ON CONSISTENCY OF THE SELF-CONSISTENT ESTIMATOR OF SURVIVAL FUNCTIONS WITH INTERVAL-CENSORED DATA

By Qiqing Yu, Linxiong Li and George Y. C. Wong

Department of Mathematics Sciences, State University of New York at Binghamton, NY 13902, USA

Department of Mathematics, University of New Orleans New Orleans, LA 70148, USA

and

Strang Cancer Prevention Center, New York, NY 10021, USA

Abstract The self-consistent estimator is commonly used for estimating a survival function with interval-censored data. Recent studies on interval censoring have focused on case 2 interval censoring, which does not involve exact observations, and double censoring, which involves only exact, right-censored or left-censored observations. In this paper, we consider an interval censoring scheme that involves exact, left-censored, right-censored and strictly interval-censored observations. Under this censoring scheme, we prove that the self-consistent estimator is strongly consistent under certain regularity conditions.

Key words and phrases: Case 2 interval-censored data, exact observations, nonparametric maximum likelihood estimator, self-consistent algorithm, strong consistency.

1. Introduction

Recent studies of interval censoring have focused on case 2 interval-censored (IC) data, which involve a time-to-event variable X whose value is never observed but is known to lie in the time interval between two consecutive inspection times Y and Z. Case 2 interval censoring arises naturally in a longitudinal follow-up study in which the event of interest cannot be easily observed (for instance, cancer recurrence, elevation of levels of a biomaker without any noticeable symptoms).

In this paper, we consider IC data which consist of both case 2 IC data and exact observations. We call such data *mixed IC data*. Mixed IC data do arise in clinical follow-up studies. In a cancer follow-up study in which a tumor marker (for instance, CA 125 in ovarian cancer) is available, a patient whose marker value 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 exact 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 such mixed IC 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, time of occurrence of coronary heart disease (CHD) is recorded for almost every participant. However, time of first occurrence of the CHD subcategory angina pectoris may be specified only as between two clinical visits, several years apart, for some of the participants who suffered from angina pectoris.

For case 2 IC data, Groeneboom and Wellner (1992) proposed the iterative convex minorant algorithm for obtaining the nonparametric MLE (NPMLE) of the distribution function, F, of X. The consistency of the NPMLE and the asymptotic distribution of an alternative estimator are obtained under the assumption that F and the inspection time distribution are both continuous and some additional regularity assumptions. Under the only assumption that the random inspection times are discrete, Yu *et al.* (1998) proved the strong consistency of the NPMLE. They further established the asymptotic normality of the NPMLE by requiring that the inspection times to take on only finitely many values.

Another commonly discussed interval censoring scheme is double censoring. Data are said to be subject to double censoring if they are exact, left censored or right censored; however, they are not to be strictly interval censored. For doubly-censored data, the consistency and asymptotic normality of the self-consistent estimator (SCE) have been established by Turnbull (1974), Chang and Yang (1987), and Gu and Zhang (1993) under different assumptions.

For mixed IC data, Peto (1973) obtained the NPMLE of F using a Newton-Raphson type algorithm. Turnbull (1976) proposed a self-consistent algorithm for estimating F and showed that the associated SCE is also the NPMLE. This SCE has been widely employed in medical applications. See, for example, Finkelstein (1986) and Becker and Melbye (1991). In this paper, we shall establish the strong consistency of the SCE under the assumption that F is arbitrary but the support of the inspection times is finite. Although the NPMLE is consistent with case 2 interval-censored data (Groeneboom and Wellner, 1992), counter example does exist and shows that the SCE may not be consistent with case 2 interval-censored data when the inspection times only take on finitely many values (Yu, 1997). Intuitively, the proof for the consistency of the SCE should be different from that of the NPMLE. We shall show that it is indeed the case in Sections 3 and 4.

The organization of the paper is as follows. Section 2 presents models to describe the mixed IC data and two algorithms for computing the SCE. The strong consistency of the SCE is established in Sections 3 and 4. Some proofs are put in the Appendix.

2. Models For Mixed IC Data

We shall discuss two models for mixed IC data in this section. The one in Section 2.2 is more general than the one in Section 2.1, but we shall show that in terms of the properties of the SCE, it suffices to consider the one in Section 2.1.

