F is defined to be a solution H_n of the equation

$$(2.2)\quad H_n(x) = \int_{l<x<r} \frac{H_n(x) - H_n(l)}{H_n(r) - H_n(l)} dQ_n(l,r) + \int_{r\le x} dQ_n(l,r) \qquad \text{and} \qquad H_n \in \Theta,$$

where $\Theta = \{h$: h is a nondecreasing function from $[-\infty, \infty]$ to $[0, 1]$, $h(-\infty) = 0$ and $h(\infty) = 1\}$ and Q_n is the empirical version of Q (Li *et al.* (1997)),

Remark 2.1. There are two more ways to define an SCE of F: (1) a subdistribution function which is a solution to the integral equation in (2.2) (Turnbull (1976)); (2) a solution to the integral equation in (2.2) without additional assumption. Since there exists a non-monotone solution to the integral equation in (2.2) (see Yu and Li (1999)), definition (2) is not suggested. It is desirable that an estimator of the cdf is right continuous. However, a solution of (2.2) may not be right continuous. Even if we restrict our attention to all right continuous solutions of (2.2), their limiting functions still cannot be guaranteed to be right continuous based on Helly's selection theorem. Furthermore, by Theorem 2.1, the non-right-continuous solutions of (2.2) are consistent estimators of F. In view of these facts, we define Θ as is.

For each SCE, there exists a unique SCE of F that is a right continuous step function with discrete points only at R_i's and the two SCEs are identical at all L_i's and R_i's.

THEOREM 2.1. *Let H_n be a solution of* (2.2). *Suppose that*

(AS1) (a) $\tau_v \le \tau_t$, *and* (b) *if* $F(\tau_t-) < 1$ *then* $P\{T \text{ or } V = \tau_t\} > 0$.

Then $\lim_{n\to\infty} \sup_{x\ge 0} |H_n(x) - F(x)| = 0$ *a.s. if* $F(\tau) = 1$, *and* $\lim_{n\to\infty} \sup_{x\le\tau} |H_n(x) - F(x)| = 0$ *a.s.*

In clinical follow-ups, a study typically lasts for a certain period of time. Thus it is often true that $F(\tau-) < 1$. In this regard, Gentleman and Geyer ((1994), Theorem 2) claimed a vague convergence result, and Huang ((1996), Theorem 3.1) claimed a uniform strong consistency result for IC data or case 1 data. Schick and Yu (2000) showed that both theorems as stated are false and can be corrected by adding assumption AS1.b to their theorems. A counterexample similar to that in Schick and Yu (2000) can also be constructed to show that the GMLE is not consistent if AS1.b is deleted from our Theorem 2.1.

Remark 2.2. AS1.a basically says that if we are not able to observe exact observations in the region $(\tau_t, +\infty)$, we would not observe IC observations neither. For Theorem 2.1, this assumption may not needed, but it is needed for asymptotic normality.

A point x is called a *support point of* a function f if there exists a sequence of points $x_k \to x$ such that $|f(x_k) - f(x)| > 0$. Denote $\mathcal{S}_f$ the set of all support points of f. It is well known (see Peto (1973)) that a GMLE $\hat{F}_n(t)$ is not uniquely determined for $t \in (L_i, R_j)$ if $L_i < R_j$, $(L_i, R_j) \cap \{L_1, \ldots, L_n, R_1, \ldots, R_n\} = \emptyset$ and $\hat{F}_n(L_i) > \hat{F}_n(R_j)$. For the convenience of our proof of normality, we restrict our attention to the following SCEs:

$$(2.3) \qquad H_n \text{ is right continuous}, \quad H_n(\infty) = 1 \quad \text{and} \quad \mathcal{S}_{H_n} \subset \{R_1, \ldots, R_n\}.$$

Under convention (2.3) the GMLE $\hat{F}$ is uniquely determined. However there are still SCEs that satisfy (2.3) but are not the GMLE. For convenience, we say a constant c is a normal variate with mean c and variance 0.

THEOREM 2.2. *Let H_n satisfy* (2.2) *and* (2.3). *Suppose that* AS1 *holds and* (AS2) *either* (a) $F(\tau) > 0$ *and* $(\mathcal{S}_{G_U} \cup \mathcal{S}_{G_V}) \subset \mathcal{S}_F$

or (b) $\mathcal{S}_{Q_L} \cup \mathcal{S}_{Q_R}$ *is finite and* $F(r) > F(l)$ *if* $l \in \mathcal{S}_{Q_L}, r \in \mathcal{S}_{Q_R}$ *and* $l < r$.

Then for $x \le \tau$, $\sqrt{n}(H_n(x) - F(x))$ *converges in distribution to a normal variate.*

The theorem under assumptions AS1 and AS2.b was established in Yu *et al.* (1998*a*). The rest proof of the theorem follows from Theorem 4.1, in which the SCE is considered as a process. AS1 and AS2 are much weaker than the assumptions made in Petroni and Wolfe (1994), Huang (1999) and Yu *et al.* (1998*a*). If F is strictly monotone on $[0, \infty)$, then AS2 holds. We suspect that AS2 can be replaced by the assumption

$$F(r) > F(l) \quad \text{if} \quad l \in \mathcal{S}_{Q_L}, \ r \in \mathcal{S}_{Q_R} \text{ and } l < r.$$

For technical reasons, we replace it by AS2.a.

Remark 2.3. Under Assumptions AS1 and AS2, H_n is also efficient. The proof is analogous to that of Theorem 3 of Gu and Zhang (1993) and is skipped here.

Remark 2.4. Without AS1.a, the convergence rate of the GMLE on (τ_t, ∞) maybe $n^{1/3}$, as in the situation of case 1 model (see Groeneboom and Wellner (1992)), or maybe $n^{1/2}$ as in the situation considered by Yu *et al.* (1998*b*,1998*c*). If AS2 fails then the GMLE is not asymptotically normally distributed, as in the case when the random vector (L, R) only takes values $(2,2)$, $(6,6)$, $(1,4)$, $(3,7)$, $(-\infty, 3)$ and $(4, \infty)$.

Remark 2.5. In a follow-up study, each patient has $\mathcal{N}$ visits, where $\mathcal{N} \geq 1$ is a random integer (rather than assuming that each patient has exactly 2 visits ($\mathcal{N} \equiv 2$) as in the case 2 model). The inspection times are $Y_1 < \cdots < Y_{\mathcal{N}}$. It is reasonable to assume that X and $(\mathcal{N}, \{Y_i : i \geq 1\})$ are independent. Then, on the event $\{\mathcal{N} = k\}$, modify (U, V) in (2.1) as

$$(2.4) \quad (U,V) = (Y_1, Y_2)\mathbf{1}_{(X \leq Y_1)} + (Y_{k-1}, Y_k)\mathbf{1}_{(X > Y_k)} + \sum_{i=2}^{k} (Y_{i-1}, Y_i)\mathbf{1}_{(Y_{i-1} < X \leq Y_i)},$$

where $Y_0 = 0$. Thus, a more realistic model for MIC data is the model of a mixture of a right censorship model and a modified case 2 model where (U, V) is specified by (2.4), instead of assuming that X and (U, V) are independent. This model includes our model (2.1) (in which $\mathcal{N} = 2$ with probability one) as well as Huang's model (in which $\mathcal{N}$ is a fixed positive integer and $T = \infty$). It is reasonable to assume that $\mathcal{N}$, the number of visits, is bounded. In such a model the Proofs of Theorems 2.1 and 2.2 are similar to the proofs given in Sections 3 and 4. Thus it suffices to study model (2.1).

3. Strong consistency

We shall prove Theorem 2.1. In our proof, we need Theorem 3.1 below to establish consistency and Proposition 3.1 to establish uniform consistency.

THEOREM 3.1. *Suppose that $F \in \Theta$, F is right continuous and H is a solution of*

$$(3.1) \qquad H(x) = \int_{l<x<r} \frac{H(x) - H(l)}{H(r) - H(l)} dQ(l, r) + \int_{r \leq x} dQ(l, r), \qquad H \in \Theta.$$

Then $H(x) = F(x)$ for all $x \leq \tau$ if AS1 *holds; and $H(x) = F(x)$ for all $x < \tau_t$ if*

$$(\text{AS3}) \qquad F(\tau_t) < 1, \tau_v \leq \tau_t \quad \text{and} \quad F = F(\tau_t) \quad \text{on } [x_o, \infty), \text{ where } x_o < \tau_t.$$

In (3.1), if $H(x) = H(r) = H(l)$, then we encounter $\frac{0}{0}$ in the integrand. Hereafter, define $\frac{0}{0} = 1$ and $\frac{0}{0} \cdot 0 = 0$. If F satisfies AS3, it can viewed as the cdf of an extended random variable X which equals ∞ with positive probability.

PROPOSITION 3.1. *Suppose that $\{f_n\}_{n \geq 1}$ is a sequence of monotone functions on an interval $[a, b)$ and $f(x)$ is a bounded monotone and right continuous function on the same interval. If* $\lim_{n\to\infty} f_n(x) = f(x)\ \forall\, x \in [a,b)$ *and* $\lim_{n\to\infty} f_n(x-) = f(x-)\ \forall\, x \in (a, b]$, *then* $\lim_{n\to\infty} \sup_{x\in[a,b)} |f_n(x) - f(x)| = 0$.

