

# ALGORITHMS FOR COMPUTING SELF-CONSISTENT AND MAXIMUM LIKELIHOOD ESTIMATORS WITH DOUBLY CENSORED DATA

BY PER A. MYKLAND<sup>1</sup> AND JIAN-JIAN REN<sup>2</sup>

*University of Chicago and University of Nebraska*

The paper investigates the structure of the self-consistent estimators (SCE) and the nonparametric maximum likelihood estimator (NPMLE) for doubly censored data. An explicit sufficient and necessary condition for an SCE to be the NPMLE is given. Based on this, algorithms for computing the SCE and the NPMLE are provided. The relation between our algorithms and the EM algorithm is studied.

**1. Introduction.** In medical follow-up studies and in biometry and reliability studies, incomplete data are frequently encountered. Examples of the lifetime samples' being censored either from above or below, called *doubly censored samples*, have been given by Gehan (1965), Mantel (1967), Peto (1973), Turnbull (1974) and others. For doubly censored samples, Turnbull (1974) and Tsai and Crowley (1985) have given the *self-consistent estimators* (SCE) for the survival function  $S_X$  with grouped data and ungrouped data, respectively. The asymptotic properties of the self-consistent estimators, such as strong consistency and weak convergence, have been studied by Chang and Yang (1987), Chang (1990) and Gu and Zhang (1993).

However, a satisfying procedure to find these self-consistent estimators and the *nonparametric maximum likelihood estimator* (NPMLE) of  $S_X$  for doubly censored data is not available. For grouped data, Turnbull (1974) gave an algorithm to compute the SCE and showed that in his case, the SCE is the NPMLE. However, Turnbull's method has its limitations, because a natural discrete time scale, which is used in Turnbull's method, may not exist for some data. For instance, an example occurred in a recent study of age-dependent growth rate of primary breast cancer by Peer, Van Dijck, Hendricks, Holland and Verbeek (1993), where the doubly censored data are clearly not "grouped." For ungrouped data, the problem is more complicated. In such a case, it is shown that an SCE is not unique and is not necessarily the NPMLE [see examples given by Gu and Zhang (1993)], and it is not clear how the SCE are related to the NPMLE. All these situations can also happen in the grouped data case if there are groups that contain censored observations only.

---

Received June 1994; revised August 1995.

<sup>1</sup> Research partially supported by NSF Grants DMS-93-05601 and DMS-96-26266, and ARO Grant DAAH04-95-1-0105.

<sup>2</sup> Research partially supported by NSF Grants DMS-95-10376 and DMS-96-26532.

AMS 1991 subject classifications. Primary 62G05, 62G30, 65U05.

Key words and phrases. EM algorithm, fixed point problem, survival function.

In this paper, we give a simple algorithm, which is shown in Theorem 6 to be the EM algorithm with restrictions on the initial point, to compute a particular SCE for ungrouped doubly censored data. Our Theorem 6 and the principle of the EM algorithm imply that SCE yield the local maxima of the “likelihood function” (2.1). Our Theorem 1 shows that the global maximum of the likelihood function (2.1)—the NPMLE—exists uniquely for ungrouped doubly censored data. In Theorem 2, we give a set of sufficient and necessary conditions to characterize the relation between the SCE and the NPMLE, which shows that the NPMLE is an SCE for ungrouped doubly censored data. Based on these conditions, a simple algorithm for searching for the NPMLE is given.

Theorem 3 by Gu and Zhang (1993) along with our Theorem 2 implies that the SCE and NPMLE are asymptotically equivalent. This means that searching for the NPMLE for a large sample may not be necessary. However, since SCE are the local maxima of the likelihood function, the difference between SCE and the NPMLE may not be negligible for a sample with moderate size. We present some simulation results at the end of Section 2 to illustrate this point. The simulation results suggest that the NPMLE is a better estimator than an arbitrary SCE for a moderate sample size.

Another aspect of the SCE and the NPMLE for doubly censored data that is worth investigating is the structure of the estimators. Since the SCE are given implicitly through some estimating equations, the structure of the estimators is not clear. Based on the SCE for doubly censored data, some statistical inference problems, such as estimation, hypothesis testing and density estimation, have been studied by Ren and Zhou (1993) and Ren (1994, 1995). In these studies, the structure of the SCE plays a very important role. For right censored data, Efron (1967) gave the clear structure of the SCE (NPMLE). In this paper, we give the version (of a particular SCE) for doubly censored data. We also give the structure of the NPMLE for doubly censored data in some special situations.

We present our main results in Section 2, and give the proofs in Sections 3 and 4.

**2. Nonparametric MLE.** Let  $X$  be a random variable (r.v.) with a survival function  $S_X$  and let  $X_i, i = 1, \dots, n$ , be  $n$  independent observations on  $X$ . In this research, one observes not  $\{X_i\}$ , but a doubly censored sample,

$$V_i = \max\{\min\{X_i, Y_i\}, Z_i\}$$

with the index

$$\delta_i = \begin{cases} 1, & \text{if } Z_i < X_i \leq Y_i, \\ 2, & \text{if } V_i = Y_i < X_i \text{ (right censoring),} \\ 3, & \text{if } V_i = Z_i \geq X_i \text{ (left censoring),} \end{cases}$$

where  $(Y_i, Z_i), i = 1, \dots, n$ , are independent from  $X_i$  and are independent observations on  $(Y, Z)$  for random variables  $Y$  and  $Z$  with  $P\{Y \geq Z\} = 1$ . The

r.v.'s  $Y_i$  and  $Z_i$  are called *right* and *left censoring variables*, respectively. We study the nonparametric maximum likelihood estimator (NPMLE) of  $S_X$  based on  $(V_i, \delta_i)$ ,  $1 \leq i \leq n$ .

Let  $W_1, W_2, \dots, W_m$  be all the distinct points of the set  $\{V_1, \dots, V_n\}$  with  $W_1 < \dots < W_m$  and let

$$\begin{aligned}\alpha_k &= \sum_{i=1}^n I\{\delta_i = 1, V_i = W_k\}, \\ \beta_k &= \sum_{i=1}^n I\{\delta_i = 2, V_i = W_k\}, \\ \gamma_k &= \sum_{i=1}^n I\{\delta_i = 3, V_i = W_k\},\end{aligned}$$

where  $k = 1, \dots, m$ . Suppose that each  $(V_i, \delta_i)$  is identically distributed as  $(V, \delta)$ . We note that

$$\begin{aligned}P\{V = v, \delta = 1\} &= P\{X = v, Z < X \leq Y\} = P\{X = v\}P\{Z < v \leq Y\}, \\ P\{V = v, \delta = 2\} &= P\{Y = v, X > Y\} = P\{X > v\}P\{Y = v\}, \\ P\{V = v, \delta = 3\} &= P\{Z = v, X \leq Z\} = P\{X \leq v\}P\{Z = v\}, \\ \prod_{i=1}^n P\{V = V_i, \delta = \delta_i\} &= C \prod_{i=1}^n (P\{X = V_i\})^{I\{\delta_i=1\}} (P\{X > V_i\})^{I\{\delta_i=2\}} \\ &\quad \times (P\{X \leq V_i\})^{I\{\delta_i=3\}},\end{aligned}$$

where  $C$  is the term that depends only on  $(Y, Z)$ . Hence, if a possible candidate  $S$  for the NPMLE of  $S_X$  assigns all its mass to the points  $W_1, \dots, W_m$  and the interval  $(W_m, \infty)$ , we know that the likelihood function of  $(V_i, \delta_i)$ ,  $1 \leq i \leq n$ , for  $S$  is proportional to the function

$$(2.1) \quad L(S) = \prod_{k=1}^m (P_{k-1} - P_k)^{\alpha_k} P_k^{\beta_k} (1 - P_k)^{\gamma_k},$$

where  $P_0 = 1$  and  $P_k = P_S\{X > W_k\}$ . The NPMLE  $\hat{S}_X^{(n)}$  of  $S_X$  is the estimator  $\hat{P}_k = \hat{S}_X^{(n)}(W_k)$ ,  $1 \leq k \leq m$ , which maximizes  $L(S)$ . With the proof deferred to the Appendix, the following theorem shows that the NPMLE is unique.