2.1. A Simple Model For Mixed IC Data

Let (Y, Z) denote a pair of extended random censoring times (∞ allowed). Assume $Y < Z$ with probability one (w.p.1), and X and (Y, Z) are independent. The observable mixed IC data are equivalent to a random interval

$$\lfloor L, R \rfloor = \begin{cases} (Y, \infty) & \text{if } Y < X \text{ and } Z = \infty \text{ (right censored)}, \\ (-\infty, Z] & \text{if } X \le Z \text{ and } Y = -\infty \text{ (left censored)}, \\ (Y, Z] & \text{if } -\infty < Y < X \le Z < \infty \text{ (strictly interval censored)}, \\ [X, X] & \text{if } X \notin (Y, Z] \text{ (exact)}. \end{cases} \tag{2.1}$$

Let $\lfloor L_i, R_i \rfloor$, $i = 1, ..., n$, be a random sample from $\lfloor L, R \rfloor$ and $\lfloor l_i, r_i \rfloor$ be a realization of $\lfloor L_i, R_i \rfloor$. Further, let $Q(l, r) = P(L \le l, R \le r)$. Following Peto (1973) and Turnbull (1976), define *sample innermost intervals*, denoted by $\lfloor p_j, q_j \rfloor$'s, to be the nonempty intersections of the intervals $\lfloor l_i, r_i \rfloor$ so that for any pair of intervals $\lfloor p_j, q_j \rfloor$ and $\lfloor l_i, r_i \rfloor$, either $\lfloor p_j, q_j \rfloor \subseteq \lfloor l_i, r_i \rfloor$ or $\lfloor p_j, q_j \rfloor \cap \lfloor l_i, r_i \rfloor = \emptyset$. Note that $\lfloor p_j, q_j \rfloor$ denotes an half open interval if $p_j < q_j$

and a closed interval if $p_j = q_j$. Moreover, every exact observation constitutes an innermost interval. We demonstrate the concept of innermost interval by an example.

EXAMPLE Suppose n=5 and the observed intervals $\lfloor l_i, r_i \rfloor$ are $(0,3], (2,5], [4,4], (2,\infty)$, and (6, 7]. Then there are three innermost intervals: (2, 3], [4, 4] and (6, 7].

Suppose there are m ($\leq n$) such distinct intervals: $\lfloor p_1, q_1 \rfloor, \lfloor p_2, q_2 \rfloor, \ldots, \lfloor p_m, q_m \rfloor$, where $p_1 \leq q_1 \leq p_2 \leq q_2 \leq \cdots \leq q_m$. Define $\delta_{ij} = I(\lfloor p_j, q_j \rfloor \subseteq \lfloor l_i, r_i \rfloor)$, where $I(A)$ denotes the indicator function of the set A.

The self-consistent algorithm (Turnbull, 1976) for obtaining the SCE $\hat{F}_n$ (which assigns weight to innermost intervals only) of F is given by

$$\hat{F}_n(x) = \sum_{j: q_j \leq x} s_{nj}, \ x \geq 0,$$

where $\{s_{n1}, \ldots, s_{nm}\}$ are the probability masses assigned to the corresponding innermost intervals, and satisfy the self-consistent equations

$$s_{nj} = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_{ij} s_{nj}}{\sum_{k=1}^{m} \delta_{ik} s_{nk}}, \ j = 1, ..., m. \tag{2.2}$$

Li, Watkins and Yu (1997) proposed an alternative approach based on the EM algorithm for obtaining the SCE $\hat{F}_n$ and expressing $H_n = \hat{F}_n$ as a solution of an integral equation

$$H_n(x) = \int \frac{H_n(x) - H_n(l)}{H_n(r-) - H_n(l)} I(l < x < r) dQ_n(l,r) + \frac{1}{n} \sum_{i=1}^{n} I(R_i \leq x), \ H_n \in \Theta, \tag{2.3}$$

where $\Theta = \{h$: h is a nondecreasing function from $[-\infty, \infty]$ to $[0,1]$ such that $h(-\infty) = 0$ and $h(\infty) = 1\}$ and $Q_n(l,r)$ is the empirical version of $Q(l,r)$. They showed that with proper initial values, algorithms (2.2) and (2.3) give the same weight $s_{nj} = H_n(q_j) - H_n(p_j)$ to the innermost interval $\lfloor p_j, q_j \rfloor$, when it is not closed, or the same weight $H_n(q_j) - H_n(p_j-)$ to $\lfloor p_j, q_j \rfloor$, when it is closed. That is, H_n and $\hat{F}_n$ are equivalent. We shall make use of expression (2.3) to establish the strong consistency of $\hat{F}_n$ in Sections 3 and 4.