We shall present the proof of Theorem 3.1 after we prove Theorem 2.1. We omit the proof of Proposition 3.1 as it is similar to Lemma 3 of Yu and Li (1994).

PROOF OF THEOREM 2.1. Let Ω be the event $\{\lim Q_n(l,r) = Q(l,r)$ uniformly $\forall\, l < r\}$. For each $\omega \in \Omega$, let H_n be a solution of (2.2). We shall prove the theorem in 2 steps.

Step 1. ($\lim_{n\to\infty} H_n(x) = F(x)$ and $\lim_{n\to\infty} H_n(x-) = F(x-) \forall x \le \tau$) Since $\{H_n\}_{n\ge1}$ is bounded and monotone, for each subsequence of natural numbers, by Helly's selection theorem, there exists a further subsequence, say $\{n_k\}$, such that $\lim_{n_k\to\infty} H_{n_k}(x) = H(x)$ and $\lim_{n_k\to\infty} H_{n_k}(x-) = H^*(x)$ pointwisely for some H and $H^* \in \Theta$, respectively. Thus it suffices to show that $H(x) = F(x)$ and $H^*(x) = F(x-)$ for all $x \le \tau$.

Since Q_n converges uniformly to Q, and H_n satisfies (2.2), by the bounded convergence theorem (BCT) H satisfies (3.1) and H^* satisfies a similar equation like (3.1). Theorem 3.1 yield the first desired equation $H(x) = F(x)$ on $(-\infty, \tau]$.

By AS1.a, $r > \tau \Rightarrow r = \infty$ and thus $H(r) = F(r) = 1$ as $H \in \Theta$. Then equation (3.1) and its analog for H^* yield

$$F(x-) = \int_{l<x\le r} \frac{F(x-) - F(l)}{F(r) - F(l)} dQ(l,r) + \int_{r<x} dQ(l,r)$$
$$\text{(as } \textstyle\int_{l<x<r} + \int_{r\le x} = \int_{l<x\le r} + \int_{r<x}\text{)},$$
$$H^*(x) = \int_{l<x\le r} \frac{H^*(x) - F(l)}{F(r) - F(l)} dQ(l,r) + \int_{r<x} dQ(l,r),$$

as $H = F$ on $(-\infty, \tau] \cup \{\infty\}$. The latter two equations yield

$$H^*(x) - F(x-) = (H^*(x) - F(x-))c(x), \qquad \text{where} \tag{3.2}$$
$$c(x) = \int_{l<x\le r} \frac{1}{F(r) - F(l)} dQ(l,r).$$

By AS1,

$$c(x) = \begin{cases} 1 - \pi_0 P(T > x) < 1 & \text{if } \ x < \tau, \\ 1 - P(L = \tau) < 1 & \text{if } \ x = \tau \text{ and } F(\tau-) < 1. \end{cases}$$

It follows from equation (3.2) and $c(x) < 1$ that $H^*(\tau) = F(\tau-)$ if $F(\tau-) < 1$, and $H^*(x) = F(x-)\ \forall\ x < \tau$. In order to show that $H^*(\tau) = F(\tau-)$ if $F(\tau-) = 1$, let $x_k \uparrow \tau$. Note $H_n(x_k) \le H_n(\tau-) \le 1$. It yields $H(x_k) \le H^*(\tau) \le 1$. Now $\lim_{k\to\infty} H(x_k) = \lim_{k\to\infty} F(x_k) = 1$. Thus $H^*(\tau) = 1 = F(\tau-)$.

Step 2. (Conclusion) By step 1 the sequence $\{H_n\}_{n\ge1}$ and F satisfy all the conditions for $\{f_n\}_{n\ge1}$ and f in Proposition 3.1, respectively, where $(a,b) = (-\infty, \tau)$. By Proposition 3.1, $\lim_{n\to\infty} \sup_{x\le\tau} |H_n(x) - F(x)| = 0\ \forall\ \omega \in \Omega$. Since $P\{\Omega\} = 1$ by Glivenko-Cantelli theorem, Theorem 2.1 follows. □

The solution $H(x)$ to (3.1) is unique for $x < \tau_t$ if AS3 holds by Theorem 3.1, but Theorem 2.1 is false if only AS3 holds, as $\int_{l<\tau_t\le r} \frac{1}{F(r)-F(l)} dQ(l,r) = 1$ if $\mathrm{P}(T < \tau_t) = 1$ and $P(V < \tau_t) = 1$. The rest of the section is devoted to prove Theorem 3.1.

The theorem is trivially true if $F(\tau) = 0$, so without loss of generality (WLOG), we can assume $F(\tau) > 0$. The outline of the proof is as follows. We first define a functional $\psi(h)$ for $h \in \Theta$. We then show that $h = F$ uniquely maximizes $\psi(h)$ for $h \in \Theta$ (Lemma 3.1) and that each solution H of (3.1) in Θ is a maximum point of $\psi(\cdot)$. Thus H must equal F. To this end, some notations and lemmas are needed.

Verify that there are at most countably many intervals (y, z) such that (1) $y < z$ and $y \leq \tau$, (2) $F(y) = F(z-)$, and (3) $y, z \in \mathcal{S}_F$. Let $\mu(x) = [F(x) + G_U(x) + G_V(x) + G_T(x)]/4$. For $i \geq 1$, denote D_i the collection of intervals (y, z) satisfying (1), (2), (3) above and $\mu(z-) - \mu(y) \geq 1/i$, then D_i contains finitely many intervals since $\mu(\cdot)$ is a cdf. Thus $\cup_i D_i$, the collection of all such intervals, is countable. Denote D_i^e the set of left endpoints of intervals in D_i.

For $\alpha = 1, 2, \ldots$, denote $B_{\alpha,1}$ the collection of all possible $j2^{-\alpha} \times 100$ percentiles of the distribution μ $(1 \leq j \leq 2^\alpha)$ which are contained in $(-\infty, \tau]$. Note that for each j such that $j2^{-\alpha} \leq \mu(\tau)$ the corresponding percentile is given by $y = \sup\{x : \mu(x) < j2^{-\alpha}\}$. Let $B_\alpha = (B_{\alpha,1} \cup D_\alpha^e) \cup \{\tau\}$ and denote $b_1 < \cdots < b_\beta = \tau$ to be the elements of B_α. Verify that

$$\mu(b_i-) - \mu(b_{i-1}) \leq 2^{-\alpha}, \qquad i = 2, \ldots, \beta. \tag{3.3}$$

Define $b_{1*} = b_1$ and $b_{i*} = \sup\{x : x \leq b_i, F(x) = F(b_{i-1})\}$, $i = 2, \ldots, \beta$. Moreover, if $\tau < \infty$, then denote $b_{\beta+1*} = \tau$ and $b_{\beta+1} = \infty$. For $b_i, b_j \in B_\alpha$, define

$$T_\alpha = \sum_{i=1}^{\beta+1} b_i \mathbf{1}_{(T \in [b_i, b_{i+1}))}, \qquad \lfloor b_{i*}, b_i] = \begin{cases} [b_{i*}, b_i] & \text{if } b_{i*} > b_{i-1}, \\ (b_{i*}, b_i] & \text{if } b_{i*} = b_{i-1}, \end{cases} \text{ and} \tag{3.4}$$

$$(U_\alpha, V_\alpha) = (b_i, b_j) \qquad \text{if} \quad b_i \leq U < b_{i+1}, \ b_{j-1} < V \leq b_j, i \leq j \leq \beta.$$

Then $P\{X \in (b_{i-1}, b_i]\} = P\{X \in \lfloor b_{i*}, b_i]\}$ as $P\{X \in (b_{i-1}, b_{i*})\} = 0$. Define an interval

$$I_\alpha = \begin{cases} (-\infty, b_i] & \text{if } \mathcal{K} = 2 \text{ and } X \leq b_i = U_\alpha, \\ (b_i, b_j] & \text{if } \mathcal{K} = 2, \ X \in (b_i, b_j] \text{ and } (U_\alpha, V_\alpha) = (b_i, b_j), \\ (b_i, \infty] & \text{if } X > b_i \text{ and either } \mathcal{K} = 2 \text{ and } V_\alpha = b_i \text{ or } \mathcal{K} = 0 \text{ and } T_\alpha = b_i, \\ \lfloor b_{i*}, b_i] & \text{if } X \in \lfloor b_{i*}, b_i], \ \mathcal{K} = 0 \text{ and } T_\alpha \geq b_i. \end{cases} \tag{3.5}$$

Then the number of distinct realizations $I_{\alpha,h}$ of the random interval I_α is finite. Denote the joint cdf of the random vector (U_α, V_α) by G_α and the cdf of T_α by G_{T_α}. Let L^α and R^α be the endpoints of the interval I_α, $Q_\alpha(l, r, k)$ the joint cdf of $(L^\alpha, R^\alpha, \mathcal{K})$, and $q_{\alpha,h,k} = P(I_\alpha = I_{\alpha,h}, \mathcal{K} = k)$. Abusing notations, let $Q(l, r, k)$ be the joint cdf of $(L, R, \mathcal{K})$. Thus $Q(l, r)$ can be viewed as the marginal cdf of the random vector (L, R).