**THEOREM 1.** *The likelihood, or log  $L(S)$ , is strictly concave on the set of  $P_k$ 's satisfying:*

- (i)  $1 \geq P_1 \geq \dots \geq P_m \geq 0$ .
- (ii)  $P_{k-1} > P_k$  if  $\alpha_k > 0$ .
- (iii)  $P_k > 0$  if  $\beta_k > 0$ .
- (iv)  $P_k < 1$  if  $\gamma_k > 0$ .

One may note that Theorem 1 considers a more general case than the one studied by Turnbull (1974), who required  $\alpha_k > 0$  for all  $1 \leq k \leq m$ . However, it is not easy to directly study the structure of the NPMLE through the likelihood  $L(\cdot)$ . In this paper, we give an algorithm to compute the NPMLE, from which the structure of the estimator follows in some situations.

By Tsai and Crowley (1985), a self-consistent estimator  $S_X^{(n)}$  of  $S_X$  is given by a solution of the equation

$$(2.2) \quad S_X^{(n)}(t) = Q^{(n)}(t) - \int_{u \leq t} \frac{S_X^{(n)}(t)}{S_X^{(n)}(u)} dQ_2^{(n)}(u) + \int_{t < u} \frac{1 - S_X^{(n)}(t)}{1 - S_X^{(n)}(u)} dQ_3^{(n)}(u),$$

where

$$Q_j^{(n)}(t) = \frac{1}{n} \sum_{i=1}^n I\{V_i > t, \delta_i = j\}, \quad j = 1, 2, 3,$$

$$Q^{(n)}(t) = \sum_{j=1}^3 Q_j^{(n)}(t), \quad t \geq 0,$$

with  $\int_{u \leq t} = 0$  if  $S_X^{(n)}(t) = 0$  and  $\int_{t < u} = 0$  if  $S_X^{(n)}(t) = 1$ . (Even though we do not consider the asymptotic properties of the SCE in this paper, we keep  $n$  in  $S_X^{(n)}$  and  $Q^{(n)}$  for consistency with earlier authors.) As pointed out by Gu and Zhang (1993), a self-consistent estimate is not necessarily an NPMLE. In Turnbull's grouped data case (1974), the NPMLE of  $S_X$  is self-consistent. Later on, we show that the NPMLE for ungrouped data is also self-consistent.

Observing the nature of the integral equation (2.2), it suffices to find the value of  $S_X^{(n)}$  at each  $W_k$ ,  $1 \leq k \leq m$ , since a solution of (2.2) must be a step function with possible jumps at  $W_k$ . Denote  $x_k^* = S_X^{(n)}(W_k)$ . From (2.2), we must have

$$(2.3) \quad x_k^* = Q^{(n)}(W_k) + \frac{x_k^*}{n} \sum_{i=1}^n \frac{I\{\delta_i = 2, V_i \leq W_k\}}{x_i^*} - \frac{1 - x_k^*}{n} \sum_{i=1}^n \frac{I\{\delta_i = 3, V_i > W_k\}}{1 - x_i^*},$$

where  $k = 1, \dots, m$ , and  $x_i^* = x_j^*$  if  $V_i = W_j$ . Let

$$\Psi_k(\mathbf{x}) = Q^{(n)}(W_k) + \frac{x_k}{n} \sum_{i=1}^n \frac{I\{\delta_i = 2, V_i \leq W_k\}}{x_i} - \frac{1 - x_k}{n} \sum_{i=1}^n \frac{I\{\delta_i = 3, V_i > W_k\}}{1 - x_i}$$

and

$$\Psi = (\Psi_1, \dots, \Psi_m),$$

where  $\mathbf{x} \in E_m = [0, 1]^m$  and  $x_i = x_j$  if  $V_i = W_j$ . Then a self-consistent estimate  $\mathbf{x}^* = (x_1^*, \dots, x_m^*)$  is a fixed point of the function  $\Psi$ , that is,

$$(2.4) \quad \mathbf{x}^* = \Psi(\mathbf{x}^*).$$

We observe that a naive and simple algorithm for computing  $S_X^{(n)}$  can be given based on (2.4) as follows:

$$(2.5) \quad \mathbf{x}_{N+1} = \Psi(\mathbf{x}_N), \quad N = 1, 2, \dots,$$

where  $\mathbf{x}_1 \in E_m$ . Any limit point of  $\{\mathbf{x}_N\}$  should give a self-consistent estimator. This algorithm (2.5) is evidently quite appealing because it is simple and easy to understand. Naturally, a few questions need to be answered.

1. Does algorithm (2.5) converge?
2. Does algorithm (2.5) converge to the NPMLE?
3. How is algorithm (2.5) related to the EM algorithm?

In Theorem 6, we show that algorithm (2.5) is in fact the EM algorithm. From our simulation study, we find that the algorithm may converge to different points depending on the initial point  $\mathbf{x}_1$ . Hence, algorithm (2.5) does not always give the NPMLE, even when it converges.

To search for the NPMLE using algorithm (2.5), from the likelihood  $L(S)$  we observe that the following facts should be satisfied by the NPMLE:

1.  $\hat{S}_X^{(n)}(t) = 1$  for  $t < A_n$ .
2.  $\hat{S}_X^{(n)}(t) = 0$  for  $t \geq B_n$  when  $\beta_B = 0$ .
3.  $\hat{S}_X^{(n)}(t) = 0$  for  $t \geq C_n$  when  $B_n < W_m$  and  $\beta_B > 0$ .
4.  $\hat{S}_X^{(n)}(t) = c > 0$  for  $t \geq B_n$  when  $B_n = W_m$  and  $\beta_B > 0$ , where  $0 < c < 1$  is a constant,

$$\begin{aligned} A_n &= \min\{W_k; \alpha_k + \gamma_k > 0\}, \\ B_n &= \max\{W_k; \alpha_k + \beta_k > 0\}, \\ C_n &= \min\{W_k; W_k > B_n\} \end{aligned}$$

and

$$\begin{aligned} \alpha_L &= \sum_{i=1}^n I\{\delta_i = 1, V_i = L_n\}, \\ \beta_L &= \sum_{i=1}^n I\{\delta_i = 2, V_i = L_n\}, \\ \gamma_L &= \sum_{i=1}^n I\{\delta_i = 3, V_i = L_n\}, \end{aligned} \tag{2.6}$$

for  $L = A, B, C$ . We also observe that  $A_n$  is the first jump point of the NPMLE  $\hat{S}_X^{(n)}$  and that  $B_n$  or  $C_n$  is the last jump point of  $\hat{S}_X^{(n)}$  when  $\beta_B = 0$  or  $B_n < W_m$ . In Section 3, we will show that all these jumps sizes are no less than  $1/n$ . Based on these observations, we restrict the selection of the initial point  $\mathbf{x}_1$  for algorithm (2.5) on the space

$$\Omega_n = \{\mathbf{x} = (x_1, \dots, x_m) \in E_m | x_k \geq x_{k+1} \text{ and satisfy (A1)-(A4)}\},$$

where:

- (A1)  $x_k = 1$  for those  $W_k < A_n$ .
- (A2) When  $\beta_B = 0$ ,  $1/n \leq x_k \leq (n - 1)/n$  for those  $A_n \leq W_k < B_n$  and  $x_k = 0$  for those  $W_k \geq B_n$ .
- (A3) When  $B_n < W_m$  and  $\beta_B > 0$ ,  $1/n \leq x_k \leq (n - 1)/n$  for those  $A_n \leq W_k < C_n$  and  $x_k = 0$  for those  $W_k \geq C_n$ .

(A4) When  $B_n = W_m$  and  $\beta_B > 0$ ,  $1/n \leq x_k \leq (n - 1)/n$  for those  $A_n \leq W_k$ .

THEOREM 2. (i)  $\Psi$  has at least one fixed point  $\hat{\mathbf{x}}^*$  in space  $\Omega_n$ .