Following the identifiability assumption given in Chang and Yang (1987), we define $K(x) = P\{X$ is not censored $|X = x\}$ for each x. Let $\tau_l = \inf\{x : K(x) = 0\}$ and $\tau_r = \sup\{x : K(x) = 0\}$, if $\{x : K(x) = 0\} \neq \emptyset$. Otherwise, define $\tau_l = \tau_r = \infty$. For each $x \in (\tau_l, \tau_r)$, either $K(x) = 0$ or $K(x)$ is not defined. To see this, it suffices to show that for any two points $a < b$ satisfying $K(a) = K(b) = 0$, there do not exist $x \in (a, b)$ such that $K(x) > 0$. In fact, $K(a) = 0$ implies that $P\{Y < a\} \geq P\{Y < a \leq Z\} = 1$. Also, $K(b) = 0$ implies that $P\{Z \geq b\} \geq P\{Y < b \leq Z\} = 1$. Thus

$$P\{Y \leq \tau_l \leq \tau_r \leq Z\} = 1 \text{ and } K(x) = 0 \; \forall \; x \in (\tau_l, \tau_r). \tag{2.4}$$

There are only four possible cases that model (2.1) implies: (1) $\tau_l > -\infty$ and $\tau_r = \infty$, (2) $\tau_l = -\infty$ and $\tau_r < \infty$, (3) $-\infty < \tau_l < \tau_r < \infty$, and (4) $\{x : K(x) = 0\} = \emptyset$. Case (1) is a right censorship model as $\mathrm{P}(Z = \infty) = 1$ by (2.4). Moreover, case (2) is a left censorship model. Both of them do not allow strictly interval-censored observations. If case (3) is true, then $Y \leq \tau_l$ and $Z \geq \tau_r$ w.p.1, which is not practically realistic. Thus case (4) is the only practical case in model (2.1) that includes both strictly interval-censored observations and exact observations. In next subsection we see how to extend model (2.1) to cover more general situations.

2.2. A Model for More General Mixed IC Data.

Even though case (4) in Section 2.1 does not have the drawback as in the first three cases, it implies that $\mathrm{P}\{K(X) > 0\} = 1$. It is often the case that a study can only last for a certain period of time, say, a time interval $[a, b]$, where $0 \leq F(a) < F(b) < 1$. In such a case, the mixed interval-censored observation $\lfloor L, R \rfloor$ satisfies

$$\{L \text{ or } R \in (-\infty, a) \cup (b, +\infty)\} = \emptyset. \tag{2.5}$$

Consequently, $\mathrm{P}\{K(X) > 0\} \leq P\{a < X \leq b\} < 1$. Thus, model (2.1) cannot specify such mixed IC data. Note that (2.1) is equivalent to

$$\lfloor L, R \rfloor = \begin{cases} (Y, Z] & \text{if } X \in (Y, Z], \\ [X, X] & \text{if } X \notin (Y, Z]. \end{cases} \tag{2.6}$$

We now formulate a model for mixed IC data satisfying (2.5). Assume $Y < Z$ w.p.1., and $\{Y \text{ or } Z \in (-\infty, a) \cup (b, \infty)\} = \emptyset$. Suppose that X and (Y, Z) are independent and the observable random vector

$$\lfloor L, R \rfloor = \begin{cases} (Y, Z] & \text{if } X \in (Y, Z], \\ [X, X] & \text{if } X \notin (Y, Z] \text{ and } a < X \leq b, \\ (-\infty, a] & \text{if } X \notin (Y, Z] \text{ and } X \leq a, \\ (b, \infty) & \text{if } X \notin (Y, Z] \text{ and } X > b\;. \end{cases} \tag{2.7}$$

In the case of (2.5) or (2.7), we can only estimate $F(x)$ for x in $[a, b]$, or equivalently, the cdf F^* of X^*, where $X^* = aI(X \leq a) + XI(a < X \leq b) + 2bI(X > b)$. Note that X^* and (Y, Z) are independent. Due to (2.5) or (2.7), the right-censored observation (b, ∞) will always be an innermost interval. The NPMLE (or an SCE) $\hat{F}(x)$ is not uniquely determined for $x \in (b, \infty)$ (see, e.g., Peto, 1973), though the total mass assigned by the NPMLE (or the SCE) to the interval (b, ∞) is uniquely determined. Thus we can, without loss of generality (WLOG), assume that the mass is put at the point $2b$ $(\in (b, \infty))$. In other words, (b, ∞) can be treated as an exact observation $[2b, 2b]$. For a similar reason, the left-censored observation $(-\infty, a)$ can be treated as an exact observation $[a, a]$. Thus model (2.7) is equivalent to

$$\lfloor L, R \rfloor = \begin{cases} (Y, Z] & \text{if } X^* \in (Y, Z], \\ [X^*, X^*] & \text{if } X^* \notin (Y, Z], \end{cases} \tag{2.8}$$

If $F(a) = 0$ and $F(b) = 1$, then models (2.7) and (2.8) are the same as (2.1) (or (2.6)).