For $H \in \Theta$, define μ_H to be the measure induced by H and

$$\psi_\alpha(H) = E[\ln(\mu_H(I_\alpha)/\mu_F(I_\alpha))] \quad \left(= \sum_{h,k} q_{\alpha,h,k} \ln[\mu_H(I_{\alpha,h})/\mu_F(I_{\alpha,h})] \right). \tag{3.6}$$

Here we interpret $\ln 0 = -\infty$, $0\ln 0 = 0$ and $0 \ln\infty = 0$. It is obvious by construction (see (3.3), (3.4) and (3.5)) and by AS1.a that the measures dG_α, dG_{T_α} and dQ_α converge setwisely to dG, dG_T and dQ, respectively. We call $\psi(H)$ a *limit* of $\{\psi_\alpha(H), \alpha \geq 1\}$ if a subsequence of $\{\psi_\alpha(H)\}$ converges to $\psi(H)$, where $\psi(H)$ may be ∞.

The proofs of the following 2 lemmas are given in Appendix A.

LEMMA 3.1. *Suppose that $H \in \Theta$ and either* AS1 *or* AS3 *holds. Let $\psi(H)$ be a limit of $\{\psi_\alpha(H)\}$. Then (1) $\psi(H) = 0$ if and only if $H(x) = F(x)$ for all $x < \tau$, and $H(\tau_t) = F(\tau_t)$ in the case $F(\tau_t-) < 1$ and $P(T$ or $V = \tau_t) > 0$; (2) $\psi(H) \leq 0$.*

A real number $x \in [\tau_o, \tau]$ is called a *left point of increase of* $F \in \Theta$ if $F(x) - F(x - \epsilon) > 0$ for each $\epsilon > 0$. Let $\mathcal{L}_F$ be the set of all left points of increase of F. Denote

$$\gamma_H(a,b) = \frac{P(X \in (a,b], X \leq T, \mathcal{K} = 0)}{H(b) - H(a)}.$$

LEMMA 3.2. *Suppose that H is a solution of* (3.1), AS1 *or* AS3 *holds, and* $b \in \mathcal{L}_F$. *Then*

$$\text{(E.1)} \qquad \int \frac{\mathbf{1}_{(l \leq \tau_t < r)}}{H(r) - H(l)} dQ(l,r) = 1 \quad \textit{if} \quad F(\tau) < 1;$$

$$\text{(E.2)} \qquad \gamma_H(a,b) \leq 1 \quad \textit{for each} \quad a < b;$$

$$\text{(E.3)} \qquad \int_{l<r} \frac{\mathbf{1}_{(b \in (l,r])}}{H(r) - H(l)} dQ(l,r) + \lim_{a \uparrow b} \gamma_H(a,b) - 1 = 0.$$

PROOF OF THEOREM 3.1. Let H be a solution of (3.1). We shall assume that $H(x) \neq F(x)$ for some $x \leq \tau$ but AS1 holds, or $H(x) \neq F(x)$ for some $x < \tau$ but AS3 holds; and show that it leads to a contradiction.

Let $\psi(H)$ be a limit of $\psi_\alpha(H)$. WLOG, assume $\lim_{\alpha \to \infty} \psi_\alpha(H) = \psi(H)$. Since $H \neq F$ for some $t_0 \leq \tau$, $\psi(F) = 0 > \psi(H)$ by Lemma 3.1. Therefore, there exists an integer α_1 such that $\psi_\alpha(F) > \psi_\alpha(H) + \delta$, for all $a \geq \alpha_1$, where $\delta = -\psi(H)/2 > 0$. For each $\alpha \geq \alpha_1$, let $p_i = \mu_F(\lfloor b_{i*}, b_i])$, $i = 1, \dots, \beta$, and $p_{\beta+1} = 1 - F(\tau)$. It is seen that b_i, β, and p_i all are functions of α. Then, for $\alpha \geq \alpha_1$, the above inequality yields

$$\begin{aligned}
\text{(3.7)}\ \delta &\leq -\psi_\alpha(H) + \psi_\alpha(F) \\
&= \lim_{u \downarrow 0} \frac{\frac{1}{1+u}\psi_\alpha(H) + \frac{u}{1+u}\psi_\alpha(F) - \psi_\alpha(H)}{u} \\
&\leq \underline{\lim}_{u \downarrow 0} \frac{\psi_\alpha(\frac{1}{1+u}H + \frac{u}{1+u}F) - \psi_\alpha(H)}{u} \quad \text{(since } -\ln(\cdot) \text{ and hence } -\psi_\alpha(\cdot) \text{ is convex)} \\
&= \underline{\lim}_{u \downarrow 0} \frac{\sum_{j,k} q_{\alpha,j,k} \ln \frac{\mu_{\frac{H+uF}{1+u}}(I_{\alpha,j})}{\mu_F(I_{\alpha,j})} - \sum_{j,k} q_{\alpha,j,k} \ln \frac{\mu_H(I_{\alpha,j})}{\mu_F(I_{\alpha,j})}}{u} \quad \text{(by (3.6))} \\
&= \underline{\lim}_{u \downarrow 0} \frac{\sum_{j,k} q_{\alpha,j,k}[\ln(1 + \frac{u\mu_F(I_{\alpha,j})}{\mu_H(I_{\alpha,j})}) - \ln(1+u)]}{u} \quad \left(\text{as } \frac{H+uF}{1+u} = \frac{1}{1+u}H + \frac{u}{1+u}F\right) \\
&= \sum_{j,k} q_{\alpha,j,k} \frac{\mu_F(I_{\alpha,j})}{\mu_H(I_{\alpha,j})} - 1 \\
&= \int_{l<r \text{ and } k=2, \text{ or } r=\infty} \frac{F(r) - F(l)}{H(r) - H(l)} dQ_\alpha(l,r,2) + \sum_{i=1}^{\beta} p_i \frac{q_{\alpha,j_i,0}}{\mu_H(\lfloor b_{i*}, b_i])} - 1,
\end{aligned}$$

where j_i is such that $I_{\alpha,j_i} = \lfloor b_{i*}, b_i]$, $i = 1, \dots, \beta$. Let $h_1(l,r) = \frac{F(r)-F(l)}{H(r)-H(l)}$ and $h_2(b_{i*}, b_i) = \frac{q_{\alpha,j_i,0}}{\mu_H(\lfloor b_{i*}, b_i])}$. By (E.1), (E.2) and (E.3) in Lemma 3.2, $\int_{l<r} \frac{\mathbf{1}_{(b_j \in (l,r])}}{H(r)-H(l)} dQ(l,r) \leq$

1, thus

$$\begin{aligned}
(3.8)\qquad \infty &> \varliminf_{\alpha\to\infty}\sum_{j=1}^{\beta+1} p_j \int_{l<r} \frac{\mathbf{1}_{(b_j\in(l,r])}}{H(r)-H(l)}dQ(l,r) \qquad \text{(by the BCT)}\\
&\ge \int_{l<r} \frac{\lim_{\alpha\to\infty}\sum_{j=1}^{\beta+1} p_j \mathbf{1}_{(b_j\in(l,r])}}{H(r)-H(l)}dQ(l,r) \qquad \text{(by Fatou's lemma)}\\
&= \int_{l<r} h_1(l,r)dQ(l,r).
\end{aligned}$$

Since h_1 is a nonnegative measurable function, (3.8) implies that it is integrable. Since

$$(3.9)\qquad q_{\alpha,j_i,0} \le P(X\in \lfloor b_{i*}, b_i], X\le T_\alpha, \mathcal{K}=0)$$

by the definition of U_α, V_α, T_α and I_α (see (3.4) and (3.5)), (E.2) and (3.9) imply that $|h_2(b_{i*},b_i)|\le 1$, and thus $\sum_{i=1}^{\beta} p_i h_2(b_{i*},b_i)$ converges by the BCT as $\alpha\to\infty$. Then

$$\begin{aligned}
0<\delta &\le \text{expression (3.7)}\\
&\le \varliminf_{\alpha\to\infty}\left[\int_{l<r \text{ and } k=2, \text{ or } r=\infty} h_1(l,r)dQ_\alpha(l,r,k) + \sum_{i=1}^{\beta} p_i h_2(b_{i*},b_i) - 1\right]\\
&= \int_{l<r} h_1(l,r)dQ(l,r) + \lim_{\alpha\to\infty}\sum_{i=1}^{\beta} p_i h_2(b_{i*},b_i) - 1 \qquad \text{(since } dQ_\alpha\to dQ \text{ setwisely)}\\
&\le \int_{l<r} h_1(l,r)dQ(l,r) + \varlimsup_{\alpha\to\infty}\sum_{i=1}^{\beta} p_i \gamma_H(b_{i*},b_i) - 1 \qquad \text{(by (3.9))}\\
&\le \varliminf_{\alpha\to\infty}\sum_{i=1}^{\beta+1} p_i \int \frac{\mathbf{1}_{(b_i\in(l,r])}}{H(r)-H(l)}dQ(l,r) + \varlimsup_{\alpha\to\infty}\sum_{i=1}^{\beta} p_i\gamma_H(b_{i*},b_i) - 1 \qquad \text{(by (3.8))}\\
&\le \varlimsup_{\alpha\to\infty}\sum_{i=1}^{\beta} p_i\left[\int_{l<r}\frac{\mathbf{1}_{(b_i\in(l,r])}}{H(r)-H(l)}dQ(l,r) + \gamma_H(b_{i*},b_i) - 1\right] \qquad \text{(by (E.1))}\\
&= \int_{[\tau_o,\tau]}\left[\int\frac{\mathbf{1}_{(b\in(l,r])}}{H(r)-H(l)}dQ(l,r) + \lim_{a\uparrow b}\gamma_H(a,b) - 1\right]dF(b) \qquad \text{(by the BCT)}\\
&= 0 \qquad \text{(by (E.3))}.
\end{aligned}$$

Thus we reach a contradiction $0<\delta\le 0$. This concludes the proof of Theorem 3.1. □

4. Asymptotic normality

If $F(\tau)=0$, the GMLE $\hat F(\tau)=0$ w.p.1. If $F(\tau)<1$, $F(t)$ is not identifiable for $t>\tau$. Thus it suffices to estimate F_τ defined by

$$F_\tau(t)=\begin{cases} F(t) & \text{if } t\le\tau\\ F(\tau) & \text{if } \tau<t<\infty,\\ 1 & \text{if } t=\infty\end{cases}$$

and assume that $F(\tau)>0$. Here $F_\tau\in\Theta$ but may not be a cdf and Theorem 3.1 does not require F be a cdf.