(ii)  $\hat{S}_X^{(n)}$  is the NPMLLE if and only if (1)  $\hat{S}_X^{(n)}$  is a fixed point of  $\Psi$  in  $\Omega_n$  and (2)  $\hat{S}_X^{(n)}$  satisfies the condition  $\Delta_k \geq 0$  for  $1 \leq k \leq m$ , where for  $\hat{P}_k = \hat{S}_X^{(n)}(W_k)$ ,

$$\Delta_k = 1 - \frac{1}{n} \left\{ \sum_{j=1}^{k-1} \frac{\beta_j}{\hat{P}_j} + \sum_{j=k}^m \frac{\gamma_j}{[1 - \hat{P}_j]} \right\}.$$

The proof of Theorem 2 is given in Section 3. Clearly, Theorem 2 implies that the NPMLLE is self-consistent and that from the structure of the space  $\Omega_n$ , the NPMLLE  $\hat{S}_X^{(n)}$  is a nonincreasing step function.

From the proof of Theorem 2, we find that if we impose the following additional condition on the initial point  $\mathbf{x}_1 \in \Omega_n$ , then the detailed structure of a particular self-consistent estimator  $\bar{S}_X^{(n)}$  can be obtained.

(A5) Choose  $x_{k-1} = x_k$  if  $\alpha_k = 0$  and  $A_n < W_k < J_n$ , where

$$J_n = \begin{cases} B_n & \text{if } \beta_B = 0, \\ C_n, & \text{if } B_n < W_m, \beta_B > 0, \\ W_m, & \text{if } B_n = W_m, \beta_B > 0. \end{cases}$$

We denote  $\Omega_n^+ = \{\mathbf{x} \in \Omega_n | \mathbf{x} \text{ satisfies (A5)}\}$  and state the results in the next corollary with the proof deferred to Section 3.

Let  $S_Y(t) = P\{Y > t\}$  and  $S_Z(t) = P\{Z > t\}$ . Chang and Yang (1987) gave the self-consistent estimators of  $S_Y$  and  $S_Z$  as

$$\begin{aligned} (2.7) \quad S_Y^{(n)}(t) &= 1 + \int_{u \leq t} \frac{dQ_2^{(n)}(u)}{S_X^{(n)}(u)} \\ &= 1 - \frac{1}{n} \sum_{i=1}^n \frac{I\{\delta_i = 2, V_i \leq t\}}{S_X^{(n)}(V_i)}, \quad t \leq D_n, \end{aligned}$$

$$\begin{aligned} (2.8) \quad S_Z^{(n)}(t) &= - \int_{t < u} \frac{dQ_3^{(n)}(u)}{1 - S_X^{(n)}(u)} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{I\{\delta_i = 3, V_i > t\}}{1 - S_X^{(n)}(V_i)}, \quad t \geq A_n, \end{aligned}$$

where  $S_X^{(n)}$  is a self-consistent estimator of  $S_X$  and  $D_n = \max\{W_k; S_X^{(n)}(W_k) > 0\}$ . They showed that  $S_Y^{(n)}$  and  $S_Z^{(n)}$  have strong consistency under regularity conditions.

COROLLARY 3. *There exists a unique self-consistent estimator  $\bar{S}_X^{(n)}$  of  $S_X$  in  $\Omega_n^+$ , and for the corresponding  $\bar{S}_Y^{(n)}$  and  $\bar{S}_Z^{(n)}$  given by (2.7) and (2.8), respectively,  $\bar{S}_X^{(n)}$  can be expressed as*

$$\begin{aligned}
 \bar{S}_X^{(n)}(t) = & 1 - \frac{(\alpha_A + \gamma_A)I\{A_n \leq t\}}{n[1 - \bar{S}_Z^{(n)}(A_n)]} \\
 & - \sum_{A_n < W_k < J_n} \frac{\alpha_k I\{W_k \leq t\}}{n[\bar{S}_Y^{(n)}(W_{k-1}) - \bar{S}_Z^{(n)}(W_{k-1})]} \\
 (2.9) \quad & - \frac{\alpha_B I\{\alpha_m + \beta_m = 0, \beta_B = 0, J_n \leq t\}}{n[\bar{S}_Y^{(n)}(J'_n) - \bar{S}_Z^{(n)}(J'_n)]} \\
 & - \frac{\beta_B I\{\alpha_m + \beta_m = 0, \beta_B > 0, J_n \leq t\}}{n[\bar{S}_Y^{(n)}(J''_n) - \bar{S}_Z^{(n)}(J''_n)]} \\
 & - \frac{\alpha_J I\{\alpha_m + \beta_m > 0, J_n \leq t\}}{n[\bar{S}_Y^{(n)}(J'_n) - \bar{S}_Z^{(n)}(J'_n)]},
 \end{aligned}$$

where  $J'_n = \max\{W_k; W_k < J_n\}$ ,  $J''_n = \max\{W_k; W_k < J'_n\}$  and  $\alpha_J, \beta_J, \gamma_J$  are given by (2.6) for  $L = J$ .

Equation (2.9) shows that except at  $A_n$  and  $J_n$ , the particular SCE  $\bar{S}_X^{(n)}$  only has jumps at those points with positive  $\alpha_k$ 's. This is quite useful in the study of statistical inference problems with doubly censored data. The strong consistency and weak convergence of any self-consistent estimator of  $S_X$  have been established by Chang and Yang (1987), Chang (1990) and Gu and Zhang (1993). Equation (2.9) provides a convenient tool to further derive other asymptotic results for doubly censored data, such as the asymptotic properties of  $L$ - and  $M$ -estimators [Ren and Zhou (1993)] for doubly censored data and the asymptotic properties of density estimation for doubly censored data [Ren (1994)]. For right censored data, Efron (1967) gave the clear structure of the self-consistent estimator (NPMLE). Our (2.9) is the version for doubly censored data. The computation method for obtaining (2.9) is given in our Theorem 6.

However, (2.9) is not necessarily the NPMLE. For example, with five observations (1, 1), (2, 2), (3, 3), (4, 3) and (5, 1) from  $(V, \delta)$ ,  $\bar{S}_X^{(n)}$  is given by  $\bar{P}_1 = \bar{P}_2 = \bar{P}_3 = \bar{P}_4 = 2/5$ ,  $\bar{P}_5 = 0$ , but the NPMLE is given by  $\hat{P}_1 = \hat{P}_2 = 1/2$ ,  $\hat{P}_3 = \hat{P}_4 = 1/3$ ,  $\hat{P}_5 = 0$ . Nevertheless, when data fit a certain pattern, (2.9) is the NPMLE.

We call a point  $W_k$  an  $\alpha$ -type point if  $\alpha_k > 0$  or  $W_k = A_n$ ; a  $\beta$ -type point if  $\alpha_k = 0$  with  $\beta_{k-1} = 0$  and  $W_1 < W_k \leq W_m$ ; a  $\gamma$ -type point if  $\alpha_k = 0$  with  $\gamma_k = 0$  and  $W_1 < W_k < W_m$ .

COROLLARY 4. *Suppose for any two adjacent  $\alpha$ -type points  $W_i$  and  $W_j$  with  $W_i < W_j$ , any point  $W_k$  between  $W_i$  and  $W_j$  is either a  $\beta$ -type point with only  $\beta$ -type points between  $W_i$  and  $W_k$  or a  $\gamma$ -type point with only  $\gamma$ -type points between  $W_k$  and  $W_j$ . Then (2.9) is the NPMLE.*

Clearly, if the data only contain  $\alpha$ -type points (which is close to Turnbull's case), (2.9) is the NPMLE. For a simpler case in which there are no ties among  $V_1, \dots, V_n$ , we can characterize those non- $\alpha$ -type jump points of the NPMLE clearly.

COROLLARY 5. *Suppose there are no ties among  $V_1, \dots, V_n$  and let  $V_k = W_k$ . We call  $W_k$  a 2-3-type point if  $\delta_k = 3$  and  $\delta_{k-1} = 2$  with  $A_n < W_k$ . Then:*

(i) *The only possible non- $\alpha$ -type jump points of the NPMLE are 2-3-type points.*

(ii) *If there are no 2-3-type points among non- $\alpha$ -type points, then (2.9) is the NPMLE.*

The proofs of Corollaries 4 and 5 are given in Section 3.

Our next theorem ensures that the algorithm (2.5) converges to a self-consistent estimator of  $S_X$  in  $\Omega_n$  when the initial point  $\mathbf{x}_1$  is selected in the space  $\Omega_n$ . The proof is given in Section 4.