In view of (2.6) and (2.8), it is easy to see that in the case of (2.5) or (2.7), in order to estimate F, it suffices to estimate F^*, which reduces model (2.7) to model (2.1). Similar modification can be made to handle the situation that there are no observations L or R in a union of arbitrary intervals. In view of the above discussion, we shall focus on model (2.1) for the rest of the paper.

3. Consistency In Case Of Finite Support For F

In this section, we assume that both the support of X, say S_F, and the support of Y and Z, say S_G, contain finitely many points. The generalization of F to an arbitrary distribution function is given in Section 4. The assumption concerning S_G is a reasonable one. In practice inspections of most follow-up studies are recorded on a discrete time scale (daily, weekly, monthly, etc.), and the total study period is finite, so the number of censoring points, i.e. the support of Y and Z, is also finite. Such an assumption was adopted by Finkelstein (1986) and Becker and Melbye (1991), among others.

Suppose that X takes on values $x_1, x_2, \ldots, x_u$, and $\lfloor L, R \rfloor$ takes on values $I_1 = \lfloor l_1^o, r_1^o \rfloor$, $I_2 = \lfloor l_2^o, r_2^o \rfloor$, $\ldots$, $I_v = \lfloor l_v^o, r_v^o \rfloor$ with probability $e_i = P\{L = l_i^o, R = r_i^o\} > 0$. Based on the assumption that $K(x) > 0$ for all $x > 0$, Chang and Yang (1987) and Gu and Zhang (1993) proved the consistency of the SCE for doubly-censored observations. In this paper, we weaken this assumption and prove the consistency of the SCE on the set $\mathcal{O} = \{x; K(x) > 0\}$ with mixed IC data.

For a point x satisfying $K(x) = 0$ and $P\{X = x\} > 0$, since there are no exact observations available at this point, the distribution function F is not estimable, and hence consistency cannot be assessed. Let us consider the structure of the innermost intervals as sample size $n \to \infty$. For x_i, if $K(x_i) > 0$, then it follows from the strong law of large number that $P\{X_k \neq x_i, \text{ for all } k = 1, 2, \ldots, n\} \to 0$ as $n \to \infty$. In other words, $K(x_i) > 0$ implies that $[x_i, x_i]$ is an innermost interval w.p.1, which further implies that the union of all closed innermost intervals coincides with the set $\mathcal{O}$. Let $A_1, A_2, \ldots, A_{m_1}$ be the innermost intervals induced by the intervals $I_i, i = 1, 2, \ldots, v$, and call them *population innermost intervals*. It is seen that as $n \to \infty$ the set of sample innermost intervals induced by $\lfloor L_i, R_i \rfloor, i = 1, 2, \ldots, n$ converges almost surely to the set of population innermost intervals. Since we are only concerned with large sample properties, we can, WLOG, assume that the sample size is large enough so that $m = m_1$. For the rest of the paper, m will be used to denote both the number of population innermost intervals and the number of sample innermost intervals. Also we shall suppress the qualifier w.p.1 throughout the rest of the paper to avoid repetition.

Let $\mathbf{s}_n = (s_{n1}, s_{n2}, \ldots, s_{nm})$ be a solution of (2.2). For sufficiently large n, the s_{ni}'s are the masses assigned to the innermost intervals by the SCE. Since $\{H_n, n \geq 1\}$ is a bounded monotone sequence, it follows from Helly-Bray selection theorem that there exist a subsequence, say $\{n_k\}$, of integers, a function H and a vector $\mathbf{s}$, such that $\lim_{n_k \to \infty} H_{n_k}(x) = H(x)$ and $\lim_{n_k \to \infty} \mathbf{s}_{n_k} = \mathbf{s} = (s_1, s_2, \ldots, s_m)$. Taking the limit in (2.2) and (2.3) with respect to n_k, we obtain

$$s_j = \sum_{i=1}^{v} e_i \frac{\delta_{ij} s_j}{\sum_{k=1}^{m} \delta_{ik} s_k}, \ s_j \geq 0, \ j = 1, ..., m, \ \sum_{j=1}^{m} s_j = 1, \tag{3.1}$$

where $\delta_{ij} = I(A_j \subseteq \lfloor l_i^o, r_i^o \rfloor)$, and

$$H(x) = \int_{l \leq x < r} \frac{H(x) - H(l)}{H(r-) - H(l)} dQ(l, r) + P\{R \leq x\}, \ H \in \Theta, \tag{3.2}$$

since Q_{n_k} converges to Q almost surely as $n_k \to \infty$.