There are two equivalent forms for equation (3.1): $H = \mathcal{B}_H(Q)$ and $H = \mathcal{R}_H(F)$, where

$$(4.1)\quad \mathcal{B}_H(Q)(x) = \int_{l<x<r} \frac{H(x)-H(l)}{H(r)-H(l)} dQ(l,r) + \int_{r\le x} dQ(l,r) \qquad \text{(RHS of (3.1))},$$

$$(4.2)\quad \mathcal{R}_H(F)(x) = \int_{l\le x<r} \left\{ \frac{H(x)-H(l)}{H(r)-H(l)} [F(r)-F(l)] - [F(x)-F(l)] \right\} \cdot dG^*(l,r) + F(x),$$

$$(4.3)\quad dG^*(l,r) = \begin{cases} \pi_2 dP(V\le l) + \pi_0 dG_T(l) & \text{if } r=\infty, \\ \pi_2 dP(U\le r) & \text{if } l=-\infty, \\ \pi_2 dG(l,r) & \text{if } -\infty<l<r<\infty, \end{cases}$$

$$dQ(l,r) = [F(r)-F(l)]dG^*(l,r) = [F_\tau(r)-F_\tau(l)]dG^*(l,r) \text{ if } l<r.$$

LEMMA 4.1. *$\mathcal{B}_{H_n}(Q_n - Q) = \mathcal{R}_{H_n}(H_n - F_\tau)$ for each SCE H_n which satisfies* (2.3).

The proof of the lemma is in Appendix B.

Let $\mathcal{D}$ be the collection of all real-valued functions h defined on $[-\infty,\infty]$ that are right-continuous, have left limits at each point and satisfy that

$$(4.4)\qquad \forall\, a<b\le\infty, \quad F_\tau(a-) = F_\tau(b) \Rightarrow h(a-) = h(b).$$

Define $\mathcal{D}_0 = \{h\in\mathcal{D} : F(x)=0 \Rightarrow h(x)=0; F_\tau(x-)=1 \Rightarrow h(x-)=0\}$. Verify that $(\mathcal{D}, \|\cdot\|)$ and $(\mathcal{D}_0, \|\cdot\|)$ are both Banach spaces, where $\|\cdot\|$ is the supremum norm. Let $(\mathcal{D}_2, \|\cdot\|)$ be a Banach space of real-valued functions defined on $[-\infty,\infty]^2$ such that the Banach space contains all bivariate cdfs. Note that AS1–AS3 are basically assumptions on (F, G, G_T). We say (H, G, G_T) satisfies AS1 etc., if $H\in\Theta$ and H replaces the role of F in AS1 etc. Let $\Theta_o = \{H\in\Theta\cap\mathcal{D} :\ \mathcal{S}_H \subset \mathcal{S}_F,\ (H,G,G_T) \text{ satisfies AS1 or AS3}\}$. For each $H\in\Theta_o$, $\mathcal{R}_H(\cdot)$ and $\mathcal{B}_H(\cdot)$ are linear operators on $\mathcal{D}$ and $\mathcal{D}_2$, respectively.

THEOREM 4.1. *Suppose that* AS1, AS2 *and* (2.3) *hold. Then $\mathcal{R}_{F_\tau}^{-1}$ exists as a bounded operator from $\mathcal{D}$ to $\mathcal{D}$ and the SCE satisfies*

$$(4.5)\qquad \sqrt{n}(H_n - F_\tau) \xrightarrow{D} \mathcal{R}_{F_\tau}^{-1}\mathcal{B}_{F_\tau}(W) \text{ in } \mathcal{D},$$

where W is the Gaussian process specified by $\sqrt{n}(Q_n(l,r) - Q(l,r)) \xrightarrow{D} W$.

We first state 3 more lemmas, with their proofs relegated to Appendix B.

For a $\tilde F\in\Theta_o$, let C_k be the collection of all the distinct points among $c_{k,i}$'s, where $c_{k,i} = \inf\{x :\ \tilde F(x) \ge i/2^k\}$, $i=0,\ldots,2^k$, $k\ge 1$. Let F_k be a step function in Θ_o such that $F_k(c) = \tilde F(c)$ for each $c\in C_k$ and its discontinuity points belong to C_k. Denote $\mathcal{D}_k$ ($\mathcal{D}_{k0}$) the subclass of $\mathcal{D}$ ($\mathcal{D}_0$) such that each member is a step function with the collection of discontinuity points being a subset of $\mathcal{S}_{F_k}$. Obviously, D_k, C_k and F_k depend on $\tilde F$.

LEMMA 4.2. *If $\tilde F\in\Theta_o$, then the linear operator $\mathcal{R}_{F_k}^{-1}$ exists as a map from $\mathcal{D}_k$ onto $\mathcal{D}_k$.*

LEMMA 4.3. *Assume that* AS1, AS2 *and* (2.3) *hold. For each $\omega\in\Omega$, $H_n\in\Theta_o$.*

LEMMA 4.4. *If $\tilde{F} \in \Theta_o$, then $\|\mathcal{R}_{F_k}^{-1}(\cdot)\| \leq 1$ for all possible k.*

PROOF. We give the proof of asymptotic normality in 4 steps.

Step 1. (Existence of $\mathcal{R}_{\tilde{F}}^{-1}$, $\tilde{F} \in \Theta_o$, as a linear operator from $\mathcal{D}$ to $\mathcal{D}$) For each $g \in \mathcal{D}$ and $k \geq 1$, let $g_k \in \mathcal{D}_k$ be such that $g_k(x) = g(x)$ if $x \in C_k$. Then $\|g_k - g\| \to 0$, since $\mathcal{S}_{g_k}$ and $\mathcal{S}_g \subset \mathcal{S}_{\tilde{F}}$ and $C = \cup_k C_k$ is dense in $\mathcal{S}_{\tilde{F}}$. By Lemma 4.2, $\mathcal{R}_{F_k}^{-1}$ exists, so there exists a unique $h_k \in \mathcal{D}_k$ such that $g_k = \mathcal{R}_{F_k}(h_k)$. $\forall\ K > k$ and $\forall\ h \in \mathcal{D}$, $\mathcal{D}_k \subset \mathcal{D}_K$, $\|F_k - F_K\| \leq 1/2^k$ and $\mathcal{R}_{F_k}(h) - \mathcal{R}_{F_K}(h)$ converges to 0 as $k \to \infty$ by the BCT. $\|\mathcal{R}_{F_k}^{-1}(\cdot)\| \leq 1$ by Lemma 4.4, thus $\lim_{k\to\infty}[\mathcal{R}_{F_k}^{-1}(h) - \mathcal{R}_{F_K}^{-1}(h)] = 0\ \forall\ h \in \mathcal{D}_k$ and $\forall\ k \geq 1$. Furthermore,

$$\begin{aligned}\|h_k - h_K\| &\leq \|\mathcal{R}_{F_k}^{-1}(g_k) - \mathcal{R}_{F_K}^{-1}(g_k)\| + \|\mathcal{R}_{F_K}^{-1}(g_k) - \mathcal{R}_{F_K}^{-1}(g_K)\| \\ &\leq \|\mathcal{R}_{F_k}^{-1}(g_k) - \mathcal{R}_{F_K}^{-1}(g_k)\| + \|\mathcal{R}_{F_K}^{-1}\| \cdot \|g_k - g_K\| \to 0 \quad \text{as} \quad k \to \infty,\end{aligned}$$

by the assumption $\|g_k - g\| \to 0$, Lemmas 4.2 and 4.4, and the BCT. That is, $\|h_k\|$ is a Cauchy sequence. Since $\mathcal{D}$ is a Banach space, there is a function $h_o \in \mathcal{D}$ such that $\|h_k - h_o\| \to 0$. By the BCT, $g = \lim_{k\to\infty} \mathcal{R}_{F_k}(h_k) = \mathcal{R}_{\tilde{F}}(h_o)$. Define $h_o = \mathcal{R}_{\tilde{F}}^{-1}(g)$.