THEOREM 6. *Let  $\Omega_0 = \{\mathbf{x} \in E_m; 1 \geq x_1 \geq x_2 \geq \dots \geq x_m \geq 0\}$ . Then:*

(i) *For  $\mathbf{x}_1 \in \Omega_0$ , algorithm (2.5) is the EM algorithm.*

(ii) *For  $\mathbf{x}_1 \in \Omega_n$ , all the limit points of  $\{\mathbf{x}_N\}$  from algorithm (2.5) are self-consistent estimators (in  $\Omega_n$ ) of  $S_X$ .*

(iii) *For  $\mathbf{x}_1 \in \Omega_n^+$ ,  $\{\mathbf{x}_N\}$  from algorithm (2.5) converges to the self-consistent estimator  $\bar{S}_X^{(n)}$  of  $S_X$  given by (2.9).*

In general, the EM algorithm does not do the job to find the NPMLE. As shown in Section 4, the algorithm can converge to any fixed point of  $\Psi$  and the fixed point of  $\Psi$  is not unique in  $\Omega_0$ . Our Theorems 2 and 6 provide a method to search for the NPMLE. The search consists of the following steps:

STEP (a) For a point  $\mathbf{x}_1 \in \Omega_n$  (see Remark 1 below on the choice of  $\mathbf{x}_1$ ), use algorithm (2.5) to find a self-consistent estimator  $S_X^{(n)}$ .

STEP (b) For  $S_X^{(n)}$ , check out condition (2) of Theorem 2.

STEP (c) If condition (2) of Theorem 2 is satisfied,  $S_X^{(n)}$  is the NPMLE; otherwise go to Step (a) with a different initial point in  $\Omega_n$ .

REMARK 1. In Step (a), one should avoid the condition (A5) for  $\mathbf{x}_1 \in \Omega_n$  if the data do not have the pattern described in Corollary 4. Also, when there



TABLE 1  
Distance between  $S_X$  and NPMLE and distance between  $S_X$  and SCE\*

|          |                                  | $X \sim \mathbf{Exp}(1), Y \sim \mathbf{Exp}(2)$<br>$Z = Y - 0.5$ | $X \sim N(0, 1), Y \sim N(0.5, 2)$<br>$Z = Y - 0.5$ |
|----------|----------------------------------|---|---|
| $n = 25$ | Average $\ S_X - \text{NPMLE}\ $ | 0.327588 (0.009189)   | 0.328770 (0.008158)                                 |
|          | Average $\ S_X - \text{SCE}\ $   | 0.367372 (0.010949)   | 0.373685 (0.011880)                                 |
| $n = 50$ | Average $\ S_X - \text{NPMLE}\ $ | 0.242737 (0.004200)   | 0.248826 (0.003884)                                 |
|          | Average $\ S_X - \text{SCE}\ $   | 0.269494 (0.006052)   | 0.279619 (0.006259)                                 |

\*  $\text{Exp}(\mu)$  is the exponential d.f. with mean  $\mu$ ;  $N(\mu, \sigma^2)$  is the normal d.f. with mean  $\mu$  and variance  $\sigma^2$ .

are no ties among  $V_1, \dots, V_n$ , one should avoid condition (A5) at those 2–3-type points; not doing so can slow down the algorithm to find the NPMLE. The reason is that  $\mathbf{x}_n \in \Omega_n^+$  implies  $\mathbf{x}_{N+1} \in \Omega_n^+$  (see discussions on this in the proof of Corollary 3 in Section 3).

REMARK 2. One may note that if the data have the pattern described in Corollary 4 or satisfy condition (ii) of Corollary 5, our Theorem 6(iii) directly finds the NPMLE.

*Simulation results.* Based on the algorithms for computing SCE and the NPMLE studied above, we conducted a small simulation study to compare SCE and the NPMLE with moderate sample sizes. In Table 1, for 1000 generated doubly censored samples, we find the NPMLE and an SCE which is not the NPMLE, and we compute the distance, in uniform norm, between  $S_X$  and NPMLE, and the distance between  $S_X$  and SCE, respectively. The simulation variances are displayed in the parentheses next to the simulation means in Table 1. The results indicate that for a sample with moderate size, there is a difference between the NPMLE and SCE and the NPMLE is better.

**3. Proofs of Theorem 2 and Corollaries.** Before proving Theorem 2, we first observe the following facts. Let  $\mathbf{y} = \Psi(\mathbf{x})$  for  $\mathbf{x} \in \Omega_n$ . Then, for each  $k$ ,

$$(3.1) \quad y_k = Q^{(n)}(W_k) + \frac{x_k}{n} \sum_{i=1}^n \frac{I\{\delta_i = 2, V_i \leq W_k\}}{x_i} - \frac{1 - x_k}{n} \sum_{i=1}^n \frac{I\{\delta_i = 3, V_i > W_k\}}{1 - x_i},$$

and moreover:

1.  $y_k = 1$  for those  $W_k < A_n$ .
2. When  $\beta_B = 0$ ,  $y_k = 0$  for those  $W_k \geq B_n$ .
3. When  $B_n < W_m$  and  $\beta_B > 0$ ,  $y_k = 0$  for those  $W_k \geq C_n$ .
4. For those  $A_n \leq W_{k-1} < W_k < B_n$  when  $\beta_B = 0$  or for those  $A_n \leq W_{k-1} < W_k < C_n$  when  $B_n < W_m$  and  $\beta_B > 0$  or for those  $A_n \leq W_{k-1} < W_k$  when  $B_n = W_m$  and  $\beta_B > 0$ , we have

$$\begin{aligned}
 y_{k-1} - y_k &= [Q^{(n)}(W_{k-1}) - Q^{(n)}(W_k)] \\
 &+ \left[ \frac{x_{k-1} - x_k}{n} \sum_{i=1}^n \frac{I\{\delta_i = 2, V_i \leq W_{k-1}\}}{x_i} \right. \\
 &\quad \left. - \frac{x_k}{n} \sum_{i=1}^n \frac{I\{\delta_i = 2, V_i = W_k\}}{x_i} \right] \\
 &- \left[ \frac{1 - x_{k-1}}{n} \sum_{i=1}^n \frac{I\{\delta_i = 3, V_i = W_k\}}{1 - x_i} \right. \\
 &\quad \left. - \frac{x_{k-1} - x_k}{n} \sum_{i=1}^n \frac{I\{\delta_i = 3, V_i > W_k\}}{1 - x_i} \right] \\
 &= [Q^{(n)}(W_{k-1}) - Q^{(n)}(W_k)] \\
 &+ \left[ \frac{x_{k-1} - x_k}{n} \sum_{i=1}^n \frac{I\{\delta_i = 2, V_i \leq W_{k-1}\}}{x_i} \right. \\
 &\quad \left. - \{Q_2^{(n)}(W_{k-1}) - Q_2^{(n)}(W_k)\} \right] \\
 &- \left[ \frac{1 - x_{k-1}}{1 - x_k} \{Q_3^{(n)}(W_{k-1}) - Q_3^{(n)}(W_k)\} \right. \\
 &\quad \left. - \frac{x_{k-1} - x_k}{n} \sum_{i=1}^n \frac{I\{\delta_i = 3, V_i > W_k\}}{1 - x_i} \right] \\
 &= [Q_1^{(n)}(W_{k-1}) - Q_1^{(n)}(W_k)] \\
 &+ \frac{x_{k-1} - x_k}{n} \sum_{i=1}^n \left\{ \frac{I\{\delta_i = 2, V_i \leq W_{k-1}\}}{x_i} \right. \\
 &\quad \left. + \frac{I\{\delta_i = 3, V_i \geq W_k\}}{1 - x_i} \right\},
 \end{aligned}
 \tag{3.2}$$

which implies  $y_{k-1} \geq y_k$ .

