The Annals of Statistics
1987, Vol. 15, No. 4, 1536–1547

STRONG CONSISTENCY OF A NONPARAMETRIC ESTIMATOR OF THE SURVIVAL FUNCTION WITH DOUBLY CENSORED DATA[1]

BY MYRON N. CHANG AND GRACE L. YANG

University of Florida and University of Maryland

A double censoring mechanism is such that each variable X in the sample is observable if and only if X is within the observation interval $[Z, Y]$. Otherwise, we can only determine whether X is less than Z or greater than Y and observe Z or Y correspondingly. This kind of censoring occurs often in collecting lifetime data. Our problem is to estimate the survival function of X, $S_X(t) = P[X > t]$, from a doubly censored sample, where X is assumed to be independent of the random interval $[Z, Y]$. We establish sufficient conditions for which $S_X(t)$ is identifiable and then prove the strong consistency of the self-consistent estimator $\hat{S}_X(t)$ for $S_X(t)$. This investigation generalizes the results available for the right censored data.

1. Introduction. Due to sampling methods and factors beyond experimental control, the measurements on lifetime have the possibility of being censored either from above or below. For example, consider a follow-up study for determining the age (X) at which a child first develops a certain skill [Leiderman, Babu, Kagia, Kraemer and Leiderman (1973)]. The age X can be determined if the child develops the skill after he is admitted to the program. However, for some children in the program, the development may have been completed before the first survey, and this results in a left censoring of X. On the other hand, a right censoring may occur when a child either is lost to follow-up before the last survey or has not developed the skill by the time of the termination of the program.

Singly censored data, particularly right censored data, have been the subject of extensive research in the literature, especially in biometry and reliability theory. Statistical inference for doubly censored data was considered by Gehan (1965), Mantel (1967), Peto (1973), Turnbull (1974) and Tsai and Crowley (1985).

Let X be a nonnegative random variable denoting the lifetime under investigation. The censoring mechanism is such that X is observable if and only if X lies in the interval $[Z, Y]$. The Z and Y are nonnegative random variables and $Z \leq Y$ with probability 1. If X is not in $[Z, Y]$, the exact value of X cannot be determined. We can only know whether X is less than Z or greater than Y and we can only observe Z or Y correspondingly. The variable X is said to be left censored if $X < Z$ and right censored if $X > Y$. The available information on X

Received July 1984; revised March 1987.
[1]Research supported in part by the USDA under contract 58-3204-0-117.
AMS 1980 *subject classifications*. Primary 62G05; secondary 62G99.
Key words and phrases. Survival distributions, doubly censored data, self-consistent estimation, uniform strong consistency.

may be expressed by a pair of random variables W and δ, where

$$W = \max(\min(X, Y), Z),$$

(1.1)
$$\delta = \begin{cases} 1, & \text{if } Z \le X \le Y, \\ 2, & \text{if } X > Y, \\ 3, & \text{if } X < Z. \end{cases}$$

Our problem is to estimate the survival function of X, $S_X(t) = P[X > t]$, for all $t > 0$ from a sample of n independent pairs, (W_i, δ_i) for $i = 1, \dots, n$, where (W_i, δ_i) is defined as in (1.1).

This problem was considered by Turnbull (1974), who used Efron's self-consistent criterion to construct an estimator for S_X, commonly called a self-consistent estimator. In this article, we prove that the model (1.1) is identifiable under a set of sufficient conditions. We then show that the self-consistent estimator is strongly consistent under an identifiable model. The result is a generalization of that of the product-limit estimator for singly censored data.

While preparing a revision for this paper, a related paper by Tsai and Crowley (1985) was brought to our attention by the Editor. In their paper, strong consistency and asymptotic normality of self-consistent estimators are studied under various censoring mechanisms. It is stated without proof in their Corollary 5.2 that in the doubly censored model S_X can be uniquely identified by the joint distribution of W and δ. However, this claim is not true without imposing appropriate probabilistic assumptions on the model and the proof of identifiability is nontrivial. The strong consistency is given in their Theorem 4.1 under the assumption that the model considered is identifiable in the family $\mathscr{P}$ of all joint survival functions of X, Y and Z. Thus, their Theorem 4.1 is not applicable to the doubly censored model since the model is not identifiable in $\mathscr{P}$ but only in a subfamily $\mathscr{P}^*$ of $\mathscr{P}$ as defined in our Assumptions A and B in Section 2. Aside from these comments, there is a gap in the proof of their Theorem 4.1, namely, the claim that any solution to $H(F_y^n, G_X) = 0$ is in the neighborhood N_ε (in their notations) does not follow from their proof.