Step 2. (Strong continuity of $\{\mathcal{R}_H^{-1} : H \in \Theta_o\}$) Let $g'_m \in \mathcal{D}$ and $H_m \in \Theta_o$ be such that $\|g'_m - g\| \to 0$ and $\|H_m - F_\tau\| \to 0$ as $m \to \infty$. Then

$$\begin{aligned}\|\mathcal{R}_{H_m}^{-1}(g'_m) - \mathcal{R}_{F_\tau}^{-1}(g)\| &\leq \|\mathcal{R}_{H_m}^{-1}(g'_m) - \mathcal{R}_{F_\tau}^{-1}(g'_m)\| + \|\mathcal{R}_{F_\tau}^{-1}(g'_m) - \mathcal{R}_{F_\tau}^{-1}(g)\| \\ &\leq \|\mathcal{R}_{H_m}^{-1} - \mathcal{R}_{F_\tau}^{-1}\| \cdot \|g'_m\| + \|\mathcal{R}_{F_\tau}^{-1}\| \cdot \|g'_m - g\| \to 0 \quad \text{as } m \to \infty.\end{aligned}$$

Step 3. (Strong continuity of $\{\mathcal{B}_H : H \in \Theta_o\}$) Let h be a simple function in $\mathcal{D}_2$. It follows from (4.1) and the BCT that $\mathcal{B}_H(h) \to \mathcal{B}_{F_\tau}(h)$ in $\mathcal{D}$ as $H \to F_\tau$. Since $\|\mathcal{B}_H\| \leq 4$ $\forall\ H \in \Theta_o$ and the collection of simple functions is dense in $\mathcal{D}_2$, we have strong continuity.

Step 4. (Conclusion) By Lemma 4.3, $H_n \in \Theta_o$. Thus $\mathcal{R}_{H_n}^{-1}$ exists by Step 1. It follows that $\sqrt{n}(H_n - F_\tau) = \mathcal{R}_{H_n}^{-1}\mathcal{B}_{H_n}(\sqrt{n}[Q_n - Q])$ by Lemma 4.1. By Theorem 2.1 $\lim_{n\to\infty} |H_n(x) - F_\tau(x)| = 0$ a.s. By Steps 2 and 3, $\{\mathcal{F}_H = \mathcal{R}_H^{-1}\mathcal{B}_H : H \in \Theta_o\}$ is strongly continuous. As a consequence of the above 4 statements, and the Banach-Steinhaus theorem, $\sup\{\|\mathcal{F}_{H_n}(h) - \mathcal{F}_{F_\tau}(h)\| : h \in A(\epsilon)\} \to 0$ a.s. as $n \to \infty$ and then $\epsilon \to 0+$ for all compact set $A \subset \mathcal{D}_2$, where $A(\epsilon) = \{h \in \mathcal{D}_2 : \|h - h'\| < \epsilon$ for some $h' \in A\}$. By the central limit theorem, $W_n = \sqrt{n}[Q_n - Q] \xrightarrow{D} W$ in $\mathcal{D}_2$, $\{W_n\}$ is uniformly tight (Pollard (1984), p. 81). As a consequence, $\|\sqrt{n}(H_n - F_\tau) - \mathcal{F}_{F_\tau}(W_n)\| = \|(\mathcal{F}_{H_n} - \mathcal{F}_{F_\tau})(W_n)\| = o_p(1)$, which implies (4.5) by the continuous mapping theorem (Pollard (1984), p. 70). □

Remark 4.1. Our proof of the normality (not the consistency) relies on the form (2.3). It can be shown that Theorem 4.1 are actually true without (2.3), and Theorem 2.1 (not Theorem 4.1) are true without AS1.a. For the sake of simplicity, we skip the details.

Theorem 2.2 is a consequence of Theorem 4.1 and Yu *et al.* (1998*a*), Theorem 2.

Acknowledgements

The authors thank two referees for helpful comments.

Appendix A

We shall prove Lemmas 3.1 and 3.2. A lemma is needed to prove Lemma 3.1.

LEMMA A.1. *Assume that* AS1 *or* AS3 *holds. Let* $\psi(H)$ *be a limit of* $\{\psi_\alpha(H)\}$, $H \in \Theta$. *Then* $\psi(H) = 0$ *if and only if* (1) $H(t) = F(t)$ *and* $H(t-) = F(t-)$ $\forall\, t \in \mathcal{S}_F \cap \cup_\alpha B_\alpha \cap (-\infty, \tau)$, (2) $H(\tau-) = F(\tau-)$ *if* $F(\tau-) < 1$ *and* (3) $H(\tau) = F(\tau)$ *if* $F(\tau-) < 1$ *and* AS1 *holds. Moreover,* $\psi(H) \le 0$.

PROOF. ($\Rightarrow$) Verify that $\psi_\alpha(F) = 0$ for all α by AS1.a, and thus $\lim_{\alpha\to\infty} \psi_\alpha(F) = 0$. Then conditions (1)–(3) above imply that $\psi_\alpha(H) = \psi_\alpha(F) = 0$ for all $\alpha \ge 1$. Thus $\psi(H) = 0$.

($\Leftarrow$) We first show that $\psi(H) = 0$ implies condition (1). It suffices to show that $\psi(H) < 0$ if for some $t_0 \in \mathcal{S}_F \cap \cup_\alpha B_\alpha \cap (-\infty, \tau)$ either (1.a) $H(t_0) \ne F(t_0)$ or (1.b) $H(t_0-) \ne F(t_0-)$. Condition (1.a) implies that for each sufficient large α, there is a point $b_h \in \mathcal{S}_F \cap B_\alpha$ such that $b_h = t_0$. Verify that

$$\begin{aligned}
\text{(A.1)}\quad \psi_\alpha(H) &= E\{E(\ln(\mu_H(I_\alpha)/\mu_F(I_\alpha)) \mid U_\alpha, V_\alpha, T_\alpha, \mathcal{K})\}\\
&= \pi_2 \iint f_{\alpha,2}(z,y)dG_\alpha(z,y) + \pi_0 \int f_{\alpha,0}(t)dG_{T,\alpha}(t), \text{ where}\\
f_{\alpha,2}(z,y) &= F(z)\ln\frac{H(z)}{F(z)} + [F(y)-F(z)]\ln\frac{H(y)-H(z)}{F(y)-F(z)} + [1-F(y)]\ln\frac{1-H(y)}{1-F(y)},\\
f_{\alpha,0}(b_j) &= F(b_1)\ln\frac{H(b_1)}{F(b_1)} + \sum_{k>1}^{j} \mu_F(\lfloor b_{k*}, b_k]) \ln\frac{\mu_H(\lfloor b_{k*}, b_k])}{\mu_F(\lfloor b_{k*}, b_k])}\\
&\quad + [1-F(b_j)]\ln\frac{1-H(b_j)}{1-F(b_j)},
\end{aligned}$$

and b_k, t_0 and $b_j \in B_\alpha$. Note t_0 is fixed but the index h of $b_h = t_0$ depends on α. Define

$$\text{(A.2)}\qquad g(k,t) = \begin{cases} F(t_0)\ln\frac{H(t_0)}{F(t_0)} + [1-F(t_0)]\ln\frac{1-H(t_0)}{1-F(t_0)} & \text{if } t_0 \le t,\\ 0 & \text{otherwise.}\end{cases}$$

Then

$$\begin{aligned}
0 &= \ln\left[F(t_0)\frac{H(t_0)}{F(t_0)} + [1-F(t_0)]\frac{1-H(t_0)}{1-F(t_0)}\right] > F(t_0)\ln\frac{H(t_0)}{F(t_0)} + [1-F(t_0)]\ln\frac{1-H(t_0)}{1-F(t_0)}\\
&= g(0,t), \qquad \text{for} \quad t \ge t_0,
\end{aligned}$$

as $-\ln(\cdot)$ is strictly convex and $F(t_0) \ne H(t_0)$. Moreover, $P\{T \text{ or } V \ge t_0\} > 0$ as $\pi_0 > 0$ and $t_0 \in (-\infty, \tau)$. It follows from the above two statements that

$$\text{(A.3)}\qquad P\{0 > g(\mathcal{K}, T)\} > 0.$$

It is obvious that (1.a.1) $g(2,t) \ge f_{\alpha,2}(u,v)$ for each (u,v,t) and (1.a.2) $0 = g(0,t) \ge f_{\alpha,0}(t)$ for $t < t_0$. We shall show that, (1.a.3) $g(0,t) > f_{\alpha,0}(t)$, for $t = b_k \ge t_0$, where $b_k \in B_\alpha$ and α is sufficiently large. Let $\int g dG^w_\alpha = \pi_2 \iint g(2,t)dG_\alpha(u,v) + \pi_0 \int g(0,t)dG_{T_\alpha}(t)$, and define $\int g dG^w$ in an obvious way. Then (1.a.1), (1.a.2) and (1.a.3) imply that $\int g dG^w_\alpha \ge \psi_\alpha(H)$. Since dG^w_α converges to dG^w setwisely by observing that dG_α (dG_{T_α})

converges to dG (dG_T) setwisely and $g(k,t)$ is a binary function in (u,v,t,k), the desired result follows from (A.3) and $0 > \int gdG^w = \overline{\lim}_{\alpha\to\infty} \int gdG_\alpha^w \geq \overline{\lim}_{\alpha\to\infty}\psi_\alpha(H) \geq \psi(H)$.