5. Denoting  $x_A$  and  $y_A$  as the components corresponding to  $A_n$  for  $\mathbf{x}$  and  $\mathbf{y}$ , respectively, by the definition of  $A_n$ , we have

$$\begin{aligned}
 y_A &= Q^{(n)}(A_n) + \frac{x_A}{n} \sum_{i=1}^n \frac{I\{\delta_i = 2, V_i \leq A_n\}}{x_i} \\
 &\quad - \frac{1 - x_A}{n} \sum_{i=1}^n \frac{I\{\delta_i = 3, V_i > A_n\}}{1 - x_i} \\
 &= Q^{(n)}(A_n) + \frac{x_A}{n} \sum_{i=1}^n \frac{I\{\delta_i = 2, V_i < A_n\}}{x_i} \\
 &\quad + \frac{x_A}{n} \sum_{i=1}^n \frac{I\{\delta_i = 2, V_i = A_n\}}{x_i} \\
 &\quad - \frac{1 - x_A}{n} \sum_{i=1}^n \frac{I\{\delta_i = 3, V_i > A_n\}}{1 - x_i} \\
 &= Q_1^{(n)}(A_n) + Q_2^{(n)}(A_n) + \frac{1}{n} \sum_{i=1}^n \frac{(x_A - x_i)I\{\delta_i = 3, V_i > A_n\}}{1 - x_i} \\
 &\quad + \frac{x_A}{n} \sum_{i=1}^n I\{\delta_i = 2, V_i < A_n\} + \frac{1}{n} \sum_{i=1}^n I\{\delta_i = 2, V_i = A_n\} \\
 &\leq Q^{(n)}(A_n) + \frac{1}{n} \sum_{i=1}^n I\{\delta_i = 2, V_i \leq A_n\} \\
 &= \frac{1}{n} \{[\# \text{ of } V_i\text{'s} > A_n] + [\# \text{ of } V_i\text{'s} \leq A_n \text{ with } \delta_i = 2]\} \\
 &= \frac{1}{n} \{[\# \text{ of } V_i\text{'s} \neq A_n] + [\# \text{ of } V_i\text{'s} = A_n \text{ with } \delta_i = 2]\} \leq \frac{(n-1)}{n}.
 \end{aligned}$$

6. When  $\beta_B = 0$ , denoting  $B'_n = \max\{W_k; W_k < B_n\}$  and denoting  $x_{B'}$  and  $y_{B'}$  as the components corresponding to  $B'_n$  for  $\mathbf{x}$  and  $\mathbf{y}$ , respectively, by the definition of  $B_n$ , we have

$$\begin{aligned}
 y_{B'} &= Q^{(n)}(B'_n) + \frac{x_{B'}}{n} \sum_{i=1}^n \frac{I\{\delta_i = 2, V_i \leq B'_n\}}{x_i} \\
 &\quad - \frac{1 - x_{B'}}{n} \sum_{i=1}^n \frac{I\{\delta_i = 3, V_i > B'_n\}}{1 - x_i} \\
 &\geq Q^{(n)}(B'_n) - Q_3^{(n)}(B'_n) = Q_1^{(n)}(B'_n) + Q_2^{(n)}(B'_n) \geq \frac{1}{n}.
 \end{aligned}$$

7. When  $B_n < W_m$  and  $\beta_B > 0$ , denoting  $\mathbf{x}_B$  and  $\mathbf{y}_B$  as the components corresponding to  $B_n$  for  $\mathbf{x}$  and  $\mathbf{y}$ , respectively, by the definition of  $B_n$ , we have

$$\begin{aligned} y_B &= Q^{(n)}(B_n) + \frac{x_B}{n} \sum_{i=1}^n \frac{I\{\delta_i = 2, V_i \leq B_n\}}{x_i} \\ &\quad - \frac{1 - x_B}{n} \sum_{i=1}^n \frac{I\{\delta_i = 3, V_i > B_n\}}{1 - x_i} \\ &\geq Q^{(n)}(B_n) - Q_3^{(n)}(B_n) + \frac{1}{n} \geq \frac{1}{n}. \end{aligned}$$

8. When  $B_n = W_m$  and  $\beta_B > 0$ , we also have  $y_B \geq 1/n$  by the same argument as above for fact 7.

PROOF OF THEOREM 2. (i) From the above discussion, we have that for any  $\mathbf{x} \in \Omega_n$ :

(a) In the case of  $\beta_B = 0$  or  $B_n < W_m$ ,

$$(3.3) \quad \Psi(\mathbf{x}) = (1, \dots, 1, \Psi_a(\mathbf{x}), \dots, \Psi_b(\mathbf{x}), 0, \dots, 0),$$

where  $W_a = A_n$ ,  $W_b = B'_n$  or  $B_n$ , and  $\mathbf{x} = (1, \dots, 1, x_a, \dots, x_b, 0, \dots, 0)$  with  $x_k \in [1/n, (n - 1)/n]$ ,  $k = a, a + 1, \dots, b$ .

(b) In the case of  $B_n = W_m$  and  $\beta_B > 0$ ,

$$(3.4) \quad \Psi(\mathbf{x}) = (1, \dots, 1, \Psi_a(\mathbf{x}), \dots, \Psi_m(\mathbf{x})),$$

where  $W_a = A_n$  and  $\mathbf{x} = (1, \dots, 1, x_a, \dots, x_m)$  with  $x_k \in [1/n, (n - 1)/n]$ ,  $k = a, \dots, m$ . Hence, we know that  $\Psi$  is a continuous function from  $\Omega_n$  to  $\Omega_n$ . Note that  $\Omega_n$  is a compact convex set. From the Brouwer fixed-point theorem [Ortega and Rheinboldt (1970), page 161], there exists a fixed point of  $\Psi$ , say  $\mathbf{x}^*$ , in  $\Omega_n$ .

(ii) *Sufficiency.* To show that a fixed point of  $\Psi$  in  $\Omega_n$  is the NPMLE, from the uniqueness of the NPMLE, we just need to show that a fixed point of  $\Psi$  in  $\Omega_n$  is an optimal point of the system

$$(3.5) \quad \begin{aligned} \max \sum_{k=1}^m \{ \alpha_k \log(P_{k-1} - P_k) + \beta_k \log P_k + \gamma_k \log(1 - P_k) \} 1 - P_1 \\ \geq 0, P_1 - P_2 \geq 0, \dots, P_{m-1} - P_m \geq 0, P_m \geq 0. \end{aligned}$$

Since all constraints of (3.5) are linear functions and since the  $\log L(S)$  is concave, by Theorem 2.15 of Zangwill [(1967), page 43], it suffices to show that a fixed point of  $\Psi$  in  $\Omega_n$ ,  $(\hat{P}_1, \hat{P}_2, \dots, \hat{P}_m)$ , satisfies the following

Kuhn–Tucker condition for problem (3.5) [Zangwill (1967), page 42]: for  $\lambda_k \geq 0, 1 \leq k \leq m + 1$ ,

$$\frac{\partial \Psi}{\partial P_k} = -\frac{\alpha_k}{P_{k-1} - P_k} + \frac{\beta_k}{P_k} - \frac{\gamma_k}{1 - P_k} + \frac{\alpha_{k+1}}{P_k - P_{k+1}} - \lambda_k + \lambda_{k+1} = 0,$$

where  $k = 1, 2, \dots, m - 1$  and  $P_0 = 1$ ;

$$\frac{\partial \Psi}{\partial P_m} = -\frac{\alpha_m}{P_{m-1} - P_m} + \frac{\beta_m}{P_m} - \frac{\gamma_m}{1 - P_m} - \lambda_m + \lambda_{m+1} = 0;$$

$$\lambda_1(1 - P_1) = 0, \lambda_2(P_1 - P_2) = 0, \dots, \lambda_m(P_{m-1} - P_m) = 0, \lambda_{m+1}P_m = 0,$$

where  $\Psi = \log L + \lambda_1(1 - P_1) + \lambda_2(P_1 - P_2) + \dots + \lambda_m(P_{m-1} - P_m) + \lambda_{m+1}P_m$ .