We derive, in Section 2, a system of integral equations which relates the survival functions of X, Y and Z to the joint distribution functions of W and δ in (1.1). This system is fundamental to our investigation. The self-consistent estimator will be derived from the sample counterpart of the system. In Section 3, the system is used to address the problem of identifying $S_X(t)$ under the model (1.1). [Identifiability in right censored data has been studied by Peterson (1977).] We give an example of a nonidentifiable model, which motivates us to introduce a set of sufficient conditions for identifiability. In Section 4, we prove the strong consistency of the estimator. The problem is nontrivial because of the lack of an explicit form for the estimator.

2. The integral equations relating distributions of X, Y and Z to that of (W, δ) and the self-consistent estimator. As given in (1.1), the information on random variables (X_i, Y_i, Z_i) is available only in a sample of (W, δ), $\{(W_i, \delta_i), i = 1, \dots, n\}$. The joint distribution of W and δ can be expressed in terms of subsurvival functions $Q_1(t)$, $Q_2(t)$ and $Q_3(t)$ defined, for $t \geq 0$, as

$$\begin{aligned} Q_1(t) &= P[W > t, \delta = 1] = P[X > t, Z \leq X \leq Y], \\ Q_2(t) &= P[W > t, \delta = 2] = P[Y > t, X > Y], \\ Q_3(t) &= P[W > t, \delta = 3] = P[Z > t, X < Z]. \end{aligned} \tag{2.1}$$

The survival function of W is

$$Q(t) = P[W > t] = \sum_{j=1}^{3} Q_j(t). \tag{2.2}$$

The $Q_j(t)$'s are to be estimated by the empirical subsurvival functions $Q_j^{(n)}(t)$ defined as

$$Q_j^{(n)}(t) = \frac{1}{n} \sum_{i=1}^{n} I_{[W_i > t,\, \delta_i = j]}, \quad \text{for } j = 1, 2, 3 \text{ and } t \geq 0. \tag{2.3}$$

The empirical subsurvival functions $Q_j^{(n)}$, $j = 1, 2, 3$, contain all the information in the sample; they are sufficient statistics for S_X. The estimation problem for S_X will be investigated in terms of the $Q_j^{(n)}$'s and the Q_j's.

The following probabilistic assumptions for the data (1.1) are imposed throughout the paper.

ASSUMPTION A.

A1. The random variable X_i and the vector (Y_i, Z_i) are independent for each i and the vectors (X_i, Y_i, Z_i), $i = 1, \dots, n$, are independently and identically distributed.

A2. $P(Z \leq Y) = 1$.

A3. S_X, S_Y and S_Z are continuous functions of t on $t \geq 0$ and $0 < S_X(t) < 1$ for $t > 0$, where $S_Y(t) = P(Y > t)$ and $S_Z(t) = P(Z > t)$.

Under Assumption A, we derive a system of integral equations which relates the survival functions S_X, S_Y and S_Z to the subsurvival functions Q_1, Q_2 and Q_3. It follows from (2.1) that

$$\begin{aligned} Q_1(t) &= P(X > t, Z \leq X \leq Y) \\ &= -\int_t^\infty P(Z \leq u \leq Y)\, dS_X(u) \\ &= -\int_t^\infty (S_Y(u) - S_Z(u))\, dS_X(u), \end{aligned} \tag{2.4}$$

and similarly for Q_2 and Q_3. The system of integral equations is given by

$$\begin{aligned} Q_1(t) &= -\int_t^\infty (S_Y(u) - S_Z(u))\, dS_X(u), \\ Q_2(t) &= -\int_t^\infty S_X(u)\, dS_Y(u), \\ Q_3(t) &= -\int_t^\infty (1 - S_X(u))\, dS_Z(u). \end{aligned} \tag{2.5}$$

This system is fundamental to our investigation. The integrals considered are Lebesgue–Stieltjes integrals. From integration by parts, it follows that

$$Q(t) = S_Z(t) + S_X(t)(S_Y(t) - S_Z(t)). \tag{2.6}$$

In estimating S_X with the empirical subsurvival functions $Q_j^{(n)}(t)$, it is reasonable to require the estimators $S_X^{(n)}$, $S_Y^{(n)}$ and $S_Z^{(n)}$ to relate to the $Q_j^{(n)}$'s in a similar way, i.e.,

$$\begin{aligned} Q_1^{(n)}(t) &= -\int_t^\infty \left(S_Y^{(n)}(u) - S_Z^{(n)}(u)\right) dS_X^{(n)}(u), \\ Q_2^{(n)}(t) &= -\int_t^\infty S_X^{(n)}(u)\, dS_Y^{(n)}(u), \\ Q_3^{(n)}(t) &= -\int_t^\infty \left(1 - S_X^{(n)}(u)\right) dS_Z^{(n)}(u). \end{aligned} \tag{2.7}$$