We now establish (1.a.3). Let $t_o = b_h \leq b_j = t$ for some integer α_o. It is easy to see by our construction that $B_{\alpha_1} \subset B_{\alpha_2}$ if $\alpha_1 < \alpha_2$ and hence $t_0, t \in B_\alpha$ for all $\alpha \geq \alpha_o$. For each $z = b_i \in B_{\alpha_1}$ such that $z < t_0$, verify

$$\text{(i)}\qquad g(0,t) = F(t_0)\ln\left\{\frac{H(z)}{F(z)}\frac{F(z)}{F(t_0)} + \frac{(H(t_0)-H(z))}{(F(t_0)-F(z))}\frac{F(t_0)-F(z)}{F(t_0)}\right\}$$

$$+[1-F(t_0)]\ln\left\{\frac{(H(t)-H(t_0))}{(F(t)-F(t_0))}\frac{F(t)-F(t_0)}{1-F(t_0)} + \frac{1-H(t)}{1-F(t)}\frac{1-F(t)}{1-F(t_0)}\right\},$$

$$\text{(ii)}\qquad \ln\frac{H(b)-H(a)}{F(b)-F(a)} \geq \frac{F(x)-F(a)}{F(b)-F(a)}\ln\frac{H(x)-H(a)}{F(x)-F(a)} + \frac{F(b)-F(x)}{F(b)-F(a)}\ln\frac{H(b)-H(x)}{F(b)-F(x)}$$

$$\text{for all}\quad x \in (a,b).$$

In view of (i) and (ii), (1.a.3) follows by an induction argument.

Now consider condition (1.b). If t_0 is a point satisfying condition (1.b), then either (1.b.1) $t_0 \in \mathcal{S}_F \cap (\cup_\alpha B_\alpha) \cap (\cup_\alpha B_\alpha^*)$, where $B_\alpha^* = \{x : x = b_{i*} > b_{i-1}, b_i, b_{i-1} \in B_\alpha\}$, or (1.b.2) $t_0 \in \mathcal{S}_F \cap (\cup_\alpha B_\alpha) \cap (\cup_\alpha B_\alpha^*)^c$, where A^c is the complement of the set A.

First assume (1.b.1). For each sufficiently large α, there exists a $b_{h*} = t_0 \in B_\alpha^*$. Thus replacing t_0 by t_0- in the proof for situation (1.a) yields $\psi(H) < 0$.

On the other hand, in view of (3.3), (1.b.2) implies that $F(t_0-) > F(t)$ for each $t < t_0$ and hence there exists a sequence of points $x_i \in \mathcal{S}_F \cap \cup_\alpha (B_\alpha \cup B_\alpha^*)$ such that $x_i \uparrow t_0$ with either $H(x_i) \neq F(x_i)$ (if $x_i = b_{j*} = b_{j-1}$) or $H(x_i-) \neq F(x_i-)$ (if $x_i = b_{j*} > b_{j-1}$). In either case, it reduces to situation (1.a) or (1.b.1). Thus, we have $\psi(H) < 0$. This concludes the proof for condition (1.b).

The proofs for conditions (3) and (2) are similar to that for conditions (1.a) and (1.b), respectively, except in the proof for condition (3) replacing in the above proof the statement $P\{t_0 \leq T\} > 0$ by $P\{T \text{ or } V = \tau_t\} > 0$ (as AS1 holds). We omit the details.

Verify that we actually show that either $\psi(H) = 0$ or $\psi(H) < 0$. Thus $\psi(H) \leq 0$. □

Proof of Lemma 3.1. Statement (2) follows from the last statement in Lemma A.1. To prove statement (1) , it suffices to show that conditions (1), (2) and (3) in Lemma A.1 imply $H(x) = F(x)\ \forall\ x \leq \tau$, i.e. the sufficient and necessary condition in Lemma 3.1.

If x is a discontinuity point of F and $x \leq \tau$, then there exists an integer N such that $F(x) - F(x-) > 2^{-\alpha}$ for all $\alpha \geq N$. This implies that x is a certain $j2^{-N} \times 100$ percentile of μ and thus $x \in B_\alpha \cap \mathcal{S}_F$. It follows that $\mathcal{S}_F \cap \cup_\alpha B_\alpha$ contains all discontinuity points of F which belong to $(-\infty, \tau]$. Thus conditions (1), (2) and (3) of Lemma A.1 imply $H(x) = F(x)$.

Suppose now x is a continuity point of F. Let $u_x = \inf\{y : F(y) = F(x)\}$ and $v_x = \sup\{y : F(y) = F(x)\}$. If both u_x and v_x belong to $\mathcal{S}_F \cap \cup_\alpha B_\alpha$, we are done, as $F(x) = F(u_x) = H(u) \leq H(x) \leq H(v_x-) = F(v_x-) = F(x)$ by conditions (1), (2) and (3) in Lemma A.1.

If neither u_x nor v_x belongs to $\mathcal{S}_F \cap \cup_\alpha B_\alpha$, then from the above discussion both u_x and v_x are continuous support points of F satisfying $F(u_x) = F(v_x) = F(x)$, and there exist two sequences of support points of F, say $\{x_i\}_{i\geq 1}$ and $\{y_j\}_{j\geq 1}$, which are contained in $\mathcal{S}_F \cap \cup_\alpha B_\alpha$ such that $x_i \uparrow u_x$ and $y_j \downarrow v_x$. Consequently, $F(x_i) = H(x_i) \leq H(x) \leq H(y_j) = F(y_j)$ by conditions (1), (2) and (3) in Lemma A.1. This yields $H(x) = F(x)$ as $F(x_i) \to F(u_x)$ and $F(y_j) \to F(v_x)$.

For simplicity, we skip the proof for the case that only u_x or v_x belongs to $\mathcal{S}_F \cap (\cup_\alpha B_\alpha)$. This concludes the proof of the lemma. □

A lemma is needed for proving Lemma 3.2.

LEMMA A.2. *Suppose that H is a solution of* (3.1) *and A is an interval* $(a,b] \subset (-\infty,\tau]$. *Then* $\mu_F(A) > 0 \Rightarrow \mu_H(A) > 0$.

PROOF. Equation (3.1) is equivalent to

$$\mu_H((a,b]) = \int_{l<r} \frac{\mu_H((a,b]\cap(l,r])}{H(r)-H(l)}dQ(l,r) + P(X\in(a,b], X\leq T, \mathcal{K}=0). \tag{A.4}$$

If H is a solution to (3.1), then for each interval $A=(a,b]\subset(-\infty,\tau]$ such that $\mu_F(A) > 0$, we have $\mu_H(A) \geq P(X\in A, X\leq T, \mathcal{K}=0) > 0$ by the assumption $\pi_0 > 0$ and $b\leq\tau$. This concludes the proof of the lemma. □

PROOF OF LEMMA 3.2. Assume that H is a solution of (3.1) and $b \in \mathcal{L}_F$. By Lemma A.2, $\mu_H((a,b]) > 0 \forall a < b$. Dividing both sides of equation (A.4) by $\mu_H((a,b])$ yields

$$1 = \int_{l<r} \frac{\mu_H((a,b]\cap(l,r])}{[H(b)-H(a)][H(r)-H(l)]}dQ(l,r) + \frac{P(X\in(a,b], X\leq T,\mathcal{K}=0)}{H(b)-H(a)}, \quad a<b. \tag{A.5}$$

For each $a<b$, (A.5) yields (E.2) as the two summands in (A.5) are nonnegative. Denote

$$\partial H(a,b,l,r) = \begin{cases} 0 & \text{if } \mu_H((a,b]\cap(l,r]) = 0, \\ \frac{\mu_H((a,b]\cap(l,r])}{[H(b)-H(a)][H(r)-H(l)]} & \text{otherwise.}\end{cases}$$

For each pair (l,r) such that $l<b\leq r$, we have $H(r)-H(l)>0$ by Lemma A.2. Moreover,

$$\partial H(a,b,l,r) \uparrow \frac{\mathbf{1}_{(b\in(l,r])}}{H(r)-H(l)} \text{ as } a\uparrow b \text{ if } b\in(l,r), \text{ and } \partial H(a,b,l,r)\downarrow 0 \text{ as } a\uparrow b \text{ if } b>r.$$

Thus by the monotone convergence theorem, as $a\uparrow b$, we have

$$\int \partial H(a,b,l,r)dQ = \int_{b\in(l,r]} \partial H(a,b,l,r)dQ + \int_{b>r} \partial H(a,b,l,r)dQ \to \int \frac{\mathbf{1}_{(b\in(l,r])}}{H(r)-H(l)}dQ.$$

The desired equation (E.3) follows from (A.5), (E.2) and the above equation.

Assume now $0 < F(\tau) < 1$ and AS1 holds. By AS1.a, $l\leq\tau_t<r \Rightarrow r=\infty$, thus

$$\begin{aligned}\mu_H((\tau,\infty]) &= \int_{l\leq\tau<r} \frac{H(r)-H(\tau)}{H(r)-H(l)}dQ(l,r) \quad \text{(by AS1 and (A.4))} \\ &\geq \int_{l=\tau<r} dQ(l,r) = (1-F(\tau))P(L=\tau) > 0 \quad \text{(by AS1.b).}\end{aligned} \tag{A.6}$$

Dividing both sides of equation (A.6) by $\mu_H((\tau_t,\infty])$ yields (E.1) under AS1.