Denote for  $1 \leq k \leq m$ ,

$$\Delta_k = 1 - \frac{1}{n} \sum_{i=1}^n \left\{ \frac{I\{\delta_i = 2, V_i \leq W_{k-1}\}}{\hat{P}_i} + \frac{I\{\delta_i = 3, V_i \geq W_k\}}{1 - \hat{P}_i} \right\},$$

where  $W_0 = -\infty$ . Since  $(\hat{P}_1, \hat{P}_2, \dots, \hat{P}_m) \in \Omega_n$ , we must have  $\hat{P}_k > 0$  if  $\beta_k > 0$  and  $\hat{P}_k < 1$  if  $\gamma_k > 0$ . Hence, (3.7) is well defined for each  $k$  and (3.2) holds for all  $W_1 \leq W_{k-1} < W_k \leq W_m$ . From (3.2), we have that for  $k \geq 2$ ,

$$\alpha_k / (\hat{P}_{k-1} - \hat{P}_k) = n\Delta_k, \quad \text{if } \alpha_k \neq 0,$$

$$(\hat{P}_{k-1} - \hat{P}_k)\Delta_k = 0, \quad \text{if } \alpha_k = 0,$$

where  $\Delta_k \geq 0$  for  $1 \leq k \leq m$  by condition (2). For  $k = 1$ , we have that by (2.3),

$$\begin{aligned} \hat{P}_1 &= Q^{(n)}(W_1) + \frac{\hat{P}_1}{n} \sum_{i=1}^n \frac{I\{\delta_i = 2, V_i \leq W_1\}}{\hat{P}_i} - \frac{1 - \hat{P}_1}{n} \sum_{i=1}^n \frac{I\{\delta_i = 3, V_i > W_1\}}{1 - \hat{P}_i} \\ &= 1 - \frac{\alpha_1 + \beta_1 + \gamma_1}{n} + \frac{\beta_1}{n} + \frac{\gamma_1}{n} - (1 - \hat{P}_1)(1 - \Delta_1) \\ &= 1 - \frac{\alpha_1}{n} - (1 - \hat{P}_1)(1 - \Delta_1), \end{aligned}$$

which implies

$$\alpha_1 = (1 - \hat{P}_1)n\Delta_1 \quad \text{and} \quad \Delta_1 \geq 0.$$

Another obvious fact implied by (3.7) is

$$n\Delta_k = n\Delta_{k+1} + \frac{\beta_k}{\hat{P}_k} - \frac{\gamma_k}{1 - \hat{P}_k}, \quad 1 \leq k \leq m - 1.$$

Next, we show that the following give a solution of the system (3.6):

$$\begin{aligned}
 \text{for } k = 1, & \quad \hat{\lambda}_1 = 0 \text{ if } \alpha_1 + \gamma_1 > 0; \quad \hat{\lambda}_1 = n\Delta_1 \text{ if } \alpha_1 + \gamma_1 = 0, \\
 \text{for } 1 < k < m, & \quad \hat{\lambda}_k = 0 \text{ if } \alpha_k \neq 0; \quad \hat{\lambda}_k = n\Delta_k \text{ if } \alpha_k = 0; \\
 \text{for } k = m; & \quad \hat{\lambda}_m = 0, \quad \hat{\lambda}_{m+1} = \frac{\alpha_m}{\hat{P}_{m-1} - \hat{P}_m} + \frac{\gamma_m}{1 - \hat{P}_m} \\
 & \quad \text{if } \alpha_m \neq 0 \text{ and } \beta_m = 0, \\
 & \quad \hat{\lambda}_m = 0, \quad \hat{\lambda}_{m+1} = 0 \text{ if } \alpha_m \neq 0 \text{ and } \beta_m \neq 0, \\
 & \quad \hat{\lambda}_m = n\Delta_m, \quad \hat{\lambda}_{m+1} = \hat{\lambda}_m + \frac{\gamma_m}{1 - \hat{P}_m} \text{ if } \alpha_m = 0, \beta_m = 0, \\
 & \quad \hat{\lambda}_m = n\Delta_m, \quad \hat{\lambda}_{m+1} = 0 \text{ if } \alpha_m = 0 \text{ and } \beta_m \neq 0.
 \end{aligned}$$

Consider  $1 < k < m$ . For each  $k$ , we need to show

$$(3.12) \quad \frac{\partial \Psi}{\partial P_k} = 0 \quad \text{and} \quad \lambda_k (\hat{P}_{k-1} - \hat{P}_k) = 0.$$

*Case 1.* If  $\alpha_k \neq 0$ , for  $\hat{\lambda}_k = 0$ , we have  $\hat{\lambda}_k (\hat{P}_{k-1} - \hat{P}_k) = 0$ . If  $\alpha_{k+1} = 0$ , for  $\hat{\lambda}_{k+1} = n\Delta_{k+1}$ , we have that by (3.8) and (3.11),

$$\frac{\partial \Psi}{\partial P_k} = -\frac{\alpha_k}{\hat{P}_{k-1} - \hat{P}_k} + \frac{\beta_k}{\hat{P}_k} - \frac{\gamma_k}{1 - \hat{P}_k} + \hat{\lambda}_{k+1} = 0.$$

If  $\alpha_{k+1} \neq 0$ , for  $\hat{\lambda}_{k+1} = 0$ , we have that by (3.8) and (3.11),

$$\frac{\partial \Psi}{\partial P_k} = -\frac{\alpha_k}{\hat{P}_{k-1} - \hat{P}_k} + \frac{\beta_k}{\hat{P}_k} - \frac{\gamma_k}{1 - \hat{P}_k} + \frac{\alpha_{k+1}}{\hat{P}_k - \hat{P}_{k+1}} = 0.$$

*Case 2.* If  $\alpha_k = 0$ , for  $\hat{\lambda}_k = n\Delta_k$ , by (3.9) we have  $\hat{\lambda}_k (\hat{P}_{k-1} - \hat{P}_k) = 0$ . If  $\alpha_{k+1} \neq 0$ , for  $\hat{\lambda}_{k+1} = 0$ , we have that by (3.8) and (3.11),

$$\frac{\partial \Psi}{\partial P_k} = \frac{\beta_k}{\hat{P}_k} - \frac{\gamma_k}{1 - \hat{P}_k} + \frac{\alpha_{k+1}}{\hat{P}_k - \hat{P}_{k+1}} - n\Delta_k = 0.$$

If  $\alpha_{k+1} = 0$ , for  $\hat{\lambda}_{k+1} = n\Delta_{k+1}$ , we have that by (3.11),

$$\frac{\partial \Psi}{\partial P_k} = \frac{\beta_k}{\hat{P}_k} - \frac{\gamma_k}{1 - \hat{P}_k} - n\Delta_k + n\Delta_{k+1} = 0.$$

Consider  $k = 1$ . We need to show

$$\frac{\partial \Psi}{\partial P_1} = 0 \quad \text{and} \quad \lambda_1 (1 - \hat{P}_1) = 0.$$

Case 1. If  $\alpha_1 + \gamma_1 > 0$ , for  $\hat{\lambda}_1 = 0$  and  $\hat{P}_1 < 1$ , we have  $\hat{\lambda}_1(1 - \hat{P}_1) = 0$ . If  $\alpha_2 \neq 0$ , for  $\hat{\lambda}_2 = 0$ , we have that by (3.10) and (3.11),

$$\begin{aligned} \frac{\partial \Psi}{\partial P_1} &= -\frac{\alpha_1}{1 - \hat{P}_1} + \frac{\beta_1}{\hat{P}_1} - \frac{\gamma_1}{1 - \hat{P}_1} + \frac{\alpha_2}{\hat{P}_1 - \hat{P}_2} \\ &= -n\Delta_1 + \frac{\beta_1}{\hat{P}_1} - \frac{\gamma_1}{1 - \hat{P}_1} + \frac{\alpha_2}{\hat{P}_1 - \hat{P}_2} = 0. \end{aligned}$$

If  $\alpha_2 = 0$ , for  $\hat{\lambda}_2 = n\Delta_2$ , we have that by (3.10) and (3.11),

$$\frac{\partial \Psi}{\partial P_1} = -\frac{\alpha_1}{1 - \hat{P}_1} + \frac{\beta_1}{\hat{P}_1} - \frac{\gamma_1}{1 - \hat{P}_1} + n\Delta_2 = 0.$$

Case 2. If  $\alpha_1 + \gamma_1 = 0$ , for  $\hat{\lambda}_1 = n\Delta_1$  and  $\hat{P}_1 = 1$ , we have  $\hat{\lambda}_1(1 - \hat{P}_1) = 0$ . If  $\alpha_2 \neq 0$ , for  $\hat{\lambda}_2 = 0$ , we have that by (3.8) and (3.11),

$$\frac{\partial \Psi}{\partial P_1} = \frac{\beta_1}{\hat{P}_1} + \frac{\alpha_2}{\hat{P}_1 - \hat{P}_2} - n\Delta_1 = 0.$$

If  $\alpha_2 = 0$ , for  $\hat{\lambda}_2 = n\Delta_2$ , we have that by (3.11),

$$\frac{\partial \Psi}{\partial P_1} = \frac{\beta_1}{\hat{P}_1} - n\Delta_1 + n\Delta_2 = 0.$$