Imposing the conditions on $S_Y^{(n)}$ and $S_Z^{(n)}$,

$$S_Y^{(n)}(0) = 1 \quad \text{and} \quad S_Z^{(n)}(\infty) = 0,$$

system (2.7) implies that

$$S_Y^{(n)}(t) = 1 + \int_0^t \frac{dQ_2^{(n)}(u)}{S_X^{(n)}(u)}, \tag{2.8}$$

$$S_Z^{(n)}(t) = -\int_t^\infty \frac{dQ_3^{(n)}(u)}{1 - S_X^{(n)}(u)}, \tag{2.9}$$

$$Q^{(n)}(t) = S_Z^{(n)}(t) + S_X^{(n)}(t)\left(S_Y^{(n)}(t) - S_Z^{(n)}(t)\right). \tag{2.10}$$

Substituting expressions (2.8) and (2.9) in (2.10), we obtain

$$\begin{aligned} S_X^{(n)}(t) = Q^{(n)}(t) &- S_X^{(n)}(t) \int_0^t \frac{dQ_2^{(n)}(u)}{S_X^{(n)}(u)} \\ &+ \left(1 - S_X^{(n)}(t)\right) \int_t^\infty \frac{dQ_3^{(n)}(u)}{1 - S_X^{(n)}(u)}, \end{aligned} \tag{2.11}$$

which coincides with (5.1) in Tsai and Crowley (1985). Therefore, $S_X^{(n)}(t)$ is self-consistent and a maximum likelihood estimator for $S_X(t)$. For the algebraic equation (2.11), the existence and uniqueness of the solution $S_X^{(n)}(t)$ can be established by using the convexity of the log-likelihood function [Turnbull (1974)]. The solution $S_X^{(n)}(t)$ can be calculated numerically by using the EM algorithm [Turnbull (1974) and Tsai and Crowley (1985)] or by the Newton–Raphson method to find the maximum point of the log-likelihood function.

To study the consistency of the estimator $S_X^{(n)}(t)$, it is necessary to examine whether system (2.5) determines $S_X(t)$ uniquely. The relationship between the two systems (2.5) and (2.7) will be explored in Sections 3 and 4.

It is worthwhile to note that if $S_Z \equiv 0$ in (2.5), then X is subject to right censoring only. Simple calculations yield the unique solution of (2.5) with initial conditions $S_X(0) = S_Y(0) = 1$ given by

$$S_X(t) = \exp\left(\int_0^t \frac{dQ_1(u)}{Q_1(u) + Q_2(u)}\right)$$

and

$$S_Y(t) = \exp\left(\int_0^t \frac{dQ_2(u)}{Q_1(u) + Q_2(u)}\right).$$

The corresponding sample equations of (2.7) for right censored data are

$$Q_1^{(n)}(t) = -\int_t^\infty S_Y^{(n)}(u)\, dS_X^{(n)}(u), \tag{2.12}$$
$$Q_2^{(n)}(t) = -\int_t^\infty S_X^{(n)}(u)\, dS_Y^{(n)}(u).$$

From the first equation in (2.12), we obtain

$$Q_1^{(n)}(t-) - Q_1^{(n)}(t) = S_Y^{(n)}(t-)\big(S_X^{(n)}(t-) - S_X^{(n)}(t)\big). \tag{2.13}$$

Adding the two equations in (2.12), we get

$$S_X^{(n)}(t) S_Y^{(n)}(t) = Q_1^{(n)}(t) + Q_2^{(n)}(t) \quad \text{for all finite } t \geq 0.$$

Dividing (2.13) by $S_X^{(n)}(t-) S_Y^{(n)}(t-) = Q_1^{(n)}(t-) + Q_2^{(n)}(t-)$ yields

$$\frac{S_X^{(n)}(t-) - S_X^{(n)}(t)}{S_X^{(n)}(t-)} = \frac{Q_1^{(n)}(t-) - Q_1^{(n)}(t)}{Q_1^{(n)}(t-) + Q_2^{(n)}(t-)}.$$

It follows that

$$S_X^{(n)}(t) = \prod_{s \leq t}\left[1 - \frac{Q_1^{(n)}(s-) - Q_1^{(n)}(s)}{Q_1^{(n)}(s-) + Q_2^{(n)}(s-)}\right],$$

where s's are jump points of $Q_1^{(n)}(\cdot)$. This is the product-limit estimator for S_X in the right censoring case.