On the other hand, assume that $0 < F(\tau) < 1$ and AS3 holds. Note that even if we encounter $\frac{0}{0}$, $\frac{\mu_H((x_o,\infty]\cap(l,r])}{H(r)-H(l)}\mathbf{1}_{[x_o\le l<r)} = 1$ by convention. By AS3, $P(x < T < \tau_t) > 0$ for each $x < \tau_t$. (A.4) yields

$$\text{(A.7)} \qquad \begin{aligned}\mu_H((x,\infty]) &= \int_{l<r}\frac{\mu_H((x,\infty]\cap(l,r])}{H(r)-H(l)}dQ(l,r)\\ &\ge (1-F(\tau_t))P(T\in(x,\tau_t])>0, \qquad \forall x\in[x_o,\tau_t).\end{aligned}$$

Then dividing both sides of (A.7) by $\mu_H((x,\infty])$ and taking limits yield

$$1 = \lim_{x\uparrow\tau_t}\int_{l<r}\frac{\mu_H((x,\infty]\cap(l,r])}{\mu_H((x,\infty))(H(r)-H(l))}dQ(l,r) = \int\frac{\mathbf{1}_{(l\le\tau_t<r)}}{H(r)-H(l)}dQ(l,r)$$

(as $\tau_v \le \tau_t$ and thus $l \le \tau_t < r \Rightarrow r = \infty$), which is (E.1). □

Appendix B

In this appendix, we prove lemmas in Section 4.

PROOF OF LEMMA 4.1. Theorem 3.1, (4.1), (4.2) and (2.3) yield $\mathcal{B}_{H_n}(Q_n)(x) = H_n(x)$ and $\mathcal{R}_{F_\tau}(F_\tau)(x) = F_\tau(x)\forall x$. Furthermore,

$$\begin{aligned}
&\int_{l\le x<r}\frac{H_n(x)-H_n(l)}{H_n(r)-H_n(l)}\{[H_n(r)-F_\tau(r)]-[H_n(l)-F_\tau(l)]\}dG^*(l,r)\\
&= \int_{l\le x<r}[H_n(x)-H_n(l)]dG^*(l,r)\\
&\quad -\int_{l\le x<r}\frac{H_n(x)-H_n(l)}{H_n(r)-H_n(l)}[F_\tau(r)-F_\tau(l)]dG^*(l,r)\\
&= \int_{l\le x<r}[H_n(x)-H_n(l)]dG^*(l,r) - \int_{l\le x<r}\frac{H_n(x)-H_n(l)}{H_n(r)-H_n(l)}dQ(l,r) \quad \text{(by (4.3))}\\
&= \int_{l\le x<r}[H_n(x)-H_n(l)]dG^*(l,r) - \mathcal{B}_{H_n}(Q)(x) + P\{R\le x\} \quad \text{(by (4.1))}\\
&= \int_{l\le x<r}[H_n(x)-H_n(l)]dG^*(l,r)\\
&\quad -\mathcal{B}_{H_n}(Q)(x) + [\mathcal{B}_{H_n}(Q_n)(x)-H_n(x)] \quad (\text{since } \mathcal{B}_{H_n}(Q_n)=H_n)\\
&\quad +\left[F_\tau(x)-\int_{l\le x<r}F_\tau(x)-F_\tau(l)dG^*(l,r)\right] \quad (=P\{R\le x\} \text{ as } F_\tau=\mathcal{R}_{F_\tau}(F_\tau))\\
&= \mathcal{B}_{H_n}(Q_n-Q)(x) - [H_n(x)-F_\tau(x)]\\
&\quad +\int_{l\le x<r}\{[H_n(x)-F_\tau(x)]-[H_n(l)-F_\tau(l)]\}dG^*.
\end{aligned}$$

Translating certain terms in the first and last expressions of the above equations yields

$$\begin{aligned}
&\mathcal{B}_{H_n}(Q_n-Q)(x)\\
&= \int_{l\le x<r}\frac{H_n(x)-H_n(l)}{H_n(r)-H_n(l)}\{[H_n(r)-F_\tau(r)]-[H_n(l)-F_\tau(l)]\}dG^*(l,r)
\end{aligned}$$

$$+[H_n(x) - F_\tau(x)] - \int_{l\le x<r} \{[H_n(x) - F_\tau(x)] - [H_n(l) - F_\tau(l)]\} dG^*(l,r)$$

$$= \int_{l\le x<r} \left(\frac{H_n(x) - H_n(l)}{H_n(r) - H_n(l)} \{[H_n(r) - H_n(l)] - [F_\tau(r) - F_\tau(l)]\} \right.$$

$$\left. - \{[H_n(x) - H_n(l)] - [F_\tau(x) - F_\tau(l)]\} \right) dG^*(l,r) + [H_n(x) - F_\tau(x)]$$

$$= \mathcal{R}_{H_n}(H_n - F_\tau)(x) \quad \text{(by (4.2))}. \square$$

LEMMA B.1. *If $\tilde{F} \in \Theta_o$ and $\mathcal{R}_{\tilde{F}}(h) = 0$, where $h \in \mathcal{D}$, then $h \in \mathcal{D}_0$.*

PROOF. For each $h \in \mathcal{D}$, by (4.2),

$$\text{(B.1)} \quad \mathcal{R}_{\tilde{F}}(h)(x) = \int_{l\le x<r} \left\{ \frac{\tilde{F}(x) - \tilde{F}(l)}{\tilde{F}(r) - \tilde{F}(l)} [h(r) - h(l)] - [h(x) - h(l)] \right\} dG^*(l,r) + h(x).$$

If $F(x) = 0$, then $h = h(x)$ on $(-\infty, x]$ by (4.4). Thus

$$0 = \mathcal{R}_{\tilde{F}}(h)(x) = -\int_{l\le x<r} [h(x) - h(l)] dG^*(l,r) + h(x) = h(x).$$

Moreover, if $F_\tau(x-) = 1$, then $h = h(x-)$ on $[x, \infty]$ by (4.4), and $0 = \mathcal{R}_{\tilde{F}}(h)(x-) = \int_{l<x\le r} [h(r) - h(x-)] dG^*(l,r) + h(x-) = h(x-)$. Thus $h \in \mathcal{D}_0$. $\square$

PROOF OF LEMMA 4.2. Note $\tilde{F} \in \Theta_o$. Since $\mathcal{R}_{F_k}$ is a linear operator on the finite dimensional linear space $\mathcal{D}_k$, it suffices to show (1) $\mathcal{R}_{F_k}$ is 1-1 and (2) $\mathcal{R}_{F_k}(\mathcal{D}_k) \subset \mathcal{D}_k$.

Step (1). Suppose $\mathcal{R}_{F_k}(h) = 0$, where $h \in \mathcal{D}_k$. We shall show that $h = 0$. Denote $\alpha = \sum_{c\in C_k} |h(c) - h(c-)|$ and $m = \min\{m_c : m_c = F_k(c) - F_k(c-) > 0, c \in C_k\}$. Note that $m > 0$ and α is finite as C_k contains finitely many points. Choose $\gamma > 0$ such that $\gamma\alpha < m$. Let $H = F_k + \gamma h$. Since $\mathcal{R}_{F_k}(h) = 0$ and $\tilde{F} \in \Theta_o$, $h \in \mathcal{D}_{k0}$ by Lemma B.1. As consequences, (1) $H(\tau_o-) = F_k(\tau_o-) + 0 = 0$ and $H(\infty) = F_k(\infty) + 0 = 1$; (2) $H(c) > H(c-)$ for all $c \in C_k$ [as $H(c) - H(c-) > m - \gamma(h(c) - h(c-)) \ge m - \gamma\alpha > 0$]; (3) $H(x) \in \mathcal{D}_k$.

It follows from statements (1), (2) and (3) that $H = F_k + \gamma h \in \Theta_o \cap \mathcal{D}_k$. Then $\mathcal{R}_{F_k}(H)(x) = \mathcal{R}_{F_k}(F_k)(x) + \mathcal{R}_{F_k}(\gamma h)(x) = F_k(x) + 0$ for each x. That is $F_k = \mathcal{R}_{F_k}(H)$. Note that (H, G, G_T) satisfies AS1 or AS3 as $H \in \Theta_o$. Thus $F_k = H = F_k + \gamma h$ by Theorem 3.1, which implies $h = 0$ as $\gamma > 0$. As a consequence, $\mathcal{R}_{F_k}(\cdot)$ is 1-1.

Step (2). It suffices to show that $\mathcal{A} = \mathcal{R}_{F_k}(h)(b) - \mathcal{R}_{F_k}(h)(a) = 0$ if $h \in \mathcal{D}_k$ and $\mu_{F_k}((a,b]) = 0$. Define $\mu_h((a,b]) = h(b) - h(a)$. By definition of $\mathcal{D}_k$, $\mu_h((a,b]) = 0$. Then

$$\mathcal{A} = \int_{l<r} \left[\frac{\mu_{F_k}((l,r] \cap (a,b])}{\mu_{F_k}((l,r])} \mu_h((l,r]) - \mu_h((l,r] \cap (a,b]) \right] dG^*(l,r) + \mu_h((a,b]) = 0. \quad \square$$

PROOF OF LEMMA 4.3. We fix $\omega \in \Omega$, as H_n is random. We shall verify that H_n satisfies the properties of Θ_o. First, by AS2 $\tau_v \le \tau_t$.

If $a < b$ and $F(a-) = F(b)$, then $[a,b] \cap \mathcal{S}_F = \emptyset$. It follows that $\{R_1, \ldots, R_n\} \cap [a,b] = \emptyset$ by AS2 and thus H_n satisfies (4.4) and $H_n \in \mathcal{D}$ by Convention (2.3).