We state two lemmas. The proof of Lemma 1 is relegated to the appendix. Hereafter, the discussion regarding uniqueness of the solution $H(x)$ of Eq. (3.2) will be restricted to the set $\mathcal{O}$.

Lemma 1 *Let* $\mathbf{s}^o = (s_1^o, s_2^o, \ldots, s_m^o)$ *where* $s_j^o = P\{X \in A_j\}$, $j = 1, 2, \ldots, m$. *Then* $\mathbf{s} = \mathbf{s}^o$ *is the unique solution of Eq. (3.1).*

Lemma 2 (Li, Watkins and Yu, 1997) *Let dH be the measure induced by a c.d.f. H and $s_j = dH(A_j)$ for all j. Then $\mathbf{s} = (s_1, \ldots, s_m)$ is a solution to Eq. (3.1) if and only if H is a solution to Eq. (3.2).*

Theorem 1 *Suppose S_F and S_G contain finitely many points. Then (1) F is the unique solution of Eq. (3.2) for $x \in \mathcal{O}$, and (2) the SCE $H_n(x)$ of $F(x)$ satisfies $\sup_{x\in\mathcal{O}} |H_n(x) - F(x)| \to 0$, as $n \to \infty$.*

Proof. Since $s_j^o = dF(A_j)$, F is a solution of Eq. (3.2) by Lemmas 1 and 2. Meanwhile, for each solution H of (3.2), dH is uniquely determined by Lemmas 1 and 2 again. Consequently, statement (1) follows.

It follows from the convergence of H_{n_k} and (1) above that $H_{n_k}(x) \to F(x)$ as $n_k \to \infty$ for $x \in \mathcal{O}$. The convergence of $H_n(x)$ to $F(x)$ for $x \in \mathcal{O}$ follows from Helly-Bray selection theorem. The uniform convergence of H_n is immediate by the assumption that S_F and S_G contain only finitely many points. □

REMARK. Even though $F(x)$ is not estimable for $x \in (\tau_l, \tau_r)$, it is easy to see that $H(\tau_r) - H(\tau_l) = F(\tau_r) - F(\tau_l)$. This remark is also valid for model (2.7). Moreover, under model (2.7) $F(x)$ is not estimatable for $x < a$ or $x > b$.

4. Consistency Of H_n For Arbitrary F

In this section, we extend the result of the previous section to the case where G is the same as previously defined but F is arbitrary.

Theorem 2 *Suppose that (Y, Z) takes on finitely many values and F is arbitrary. Then the solution to Eq. (3.2) is unique. Furthermore, $H_n(x)$ converges to $F(x)$ uniformly for all $x \in \mathcal{O}$.*

Proof. The main idea of the proof is to partition the interval $[0, \infty)$ into finitely many subintervals, and then to prove the consistency of the SCE for every such subinterval.

WLOG, assume that the values that (Y, Z) can take are (a_i, b_i), $i = 1, ..., N_G$, for some integer N_G. Rank the $2N_G$ values $\{a_i, b_i\}$ in increasing order to obtain a sequence (ties and infinity are allowed). Let $d_1 < d_2 < \cdots < d_N$ ($N \le 2N_G$) be the distinct finite values of the sequence. We first partition $[0, \infty)$ into

$$[0, 0], (0, d_1), [d_1, d_1], (d_1, d_2), \ldots, [d_N, d_N], (d_N, \infty). \tag{4.1}$$

Note that in this partition, all exact observations in the same interval (d_j, d_{j+1}) (or $[d_j, d_j]$) carry the same weight. This is because for any observed interval, if (d_j, d_{j+1}) (or $[d_j, d_j]$) is not a subset of the interval, then it is disjoint from the observed interval, and because the weight received by an innermost interval is determined by all the observed intervals that cover the innermost interval (see (2.2)).