3. Identifiability. The problem of identifiability is to determine under what conditions system (2.5) has a unique solution for $S_X(\cdot)$, given the Q_j's.

We give an example to illustrate that, without additional assumption, $S_X(\cdot)$ is not identifiable. Suppose that $P(1 < Z < 3) = 0$ and $Y = Z + 1$. Let $X^{(1)}$ and $X^{(2)}$ be two random variables whose survival functions coincide on intervals $[0, 2]$ and $[3, \infty]$, but differ on $(2, 3)$. We further assume that $X^{(i)}$, Y and Z, $i = 1, 2$, satisfy Assumption A. By using the system (2.5), it is easy to verify that $X^{(i)}$, Y and Z, $i = 1, 2$, produce the same subsurvival functions Q_j, $j = 1, 2, 3$.

In the example, we make certain the observation window $[Z, Y]$ never covers any part of the interval $(2, 3)$. Thus, there is no chance of observing the actual value of X if in fact $X \in (2, 3)$ and, consequently, we cannot determine $S_X(t)$, for $t \in (2, 3)$. To identify $S_X(\cdot)$, we, therefore, require the observation window

$[Z, Y]$ to cover each t in $(0, \infty)$ with a positive probability. This condition will be called

ASSUMPTION B.

$$0 < P(Z < t < Y) = S_Y(t) - S_Z(t) \quad \text{for every } t \in (0, \infty). \tag{3.1}$$

Assumption B requires that the right censoring variable Y is strictly stochastically larger than the left censoring variable Z. The case where $P[Z = Y] = 1$ is studied in Ayer, Brunk, Ewing and Reid (1955).

We shall show that Assumptions A and B are sufficient conditions for identifying S_X. These conditions are conceivably easy to check in practice.

Let $\mathscr{D}$ be the class of nonincreasing functions defined on $(0, \infty)$ with values in $[0, 1]$. In $\mathscr{D}$ we look for solutions S_i, $i = 1, 2, 3$, to the system (2.5).

LEMMA 3.1. *Suppose that the random vector* (X, Y, Z) *satisfies Assumptions A and B. Let* $S_i(t)$, *for* $i = 1, 2, 3$, *belong to* $\mathscr{D}$. *If* S_i, $i = 1, 2, 3$, *satisfy the system* (2.5) *for the* Q_j*'s induced by* (X, Y, Z), *then* $S_2(0+) = 1$, $S_3(\infty -) = 0$, $(1 - S_1(0+))(1 - S_3(0+)) = 0$ *and* $S_1(\infty -)S_2(\infty -) = 0$.

PROOF. Since S_1, S_2 and S_3 satisfy (2.5), summing over three equations in (2.5) yields

$$Q(t) = -S_1S_2|_t^\infty + S_1S_3|_t^\infty - S_3(\infty) + S_3(t).$$

We do not know the values of $S_i(\infty)$, $i = 1, 2, 3$, so (2.6) cannot be applied to the S_i's. However, we can evaluate the difference of Q at t_1 and t_2, which is

$$Q(t_1) - Q(t_2) = [S_3 + S_1(S_2 - S_3)](t_1) - [S_3 + S_1(S_2 - S_3)](t_2). \tag{3.2}$$

Suppose

$$\lim_{t \to \infty} S_i(t) = \alpha_i$$

and

$$\lim_{t \to 0+} S_i(t) = \beta_i, \qquad i = 1, 2, 3.$$

Under Assumptions B and A3, it is clear, from the first equation of (2.5), that

$$0 < Q_1(t) < Q_1(0) \quad \text{for any } t \in (0, \infty). \tag{3.3}$$

We shall prove that $\alpha_2 \ge \alpha_3$ and $\beta_2 \ge \beta_3$ by contradiction. If $\beta_2 < \beta_3$, then there exists a $\delta > 0$ and an $\varepsilon > 0$ such that

$$S_2(t) - S_3(t) < -\varepsilon \quad \text{for all } t \in (0, \delta).$$

From the first equation of (2.5), for any $t \in (0, \delta)$,

$$0 < Q_1(0) - Q_1(t) = -\int_0^t (S_2 - S_3)\, dS_1 \le \varepsilon \int_0^t dS_1 = -\varepsilon(S_1(0+) - S_1(t)) \le 0,$$

which is impossible. Similarly we can prove $\alpha_2 \ge \alpha_3$.