If $H_n(\tau_t-) = 1$ then (H_n, G, G_T) trivially satisfies AS1 and thus $H_n \in \Theta_o$. Moreover, if $F(\tau_t-) < 1$, then $\mathrm{P}(T \text{ or } V = \tau_t) > 0$ and thus (H_n, G, G_T) also satisfies AS1. It follows that $H_n \in \Theta_o$. Hence, WLOG, we can assume that $F(\tau_t-) = 1$. and $H_n(\tau_t-) < 1$. Then either $P(R = \tau_t) > 0$ or $P(R = \tau_t) = 0$. If $P(R = \tau_t) > 0$, then $\mathrm{P}(V = \tau_t) > 0$ as $F(\tau_t-) = 1$. I.e., (H_n, G, G_T) satisfies AS1 and $H_n \in \Theta_o$. If $P(R = \tau_t) = 0$ then with probability one $R_i \neq \tau_t$. WLOG, we can assume that $R_i \neq \tau_t$. Let x_o be the largest R_i that is smaller than τ_t. Then (2.3) implies that $\mu_{H_n}([x_o, \tau_t]) = 0$. Moreover, $\tau_v \leq \tau_t$ by AS1. Hence, (H_n, G, G_T) satisfies AS3. It follows that $H_n \in \Theta_o$. □

Proof of Lemma 4.4. Let $o_1, \ldots, o_m$ be all the discontinuity points of F_k. Then $\mathcal{D}_k$ is an m-dimensional linear space. Define $h_i(x) = \mathbf{1}_{(x \geq o_i)}$. We shall show that

$$\mathcal{A}_{F_k} = \mathcal{R}_{F_k}(h_i)(o_j) - \mathcal{R}_{F_k}(h_i)(o_j-) \geq 0 \text{ for each } j \text{ and for each } h_i. \tag{B.2}$$

Verify that $\mathcal{R}_{F_k}(h_i) \in \mathcal{D}_k$ (by Lemma 4.2), $\mathcal{R}_{F_k}(h_i)(0-) = 0$ and $\mathcal{R}_{F_k}(h_i)(\infty) = 1$. Then

$$\mathcal{R}_{F_k}(h_i), \quad i = 1, \ldots, m, \text{ are a base of } \mathcal{D}_k \text{ and } \|\mathcal{R}_{F_k}(h_i)\| = 1, \tag{B.3}$$

by Lemma 4.2, as $\mathcal{D}_k$ is an m-dimensional linear space, and

$$h_i, i = 1, \ldots, m, \text{ are a base of } \mathcal{D}_k \text{ and } \|h_i\| = 1. \tag{B.4}$$

(B.3) and (B.4) imply that $\|\mathcal{R}_{F_k}^{-1}\| = 1$.

The proof of the lemma will be completed after we prove (B.2). Letting $x = o_j$, $h = h_i$, $F = F_k$, (B.1) yields

$$\mathcal{R}_F(h)(x) = \int_{l \leq x < r} \left[\frac{F(x) - F(l)}{F(r) - F(l)} h(r) + \frac{F(r) - F(x)}{F(r) - F(l)} h(l)\right] dG^*(l, r) + \beta(x),$$

where $\beta(x) = \pi_0(1 - G_T(x))h(x)$. Moreover, $\{l \leq x- < r\} = \{l < x \leq r\}$ and

$$\begin{aligned}
\mathcal{A}_F = & \int_{l \leq x < r} \left[\frac{F(x) - F(l)}{F(r) - F(l)} h(r) + \frac{F(r) - F(x)}{F(r) - F(l)} h(l)\right] dG^*(l, r) \\
& - \int_{l < x \leq r} \left[\frac{F(x-) - F(l)}{F(r) - F(l)} h(r) + \frac{F(r) - F(x-)}{F(r) - F(l)} h(l)\right] \\
& \cdot dG^*(l, r) + \beta(x) - \beta(x-) \\
= & \int_{l < x < r} \frac{(F(x) - F(x-))(h(r) - h(l))}{F(r) - F(l)} dG^*(l, r) + \int_{l = x < r} h(x) dG^*(l, r) \\
& - \int_{l < x = r} \left[\frac{F(x-) - F(l)}{F(x) - F(l)} h(x) + \frac{F(x) - F(x-)}{F(x) - F(l)} h(l)\right] \\
& \cdot dG^*(l, r) + \beta(x) - \beta(x-).
\end{aligned} \tag{B.5}$$

Replacing F and h by F_k and $\mathbf{1}_{(x \geq o_i)}$, respectively, equation (B.5) yields

$$\begin{aligned}
\mathcal{A}_{F_k} \geq & \int_{l = x < r} h(x) dG^*(l, r) - \int_{l < x = r} h(x) dG^*(l, r) + \beta(x) - \beta(x-) \\
= & \pi_0 P(T = x) h(x) + \pi_0(1 - G_T(x))h(x) - \pi_0(1 - G_T(x-))h(x-) \\
= & \pi_0(1 - G_T(x-))(h(x) - h(x-)) \\
\geq & 0, \text{ which is (B.2).} \ \square
\end{aligned}$$

References

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T. and Silverman, E. (1955). An empirical distribution function for sampling incomplete information, *Ann. Math. Statist.*, **26**, 641–647.

Chang, M. N. and Yang, G. (1987). Strong consistency of a self-consistent estimator of the survival function with doubly-censored data, *Ann. Statist.*, **15**, 1536–1547.

Finkelstein, D. M. and Wolfe, R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data, *Biometrics*, **41**, 933–945.

Gentleman, R. and Geyer, C. J. (1994). Maximum likelihood for interval censored data: Consistency and computation, *Biometrika*, **81**, 618–623.

Groeneboom, P. (1996). *Inverse problem in statistics*, *Proceedings of the St. Flour Summer School in Probability, 1994*, Lecture notes in Math., No. 1648, Springer, Berlin.

Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Birkhäuser, Basel.

Gu, M. G. and Zhang, C. H. (1993). Asymptotic properties of self-consistent estimator based on doubly censored data, *Ann. Statist.*, **21**, 611–624.

Huang, J. (1996). Efficient estimation for proportional hazards models with interval censoring, *Ann. Statist.*, **24**, 540–568.

Huang, J. (1999). Asymptotic properties of nonparametric estimation based on partly interval-censored data, *Statist. Sinica*, **9**, 501–520.

Li, L. X., Watkins, T. and Yu, Q. Q. (1997). An EM algorithm for smoothing the self-consistent estimator of survival functions with interval-censored data, *Scand. J. Statist.*, **24**, 531–542.

Odell, P. M., Anderson, K. M. and D'Agostino, R. B. (1992). Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model, *Biometrics*, **48**, 951–959.

Peto, R. (1973). Experimental survival curves for interval-censored data, *App. Statist.*, **22**, 86–91.

Petroni, G. R. and Wolfe, R. A. (1994). A two-sample test for stochastic ordering with interval-censored data, *Biometrics*, **50**, 77–87.

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer, New York.

Schick, A. and Yu, Q. Q. (2000). Consistency of the GMLE with mixed case interval-censored data, *Scand. J. Statist.*, **27**, 45–55.

Turnbull, B. W. (1976). The empirical distribution function with arbitrary grouped, censored and truncated data, *J. Roy. Statist. Soc. Ser. B.*, **38**, 290–295.

Van De Geer, S. (1996). Rates of convergence for the maximum likelihood estimator in mixture models, *J. Nonparametr. Statist.*, **6**, 293–310.

Wellner, J. A. (1995). Interval censoring case 2: alternative hypotheses, *Analysis of Censored Data*, Proceedings of the workshop on analysis of censored data, December 28, 1994-January 1, 1995, University of Pune, Pune, India, IMS Lecture Notes-Monograph Series, Vol. 27, Hayward, Carifornia.

Wellner, J. A. and Zhan, Y. (1997). A hybrid algorithm for computation of the NPMLE from censored data, *J. Amer. Statist. Assoc.*, **92**, 945–959.

Yu, Q. Q. and Li, L. X. (1994). On the strong consistency of the product limit estimator. *Sankhyā Ser. A*, **56**, 416–430.

Yu, Q. Q. and Li, L. X. (1999). Asymptotic properties of the GMLE of self-consistent estimators with doubly-censored data, *Acta Math. Sinica*. (accepted).

Yu, Q. Q., Li, L. X. and Wong, G. Y. C. (1998*a*). Asymptotic variance of the GMLE of a survival function with interval-censored data, *Sankhyā*, *Ser. A*, **60**, 184–197.

Yu, Q. Q., Schick, A., Li, L. X. and Wong, G. Y. C. (1998*b*). Asymptotic properties of the GMLE in the case 1 interval censorship model with discrete inspection times, *Canad. J. Statist.*, **26**, 619–627.

Yu, Q. Q., Schick, A., Li, L. X. and Wong, G. Y. C. (1998*c*). Asymptotic properties of the GMLE of a survival function with case 2 interval-censored data, *Statist. Probab. Lett.*, **37**, 223–228.

Yu, Q. Q., Li, L. X. and Wong, G. Y. C. (2000). On consistency of the self-consistent estimator of survival functions with interval-censored data, *Scand. J. Statist.*, **27**, 35–44.