For a fixed $\epsilon > 0$, if there is a value d in an open interval (d_j, d_{j+1}) such that $P\{X = d\} \geq \epsilon$, divide the interval into $(d_j, d), [d, d], (d, d_{j+1})$. Perform the partitioning for every such d. Since the set of such d values is finite, the total number of intervals partitioning $[0, \infty)$ must also be finite. WLOG, assume (4.1) is the final partition at this stage.

Consider an interval, say, (d_1, d_2), such that $F(d_2-) - F(d_1) > 0$. For this fixed ϵ, partition (d_1, d_2) into subintervals, say $(c_1, c_2), [c_2, c_2], (c_2, c_3,), \ldots, (c_k, c_{k+1})$ ($c_1 = d_1$ and $c_{k+1} = d_2$) such that $F(c_{i+1}-) - F(c_i) < \epsilon$ for $i = 1, 2, \ldots, k$. Perform this second partition for every interval (d_i, d_{i+1}) and $[d_i, d_i]$ for all $i = 0, 1, \ldots, N$, where $d_0 = 0$.

From now on, we focus our discussion on (d_1, d_2). The argument for other intervals is similar. Let c'_i be the midpoint of the interval (c_i, c_{i+1}), $i = 1, ..., k$, and construct a new (pseudo) distribution function F' with finite support, $F'(d_i) = F(d_i)$ and $F'(c'_i) - F'(c'_i-) = F(c_{i+1}-) - F(c_i)$, for all i. It is readily seen that

$$\sup_x |F(x) - F'(x)| < \epsilon. \tag{4.2}$$

It can be verified that if (τ_l, τ_r) is not an empty set, then one of (d_i, d_{i+1}) must be (τ_l, τ_r), due to the special structure of partition (4.1). In addition, since consistency is restricted to $\mathcal{O}$, it is natural to assume that $S_F \cap (d_1, d_2) \subset \mathcal{O}$. Moreover, since $F(d_2-) - F(d_1) > 0$ and $S_F \cap (d_1, d_2) \subset \mathcal{O}$, the probability of having exact observations in (d_1, d_2) converges to one as $n \to \infty$. Thus, for n large enough, we can eventually observe exact observations in the interval and hence $(d_1, d_2]$ cannot be an half open innermost interval.

Consider a pseudo random variable $X' = \sum_i [c'_i I(X \in (c_i, c_{i+1}) + c_i I(X = c_i)]$. Note that X' has the distribution function F'. Suppose there are h exact observations in (c_i, c_{i+1}), then the pseudo random variable X' will assume the value c'_i as an exact observation a total of h times. For sample size n, let w_{ni} denote the weight received by all exact observations in (c_i, c_{i+1}), and let $w'_{ni} = w_{ni}$ be the weight received by c'_i from the pseudo observations generated by X'. Let H'_n denote the corresponding SCE of F' associated with X' and H_n the SCE of F associated with X. It is easy to see that for each i and $D_i = (d_i, d_{i+1})$ or

$[d_i, d_i]$,

$$dH_n(D_i) = dH_n'(D_i). \tag{4.3}$$

By the results of the previous section and the finiteness of support of F', $H_n'(d_i) \to F'(d_i)$ for each i as $n \to \infty$. Thus, it follows from (4.2) and (4.3) that when n is large enough,

$$\sup_{x \in \mathcal{O}} |F(x) - H_n(x)| \le \epsilon, \tag{4.4}$$

which proves the consistency of the SCE H_n. By Helly-Bray selection theorem and the fact that F is a solution of (3.2), (4.4) also shows that F is the unique solution of (3.2). ◻

Appendix

In this appendix, we give a proof of Lemma 1. To show the uniqueness of the solution of (3.1), consider the generalized log-likelihood function

$$\Lambda = \Lambda(\mathbf{s}) = \sum_{i=1}^{v} e_i \big(\ln \sum_{j=1}^{m} \delta_{ij} s_j\big).$$

It follows from (2.2) that $\sum_{j=1}^m \delta_{ij} s_j > 0$ for every i, $1 \le i \le v$ and hence the function Λ is well defined. Let $\mathbf{s}^e = (s_1^e, s_2^e, \ldots, s_m^e)$ denote the solution of Eq. (3.1) and $\mathbf{s}^o = (s_1^o, s_2^o, \ldots, s_m^o)$ the probability vector with $s_i^o =$P$\{X \in A_i\}$.