Consider  $k = m$ . We need to show

$$\frac{\partial \Psi}{\partial P_m} = 0, \quad \lambda_m(\hat{P}_{m-1} - \hat{P}_m) = 0, \quad \lambda_{m+1}\hat{P}_m = 0.$$

Case 1. If  $\alpha_m \neq 0$  and  $\beta_m = 0$ , for  $\hat{\lambda}_m = 0$ ,  $\hat{P}_m = 0$  and  $\hat{\lambda}_{m+1} = \alpha_m/(\hat{P}_{m-1} - \hat{P}_m) + \gamma_m/(1 - \hat{P}_m)$ , we have that  $\hat{\lambda}_m(\hat{P}_{m-1} - \hat{P}_m) = 0$ ,  $\hat{\lambda}_{m+1}\hat{P}_m = 0$  and

$$\frac{\partial \Psi}{\partial P_m} = -\frac{\alpha_m}{\hat{P}_{m-1} - \hat{P}_m} - \frac{\gamma_m}{1 - \hat{P}_m} + \frac{\alpha_m}{\hat{P}_{m-1} - \hat{P}_m} + \frac{\gamma_m}{1 - \hat{P}_m} = 0.$$

Case 2. If  $\alpha_m \neq 0$  and  $\beta_m \neq 0$ , for  $\hat{\lambda}_m = 0$ ,  $\hat{\lambda}_{m+1} = 0$  and  $\hat{P}_m > 0$ , we have that  $\hat{\lambda}_m(\hat{P}_{m-1} - \hat{P}_m) = 0$ ,  $\hat{\lambda}_{m+1}\hat{P}_m = 0$  and

$$\frac{\partial \Psi}{\partial P_m} = -\frac{\alpha_m}{\hat{P}_{m-1} - \hat{P}_m} + \frac{\beta_m}{\hat{P}_m} - \frac{\gamma_m}{1 - \hat{P}_m} = 0,$$

because  $\hat{P}_m > 0$  and

$$\hat{P}_m = \frac{\hat{P}_m}{n} \sum_{i=1}^n \frac{I\{\delta_i = 2, V_i \leq W_m\}}{\hat{P}_i} - \frac{1 - \hat{P}_m}{n} \sum_{i=1}^n \frac{I\{\delta_i = 3, V_i > W_m\}}{1 - \hat{P}_i}$$

implies

$$1 = \frac{1}{n} \sum_{i=1}^n \frac{I\{\delta_i = 2, V_i \leq W_m\}}{\hat{P}_i}.$$

Thus

$$(3.13) \quad \begin{aligned} n\Delta_m &= n - \sum_{i=1}^n \left\{ \frac{I\{\delta_i = 2, V_i \leq W_{m-1}\}}{\hat{P}_i} + \frac{I\{\delta_i = 3, V_i \geq W_m\}}{1 - \hat{P}_i} \right\} \\ &= \frac{\beta_m}{\hat{P}_m} - \frac{\gamma_m}{1 - \hat{P}_m}. \end{aligned}$$

Case 3. If  $\alpha_m = 0, \beta_m = 0$ , for  $\hat{P}_m = 0, \hat{\lambda}_m = n\Delta_m$  and  $\hat{\lambda}_{m+1} = \hat{\lambda}_m + \gamma_m/(1 - \hat{P}_m)$ , we have  $\hat{\lambda}_m(\hat{P}_{m-1} - \hat{P}_m) = 0, \hat{\lambda}_{m+1}\hat{P}_m = 0$  and

$$\frac{\partial \Psi}{\partial P_m} = -\frac{\gamma_m}{1 - \hat{P}_m} - n\Delta_m + n\Delta_m + \frac{\gamma_m}{1 - \hat{P}_m} = 0.$$

Case 4. If  $\alpha_m = 0$  and  $\beta_m \neq 0$ , for  $\hat{\lambda}_m = n\Delta_m, \hat{\lambda}_{m+1} = 0$  and  $\hat{P}_m > 0$ , we have that  $\hat{\lambda}_m(\hat{P}_{m-1} - \hat{P}_m) = 0, \hat{\lambda}_{m+1}\hat{P}_m = 0$  and, by (3.13),

$$\frac{\partial \Psi}{\partial P_m} = \frac{\beta_m}{\hat{P}_m} - \frac{\gamma_m}{1 - \hat{P}_m} - n\Delta_m = 0.$$

This completes the proof of sufficiency.

*Necessity.* To show that the NPMLE  $\hat{S}_X^{(n)}$  must be a fixed point of  $\Psi$  in  $\Omega_n$  satisfying condition (2), we see that by Theorem 2.14 of Zangwill [(1967), page 40], it suffices to prove that a solution  $\{\hat{P}_k; 1 \leq k \leq m\}$  of (3.6) is a fixed point of  $\Psi$  in  $\Omega_n$  with all  $\Delta_k \geq 0$ .

Let  $\hat{P}_k = \hat{S}_X^{(n)}(W_k), k = 1, \dots, m$ , be a solution of (3.6). By adding the equations in (3.6), we obtain

$$(3.14) \quad -\frac{\alpha_1}{d_1} + \sum_{j=1}^m \frac{\beta_j}{\hat{P}_j} - \sum_{j=1}^m \frac{\gamma_j}{1 - \hat{P}_j} - \lambda_1 + \lambda_{m+1} = 0,$$

$$(3.15) \quad -\frac{\alpha_1}{d_1} + \frac{\alpha_k}{d_k} + \sum_{j=1}^{k-1} \frac{\beta_j}{\hat{P}_j} - \sum_{j=1}^{k-1} \frac{\gamma_j}{1 - \hat{P}_j} - \lambda_1 + \lambda_k = 0,$$

where  $d_k = \hat{P}_{k-1} - \hat{P}_k, 1 \leq k \leq m$ , with  $\hat{P}_0 = 1$ . Hence, when  $\alpha_k > 0$ , we have  $\lambda_k = 0$  and

$$(3.16) \quad \begin{aligned} \frac{\alpha_k}{d_k} &= \frac{\alpha_1}{d_1} + \lambda_1 - \sum_{j=1}^{k-1} \frac{\beta_j}{\hat{P}_j} + \sum_{j=1}^{k-1} \frac{\gamma_j}{1 - \hat{P}_j} \\ &= \sum_{j=1}^m \frac{\beta_j}{\hat{P}_j} - \sum_{j=1}^{k-1} \frac{\beta_j}{\hat{P}_j} - \sum_{j=k}^m \frac{\gamma_j}{1 - \hat{P}_j} + \lambda_{m+1} \end{aligned}$$

and when  $\alpha_k = 0$ ,

$$(3.17) \quad \begin{aligned} \lambda_k &= \frac{\alpha_1}{d_1} + \lambda_1 - \sum_{j=1}^{k-1} \frac{\beta_j}{\hat{P}_j} + \sum_{j=1}^{k-1} \frac{\gamma_j}{1 - \hat{P}_j} \\ &= \sum_{j=1}^m \frac{\beta_j}{\hat{P}_j} - \sum_{j=1}^{k-1} \frac{\beta_j}{\hat{P}_j} - \sum_{j=k}^m \frac{\gamma_j}{1 - \hat{P}_j} + \lambda_{m+1}. \end{aligned}$$



Thus, (3.6), (3.16) and (3.17) imply

$$(3.18) \quad d_k \left( D - \sum_{j=1}^{k-1} \frac{\beta_j}{\hat{P}_j} - \sum_{j=k}^m \frac{\gamma_j}{1 - \hat{P}_j} \right) = \alpha_k, \quad 1 \leq k \leq m,$$

where  $D = \sum_{j=1}^m \beta_j / \hat{P}_j + \lambda_{m+1}$ . Hence for  $k = 1$ , we have

$$(1 - \hat{P}_1) \left( 1 - \frac{1}{D} \sum_{j=1}^m \frac{\gamma_j}{1 - \hat{P}_j} \right) = \frac{\alpha_1}{D}.$$