Letting $t_1 \to 0+$ and $t_2 \to \infty$ in (3.2), we get

$$1 = [\beta_3 + \beta_1(\beta_2 - \beta_3)] - [\alpha_3 + \alpha_1(\alpha_2 - \alpha_3)].$$

Observe that

$$\alpha_3 + \alpha_1(\alpha_2 - \alpha_3) \geq \alpha_3 \geq 0 \quad \text{and} \quad \beta_3 + \beta_1(\beta_2 - \beta_3) \leq \beta_2 \leq 1.$$

Consequently,

$$\alpha_3 = 0, \qquad \alpha_1\alpha_2 = 0, \qquad \beta_2 = 1 \quad \text{and} \quad (1 - \beta_1)(1 - \beta_3) = 0. \qquad \square$$

THEOREM 3.2. *Let the conditions in Lemma* 3.1 *be satisfied. Assume that* S_2 *and* S_3 *are continuous functions on* $(0, \infty)$. *Then* $S_X = S_1$, $S_Y = S_2$ *and* $S_Z = S_3$ *on* $(0, \infty)$.

Note that the continuity condition is not assumed for S_1 in this theorem.

PROOF. Since $\{S_X, S_Y, S_Z\}$ and $\{S_1, S_2, S_3\}$ satisfy (2.5) for the same Q_j's, it follows by subtraction that

$$(3.4) \quad 0 = -\int_t^\infty (S_Y - S_Z)\,d(S_X - S_1) - \int_t^\infty [(S_Y - S_2) - (S_Z - S_3)]\,dS_1,$$

$$(3.5) \quad 0 = -\int_t^\infty S_X d(S_Y - S_2) - \int_t^\infty (S_X - S_1)\,dS_2,$$

$$(3.6) \quad 0 = -\int_t^\infty (1 - S_X)\,d(S_Z - S_3) + \int_t^\infty (S_X - S_1)\,dS_3.$$

For easy reference, we shall call the following argument

DERIVATION C. By Assumption B (Assumption A) $S_Y - S_Z > 0$ ($S_X > 0, 1 - S_X > 0$) on $(0, \infty)$. If $(S_Y - S_2) - (S_Z - S_3) \geq 0$ ($S_X - S_1 \leq 0$) on (t', t''), then from (3.4)–(3.6), we obtain $d(S_X - S_1) \geq 0$ ($d(S_Y - S_2) \leq 0$, $d(S_Z - S_3) \geq 0$) on (t', t'').

Note that S_X and S_1 belong to $\mathscr{D}$. If $S_X \not\equiv S_1$, then there are three possible cases.

CASE 1. There exist t_1 and t_2 with $0 \leq t_1 < t_2 \leq \infty$ such that $S_X(t) < S_1(t+)$ for $t \in (t_1, t_2)$, $S_X(t_1) = S_1(t_1+)$ and $S_X(t_2) \geq S_1(t_2+)$, where $S_X(\infty) = 0$ and $S_1(\infty+) = S_1(\infty)$.

CASE 2. There exists a t_1, $0 \leq t_1 < \infty$, such that $S_X(t) < S_1(t+)$ on (t_1, ∞), $S_X(t_1) = S_1(t_1+)$ and $S_1(\infty) > 0$.

CASE 3. $S_X \geq S_1$ on $(0, \infty)$.

We show that each case leads to a contradiction.

CONTRADICTION OF CASE 1. If $(S_Y - S_2) - (S_Z - S_3)$ keeps the same sign on (t_1, t_2), positive, say, then Derivation C implies that

$$d(S_X - S_1) \geq 0 \quad \text{on } (t_1, t_2).$$

But then

$$S_X(t) - S_1(t+) = \int_{(t_1, t+]} d(S_X - S_1) \geq 0 \quad \text{for any } t \in (t_1, t_2).$$