To facilitate the proof of Lemma 1, we first establish three lemmas. Following the notations of Section 3, let $I_i = \lfloor l_i^o, r_i^o \rfloor$ and $e_i = P\{\lfloor L, R \rfloor = I_i\} = \alpha_i^o g_i$, where

$$\alpha_i^o = P\{X \in I_i\} = \sum_{j=1}^{m} \delta_{ij} s_j^o, \ i = 1, ..., v, \tag{A.1}$$

and $g_i = P\{\lfloor Y, Z \rfloor = I_i\}$ if $l_i^o < r_i^o$ and $P\{l_i^o \notin (Y, Z)\}$ if $l_i^o = r_i^o$. Thus $\sum_i \alpha_i^o g_i = \sum_i e_i = 1$. It is seen that $\mathbf{s}^o$ uniquely determines $\mathbf{a}^o = (\alpha_1^o, \alpha_2^o, \ldots, \alpha_v^o)$. Let $\mathbf{a} = (\alpha_1, \ldots, \alpha_v)$, with $\alpha_i = \sum_{j=1}^m \delta_{ij} s_j$. Then Λ can be rewritten as

$$\Lambda(\mathbf{s}) = \sum_{i=1}^{v} e_i \ln \alpha_i \equiv \lambda(\mathbf{a}).$$

Thus maximizing $\Lambda(\mathbf{s})$ is equivalent to maximizing $\lambda(\mathbf{a})$. Note that α_i^o, e_i and g_i are fixed, but α_i are not.

Let X_s be a random variable such that $P\{X_s \in A_j\} = s_j$, $j \geq 1$. Define a new random interval $\lfloor L_s, R_s \rfloor$ to be the counterpart of $\lfloor L, R \rfloor$ in (2.1) with X replaced by X_s. Thus α_i satisfies $\alpha_i g_i = P\{\lfloor L_s, R_s \rfloor = I_i\}$. It follows that

$$\sum_i \alpha_i g_i = 1. \tag{A.2}$$

Lemma A1 *Suppose* $\mathbf{a}$ *satisfies (A.2). Then* $\lambda(\mathbf{a})$ *is uniquely maximized by* $\mathbf{a}^o$*, where* $\alpha_i^o = e_i/g_i$.

Proof. Let $t_i = \alpha_i g_i$, $i = 1, ..., v$. Then $t_i \in [0,1]$ and $\sum_i t_i = 1$. Let $h(\mathbf{t}) = \sum_{i=1}^v e_i \ln t_i$. Note that

$$h(\mathbf{t}) = \sum_i e_i \ln \alpha_i + \sum_i e_i \ln g_i = \lambda(\alpha) + \sum_i e_i \ln g_i$$

and $\sum_i e_i \ln g_i$ is fixed under the given assumption. Hence, maximizing $\lambda(\mathbf{a})$ is equivalent to maximizing $h(\mathbf{t})$. It can be shown that $h(\mathbf{t})$ is uniquely maximized by $t_i = e_i$, $i \geq 1$. Therefore, the unique maximizer for $\lambda(\mathbf{a})$ is $\alpha_i^o = e_i/g_i$. □

Lemma A2 $\mathbf{s}^o$ *is the unique maximum point of* $\Lambda(\mathbf{s})$.

Proof. Following the notations in Lemma A1, we have $\Lambda(\mathbf{s}) = \lambda((\mathbf{a}(\mathbf{s}))$. By Lemma A1 and the equality $\mathbf{a}(\mathbf{s}^o) = \mathbf{a}^o$ (due to (A.1)), $\mathbf{s}^o$ is a maximum point of $\Lambda(\mathbf{s})$. By the finiteness assumption on S_F and S_G, each population innermost interval is a realization of $\lfloor L, R \rfloor$, say, $A_j = I_{i_j}$, except perhaps (τ_l, τ_r) if (τ_l, τ_r) is not an empty set. Thus (A.1) implies that there are at least $m-1$ (out of v) js such that $\alpha_{i_j}^o = s_j^o$. Since $\lambda((\mathbf{a}(\mathbf{s}))$ is uniquely maximized by $\mathbf{a}^o$, $\mathbf{s}^o$ is the unique maximum point of $\Lambda(\mathbf{s})$. □

Lemma A3 $s_k^o > 0$ implies that $s_k^e > 0$.