This contradicts the assumption that $S_X < S_1$ on (t_1, t_2). If $(S_Y - S_2) - (S_Z - S_3)$ is negative on (t_1, t_2), then we can find a $\delta > 0$ such that $(S_Y - S_2) - (S_Z - S_3)$ is negative on $(t_1, t_2 + \delta)$ due to the continuity of S_Y, S_Z, S_2 and S_3. Derivation C implies that $d(S_X - S_1) \leq 0$ on $(t_1, t_2 + \delta)$. Then

$$S_X(t) - S_1(t+) = S_X(t_2) - S_1(t_2+) - \int_{(t, t_2+]} d(S_X - S_1) \geq 0$$

$$\text{for any } t \in (t_1, t_2).$$

This again contradicts with the assumption that $S_X < S_1$ on (t_1, t_2). Therefore, there exists a $t^* \in (t_1, t_2)$ such that

$$(S_Y(t^*) - S_2(t^*)) - (S_Z(t^*) - S_3(t^*)) = 0.$$

From Derivation C and the assumption that $S_X < S_1$ on (t_1, t_2), we conclude that

$$d(S_Y - S_2) \leq 0$$

and

$$d(S_Z - S_3) \geq 0 \quad \text{on } (t_1, t_2).$$

Therefore,

$$\begin{aligned}(S_Y(t) - S_2(t)) - (S_Z(t) - S_3(t)) \\ = -\int_t^{t^*} [d(S_Y - S_2) - d(S_Z - S_3)] \geq 0 \quad \text{on } (t_1, t^*).\end{aligned}$$

Again from Derivation C,

$$d(S_X - S_1) \geq 0 \quad \text{on } (t_1, t^*)$$

and, hence,

$$S_X(t) - S_1(t+) = \int_{(t_1, t+]} d(S_X - S_1) \geq 0 \quad \text{for any } t \in (t_1, t^*).$$

This contradicts the assumption that $S_X - S_1 < 0$ on (t_1, t_2).

CONTRADICTION OF CASE 2. From Lemma 3.1 and $S_1(\infty -) > 0$, we have $S_3(\infty -) = 0$ and $S_2(\infty -) = 0$. Since $S_X < S_1$ on (t_1, ∞), by Derivation C,

$$d(S_Y - S_2) \leq 0$$

and

$$d(S_Z - S_3) \geq 0 \quad \text{on } (t_1, \infty).$$

It follows that

$$\begin{aligned}(S_Y(t) - S_2(t)) - (S_Z(t) - S_3(t)) \\ = -\int_t^{\infty} [d(S_Y - S_2) - d(S_Z - S_3)] \geq 0 \quad \text{on } (t_1, \infty).\end{aligned}$$

Another application of Derivation C gives

$$d(S_X - S_1) \geq 0 \quad \text{on } (t_1, \infty).$$

As a result,

$$S_1(t+) - S_1(\infty -) = -\int_{(t,\infty)} dS_1 \geq -\int_t^\infty dS_X = S_X(t) \quad \text{on } (t_1, \infty).$$

Letting $t \downarrow t_1$, we have

$$-S_1(\infty -) \geq 0.$$

This is a contradiction.

CONTRADICTION OF CASE 3. The inequality $S_X \geq S_1$ on $(0, \infty)$ implies that $S_1(\infty -) = 0$. An argument similar to that used in Case 1 or in Case 2 would lead to a contradiction.

Therefore, $S_X \equiv S_1$ on $(0, \infty)$. To show $S_Y = S_2$ and $S_Z = S_3$ we note, by (2.5),

$$S_2(t) = 1 + \int_0^t \frac{dQ_2}{S_1} = 1 + \int_0^t \frac{dQ_2}{S_X} = S_Y(t)$$

and

$$S_3(t) = -\int_t^\infty \frac{dQ_3}{1 - S_1} = -\int_t^\infty \frac{dQ_3}{1 - S_X} = S_Z(t).$$

Here we have used the conditions $S_2(0+) = 1$ and $S_3(\infty -) = 0$, which are proved in Lemma 3.1. □

COROLLARY 3.3 (Identifiability). *Let (X_1, Y_1, Z_1) be a random vector that satisfies Assumptions A and B and (X_2, Y_2, Z_2) be another random vector that satisfies Assumption A. If both random vectors produce the same subsurvival functions Q_j, then $S_{X_1} = S_{X_2}$, $S_{Y_1} = S_{Y_2}$ and $S_{Z_1} = S_{Z_2}$ on $[0, \infty)$.*

4. Strong consistency. In this section, we assume that Assumptions A and B are satisfied. We shall prove that the estimator $S_X^{(n)}(t)$ is uniformly strongly consistent in the sense that

$$P\left[\sup_{t \in [0,\infty)} \left| S_X^{(n)}(t) - S_X(t) \right| \to 0, \text{as } n \to \infty \right] = 1. \tag{4.1}$$

The same holds for $S_Y^{(n)}(t)$ and $S_Z^{(n)}(t)$.

We shall use the fact that

$$\lim_{n \to \infty} Q_i^{(n)}(t) = Q_i(t) \quad \text{for } i = 1, 2, 3, \tag{4.2}$$

uniformly for $t \in [0, \infty)$ with probability 1 as ensured by the Glivenko–Cantelli theorem. In this section, we work with a fixed $\omega \in A$, where A is a measurable subset of the underlying sample space, $P(A) = 1$ and (4.2) holds for all $\omega \in A$. The dependence of $S_X^{(n)}(t)$ and other related functions on ω will not be indicated explicitly.

The main idea in proving consistency (4.1) is as follows. The sequences $\{S_X^{(n)}(t)\}$, $\{S_Y^{(n)}(t)\}$ and $\{S_Z^{(n)}(t)\}$ are uniformly bounded and nonincreasing functions. By Helly's theorem, for any subsequence of $\{S_X^{(n)}(t), S_Y^{(n)}(t), S_Z^{(n)}(t)\}$, we can select a further subsequence, indexed by $\{n_k\}$, $\{S_X^{(n_k)}(t), S_Y^{(n_k)}(t), S_Z^{(n_k)}(t)\}$ such that, for $t \in [0, \infty)$,

$$\lim_{k\to\infty} S_X^{(n_k)}(t) = S_X^0(t), \tag{4.3}$$

$$\lim_{k\to\infty} S_Y^{(n_k)}(t) = S_Y^0(t), \tag{4.4}$$

$$\lim_{k\to\infty} S_Z^{(n_k)}(t) = S_Z^0(t). \tag{4.5}$$

We then prove that the cluster point $\{S_X^0, S_Y^0, S_Z^0\}$ is actually the limit of the original sequence $\{S_X^{(n)}, S_Y^{(n)}, S_Z^{(n)}\}$. The proof involves showing the limit $\{S_X^0, S_Y^0, S_Z^0\}$ satisfies the system of integral equations (2.5) and the continuity condition (Lemma 4.1). According to Theorem 3.2, system (2.5) has a unique solution, $\{S_X, S_Y, S_Z\}$. Hence, we conclude in Theorem 4.2 that

$$S_X^0 \equiv S_X, \qquad S_Y^0 \equiv S_Y, \qquad S_Z^0 \equiv S_Z$$

and $\{S_X^{(n)}, S_Y^{(n)}, S_Z^{(n)}\}$ converges to $\{S_X, S_Y, S_Z\}$.

LEMMA 4.1. *S_X^0, S_Y^0 and S_Z^0 belong to class $\mathscr{D}$ and satisfy the system* (2.5). *Moreover, S_Y^0 and S_Z^0 are continuous functions on* $(0, \infty)$.

PROOF. By the self-consistent estimator criterion [Turnbull (1974)], the estimator $S_X^{(n)}$ is a nonincreasing function with values in $[0, 1]$. This implies that $S_X^0 \in \mathscr{D}$.

Following from the structure of the self-consistent estimator again, we have

$$S_X^{(n)}(t) \ge Q_1^{(n)}(t) \tag{4.6}$$

and

$$1 - S_X^{(n)}(t) \ge Q_1^{(n)}(0) - Q_1^{(n)}(t) \quad \text{for } t \in (0, \infty). \tag{4.7}$$

The equality (2.8) gives that, for t_1 and t_2, $0 < t_1 < t_2$,

$$0 \le S_Y^{(n)}(t_1) - S_Y^{(n)}(t_2) = -\int_{t_1}^{t_2} \frac{dQ_2^{(n)}}{S_X^{(n)}} \le \frac{Q_2^{(n)}(t_1) - Q_2^{(n)}(t_2)}{Q_1^{(n)}(t_2)}.$$

Letting $n_k \to \infty$, we obtain

$$0 \le S_Y^0(t_1) - S_Y^0(t_2) \le \frac{Q_2(t_1) - Q_2(t_2)}{Q_1(t_2)}.$$

The continuity of S_Y^0 on $(0, \infty)$ follows from the continuity of Q_2 and clearly S_Y^0 is nonincreasing. The continuity and the nonincreasing property of S_Z^0 can be proved in a similar way. □

Since S_Y^0 and S_Z^0 are continuous, $S_Y^{(n_k)}$ and $S_Z^{(n_k)}$ converge, respectively, to S_Y^0 and S_Z^0 uniformly on any closed interval $[t', t''] \subset (0, \infty)$. Passing to the limit as

$n_k \to \infty$ in

$$Q_1^{(n_k)}(t_1) - Q_1^{(n_k)}(t_2) = -\int_{t_1}^{t_2} \left(S_Y^{(n_k)} - S_Z^{(n_k)}\right) dS_X^{(n_k)}$$

yields

$$Q_1(t_1) - Q_1(t_2) = -\int_{t_1}^{t_2} \left(S_Y^0 - S_Z^0\right) dS_X^0. \tag{4.8}$$

As $t_2 \to \infty$ in (4.8), we see that S_X^0, S_Y^0 and S_Z^0 satisfy the first equation in (2.5). Similarly, we can prove that S_X^0, S_Y^0 and S_Z^0 satisfy the second and the third equations in (2.5).