Proof. Note that for each k, if $\mathrm{P}\{X \in A_k\} > 0$, then there is an integer h such that $I_h = A_k$ and thus $\delta_{hk} = 0$ if $h \neq k$ and 1 otherwise. For $j = k$, Eq. (3.1) yields

$$s_k^e \geq e_h \frac{s_k^e}{s_k^e} = e_h > 0 \qquad \text{(since } I_h = A_k \text{ and } e_h > 0\text{)}.$$

This completes the proof of the lemma. □

We are now ready to prove Lemma 1. By Lemma A2, $\mathbf{s}^o$ is the unique maximizer of Λ. Thus, to prove the lemma, it suffices to show that $\mathbf{s}^o = \mathbf{s}^e$. Consider the effect of increasing a particular component, s_k, by a small amount u and then dividing all the s_j, including $s_k + u$, by $1+u$ in order to ensure that the components of $\mathbf{s}$ sum to 1. Let $d_k(\mathbf{s}) = \frac{\partial \Lambda}{\partial u}\Big|_{u=0}$. Then

$$\begin{aligned}
d_k(\mathbf{s}) =& \frac{\partial}{\partial u}\Lambda(\frac{s_1}{1+u}, ..., \frac{s_k+u}{1+u}, ..., \frac{s_m}{1+u})\Big|_{u=0} \\
=& \sum_{i=1}^{v} e_i \frac{\partial}{\partial u} \ln \sum_{j=1}^{m} \delta_{ij} s_j^r\Big|_{u=0} \qquad \left(s_j^r = \begin{cases} s_j/(1+u) & \text{if } j \neq k \\ (s_k+u)/(1+u) & \text{if } j = k \end{cases}\right) \\
=& \sum_{i=1}^{v} e_i \frac{\sum_{j=1}^{m} \delta_{ij} \frac{\partial}{\partial u} s_j^r}{\sum_{j=1}^{m} \delta_{ij} s_j^r}\Big|_{u=0} \\
=& -\sum_{i=1}^{v} e_i \Big(1 - \frac{\delta_{ik}}{\sum_{j=1}^{m} \delta_{ij} s_j}\Big), \quad k = 1, ..., m.
\end{aligned} \qquad (A.3)$$

Consider two separate situations regarding the values of s_k^o.

CASE 1. $s_k^o > 0$, for all k.

If $\mathbf{s}^e$ is a solution to (3.1), then it follows that $s_k^e > 0$ for all k. Consequently,

$$0 = \sum_{i=1}^{v} e_i \Big(1 - \frac{\delta_{ik}}{\sum_{j=1}^{m} \delta_{ij} s_j^e}\Big), \quad k = 1, ..., m, \qquad (A.4)$$

since $\sum_{i=1}^{v} e_i = 1$. In view of (A.3) and (A.4), $d_k(\mathbf{s}^e) = 0$ and $s_k^e > 0$ for each k. Therefore, $\mathbf{s}^e$ is the maximum point of L. By Lemma A2, $\mathbf{s}^e = \mathbf{s}^o$.

CASE 2. $s_k^o = 0$ for some k.

WLOG, assume that $s_m^o = 0$ and $s_k^o > 0$ for $k = 1, 2, \ldots, m-1$. We shall show that $s_m^e > 0$ leads to a contradiction. If $s_m^e > 0$, then $s_i^e > 0$ for all i by Lemma A3. Consequently, (A.4) holds. By virtue of (A.3) and (A.4), $\mathbf{s}^e$ is a maximum point of Λ, and it follows that $\mathbf{s}^e = \mathbf{s}^o$. However, this contradicts the hypothesis that $s_m^e > s_m^o = 0$. This completes the proof for Case 2 and thus the proof of the lemma. □

Acknowledgements: We would like to thank the referees and the associate editor for their valuable comments. This research was supported by NSF Grant DMS-9402561 and DAMD17-94-J-4332 to Qiqing Yu, by LEQSF Grant RD-A-31 to Linxiong Li, and by DAMD17-94-J-4332 to George Y. C. Wong.

References

[1] Becker, N. and Melbye, M. (1991). Use of a log-linear model to compute the empirical survival curve from interval-censored data, with application to data on tests for HIV positivity. *Austral. J. Statist.*, 33, 125-133.

[2] 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.

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

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

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

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

[7] 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.

[8] Peto, R. (1973). Experimental survival curves for interval-censored data. *Applied Statistics*, 22, 86-91.

[9] Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *JASA,* 69, 169-173.

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

[11] Yu, Q., Schick, A., Li, L. and Wong G. (1998). Asymptotic properties of the GMLE of a survival function with case 2 interval-censored data. *Statistics & Probability Letters*, 23, 223-228.

[12] Yu, Q. (1997). Nonparametric and parametric estimation with interval-censored data. *Technical report.* Department of Mathematical Sciences, State University of New York at Binghamton.