According to (2.8) and (2.9), $S_Y^{(n)}$ and $S_Z^{(n)}$ are nonincreasing, $S_Y^{(n)} \le 1$ and $S_Z^{(n)} \ge 0$, which imply that S_Y^0 and S_Z^0 are nonincreasing, $S_Y^0 \le 1$ and $S_Z^0 \ge 0$. It remains to show that $S_Y^0 \ge 0$ and $S_Z^0 \le 1$. The first equation in (2.5) is

$$Q_1(t) = -\int_t^{\infty} \left(S_Y^0 - S_Z^0\right) dS_X^0,$$

where $0 < Q_1 < 1$ for $t \in (0, \infty)$ as established in (3.3). Consequently,

$$S_Y^0(0+) - S_Z^0(0+) \ge 0$$

and

$$S_Y^0(\infty-) - S_Z^0(\infty-) \ge 0.$$

Therefore,

$$1 \ge S_Y^0(t) \ge S_Y^0(\infty) \ge S_Z^0(\infty) \ge 0$$

and

$$1 \ge S_Y^0(0+) \ge S_Z^0(0+) \ge S_Z^0(t) \ge 0.$$

Thus S_Y^0 and S_Z^0 belong to class $\mathscr{D}$.

THEOREM 4.2 (Strong consistency). *If (X, Y, Z) satisfies Assumptions A and B, then*

$$\lim_{n\to\infty} S_X^{(n)}(t) = S_X(t), \tag{4.9}$$

$$\lim_{n\to\infty} S_Y^{(n)}(t) = S_Y(t), \tag{4.10}$$

$$\lim_{n\to\infty} S_Z^{(n)}(t) = S_Z(t) \tag{4.11}$$

uniformly for $t \in [0, \infty)$ with probability 1.

PROOF. According to Lemma 4.1 and Theorem 3.2,

$$S_X^0 = S_X, \qquad S_Y^0 = S_Y \quad \text{and} \quad S_Z^0 = S_Z \quad \text{on } (0, \infty).$$

This proves that (4.9)–(4.11) hold for every fixed t in $(0, \infty)$ almost surely. Since S_X, S_Y and S_Z are continuous, nonincreasing on $[0, \infty)$, $S_X(0) = S_Y(0) = S_Z(0) = 1$, $S_X(\infty) = S_Y(\infty) = S_Z(\infty) = 0$ and $S_X^{(n)}$, $S_Y^{(n)}$ and $S_Z^{(n)}$ are nonincreasing with values between 0 and 1, (4.9)–(4.11) hold uniformly for $t \in [0, \infty)$ with probability 1. □

Acknowledgments. We wish to thank P. J. Smith and C. Z. Wei for helpful discussions. We appreciate the comments from the referees which helped to shorten the proofs of this paper.

REFERENCES

AYER, M., BRUNK, H. D., EWING, G. M. and REID, W. T. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.* **26** 641–647.

EFRON, B. (1967). The two-sample problem with censored data. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **4** 831–853. Univ. California Press.

GEHAN, E. A. (1965). A generalized two-sample Wilcoxon test for doubly censored data. *Biometrika* **52** 650–653.

LEIDERMAN, P. H., BABU, D., KAGIA, J., KRAEMER, H. C. and LEIDERMAN, G. F. (1973). African infant precocity and some social influences during the first year. *Nature* **242** 247–249.

MANTEL, N. (1967). Ranking procedures for arbitrarily restricted observations. *Biometrics* **23** 65–78.

PETERSON, A. V., JR. (1977). Expressing the Kaplan–Meier estimator as a function of empirical subsurvival functions. *J. Amer. Statist. Assoc.* **72** 854–88.

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

TSAI, W. Y. and CROWLEY, J. (1985). A large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency. *Ann. Statist.* **13** 1317–1334.

TSIATIS, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proc. Nat. Acad. Sci. U.S.A.* **72** 20–22.

TURNBULL, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.* **69** 169–173.

DEPARTMENT OF STATISTICS
UNIVERSITY OF FLORIDA
GAINESVILLE, FLORIDA 32611

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND 20742