Since

$$\frac{1}{D} \sum_{j=1}^m \frac{\gamma_j}{1 - \hat{P}_j} = \frac{\gamma_1}{D(1 - \hat{P}_1)} + \frac{1}{D} \sum_{j=2}^m \frac{\gamma_j}{1 - \hat{P}_j},$$

we have

$$(3.19) \quad \hat{P}_1 = 1 - \frac{\pi_1}{D} + \frac{\hat{P}_1}{D} \sum_{j=1}^1 \frac{\beta_j}{\hat{P}_j} - \frac{1 - \hat{P}_1}{D} \sum_{j=2}^m \frac{\gamma_j}{1 - \hat{P}_j},$$

where  $\pi_k = \alpha_k + \beta_k + \gamma_k$ . Note that (3.19) is the equation (2.3) for  $k = 1$ . Using the derivation of (3.2), from (3.19) and (3.18), we know that for any  $1 \leq k \leq m$ ,

$$(3.20) \quad \hat{P}_k = 1 - \frac{\pi_1 + \dots + \pi_k}{D} + \frac{\hat{P}_k}{D} \sum_{j=1}^k \frac{\beta_j}{\hat{P}_j} - \frac{1 - \hat{P}_k}{D} \sum_{j=k+1}^m \frac{\gamma_j}{1 - \hat{P}_j}.$$

Note that  $\lambda_k, 1 \leq k \leq m + 1$ , satisfying (3.6) implies  $\lambda_k \geq 0, 1 \leq k \leq m + 1$ . Clearly, if  $D = n$ , we have that (3.17) implies  $\lambda_k = n\Delta_k$  when  $\alpha_k = 0$  and that (3.18) and (3.7) imply  $\Delta_k > 0$  when  $\alpha_k > 0$ . Hence, if  $D = n$ , statement (2) of Theorem 2(ii) holds for  $\hat{S}_X^{(n)}$ . Moreover, if  $D = n$ , we know that from (3.2),  $\hat{S}_X^{(n)}$  is a fixed point and that the facts listed before the proof of Theorem 2 imply  $\hat{S}_X^{(n)} \in \Omega_n$ . We complete the proof by showing  $D = n$ . In (3.20), for  $k = m$ ,

$$\hat{P}_m = 1 - \frac{n}{D} + \frac{\hat{P}_m}{D} \sum_{j=1}^m \frac{\beta_j}{\hat{P}_j}.$$

If  $\hat{P}_m = 0$ , we have  $D = n$ . If  $\hat{P}_m > 0$ , we have  $\lambda_{m+1} = 0$  and  $D = \sum_{j=1}^m \beta_j / \hat{P}_j$ ; hence we also have  $D = n$ .  $\square$

**PROOF OF COROLLARY 3.** First we notice that in (3.2), if we choose  $x_{k-1} = x_k$  when  $\alpha_k = 0$  and  $A_n < W_k < J_n$ , then we must have  $y_{k-1} = y_k$ . Hence, from the proof of Theorem 2, there is a fixed point  $\bar{\mathbf{x}}$  of  $\Psi$  in  $\Omega_n$  satisfying  $\bar{P}_{k-1} = \bar{P}_k$  for  $\alpha_k = 0$  and  $A_n < W_k < J_n$ . This implies that for  $A_n < W_k < J_n$ ,

the self-consistent estimator  $\bar{S}_X^{(n)}$  only has jumps when  $\alpha_k > 0$  and, by (3.2), the jump size is given by

$$(3.21) \quad \begin{aligned} & \bar{S}_X^{(n)}(W_{k-1}) - \bar{S}_X^{(n)}(W_k) \\ &= \alpha_k \left\{ n - \sum_{i=1}^n \left( \frac{I\{\delta_i = 2, V_i \leq W_{k-1}\}}{\bar{P}_i} + \frac{I\{\delta_i = 3, V_i \geq W_k\}}{1 - \bar{P}_i} \right) \right\}^{-1} \\ &= \frac{\alpha_k/n}{\bar{S}_Y^{(n)}(W_{k-1}) - \bar{S}_Z^{(n)}(W_{k-1})}. \end{aligned}$$

Next, we consider the jump sizes at  $A_n$  and  $J_n$ .

Case 1. The jump size at  $A_n = W_k$  (with  $\alpha_k + \gamma_k > 0$ ) is given by

$$(3.22) \quad 1 - \bar{P}_k = \frac{(\alpha_k + \gamma_k)/n}{1 - \bar{S}_Z^{(n)}(W_k)}$$

because

$$\begin{aligned} \bar{P}_k &= Q^{(n)}(W_k) + \frac{\bar{P}_k}{n} \sum_{i=1}^n \frac{I\{\delta_i = 2, V_i \leq W_k\}}{\bar{P}_i} - \frac{1 - \bar{P}_k}{n} \sum_{i=1}^n \frac{I\{\delta_i = 3, V_i > W_k\}}{1 - \bar{P}_i}. \\ &= 1 - \frac{[\# \text{ of } V_i\text{'s} < W_k]}{n} - \frac{\alpha_k + \beta_k + \gamma_k}{n} + \frac{\beta_k}{n} + \frac{[\# \text{ of } V_i\text{'s} < W_k]}{n} \\ &\quad - \frac{1 - \bar{P}_k}{n} \sum_{i=1}^n \frac{I\{\delta_i = 3, V_i > W_k\}}{1 - \bar{P}_i} \\ &= 1 - \frac{\alpha_k + \gamma_k}{n} - \frac{1 - \bar{P}_k}{n} \sum_{i=1}^n \frac{I\{\delta_i = 3, V_i > W_k\}}{1 - \bar{P}_i}. \end{aligned}$$

Case 2. If  $\beta_m > 0$ , then  $J_n = W_m$  with  $\bar{S}_X^{(n)}(W_m) > 0$ . In this case, by (3.8), the jump size of  $\bar{S}_X^{(n)}$  at  $J_n$  is given by

$$(3.23) \quad \frac{\alpha_m/n}{\bar{S}_Y^{(n)}(W_{m-1}) - \bar{S}_Z^{(n)}(W_{m-1})}.$$

Case 3. If  $\beta_m = 0$  and  $\alpha_m > 0$ , then  $J_n = W_m$  and the jump size is given by

$$(3.24) \quad \frac{\alpha_m/n}{\bar{S}_Y^{(n)}(W_{m-1}) - \bar{S}_Z^{(n)}(W_{m-1})}.$$

Case 4. If  $\beta_m = 0$  and  $\alpha_m = 0$ , then  $J_n = B_n$ , if  $\beta_B = 0$ ;  $J_n = C_n$ , if  $\beta_B > 0$ . Let  $J_n = W_j$ . When  $\beta_B = \beta_J = 0$ , the jump size is given by

$$(3.25) \quad \bar{S}_X^{(n)}(W_{j-1}) = \bar{P}_{j-1} = \frac{\alpha_J/n}{\bar{S}_Y^{(n)}(W_{j-1}) - \bar{S}_Z^{(n)}(W_{j-1})},$$

because

$$\begin{aligned} \bar{P}_{j-1} &= Q^{(n)}(W_{j-1}) - \frac{1}{n} \sum_{i=1}^n \frac{I\{\delta_i = 3, V_i \geq W_j\}}{1 - \bar{P}_i} \\ &\quad + \frac{\bar{P}_{j-1}}{n} \sum_{i=1}^n \left( \frac{I\{\delta_i = 2, V_i \leq W_{j-1}\}}{\bar{P}_i} + \frac{I\{\delta_i = 3, V_i \geq W_j\}}{1 - \bar{P}_i} \right) \\ &= \frac{\alpha_j + \beta_j + \gamma_j}{n} - \frac{\gamma_j}{n} \\ &\quad + \frac{\bar{P}_{j-1}}{n} \sum_{i=1}^n \left( \frac{I\{\delta_i = 2, V_i \leq W_{j-1}\}}{\bar{P}_i} + \frac{I\{\delta_i = 3, V_i \geq W_j\}}{1 - \bar{P}_i} \right). \end{aligned}$$

When  $\beta_B > 0$ , we have  $W_{j-1} = B_n$ ,  $W_j = C_n$ , and the jump size is given by

$$(3.26) \quad \bar{S}_X^{(n)}(W_{j-1}) = \bar{P}_{j-1} = \frac{\beta_{j-1}/n}{\bar{S}_Y^{(n)}(W_{j-2}) - \bar{S}_Z^{(n)}(W_{j-1})},$$

because  $\bar{P}_{j-1} > 0$  and

$$\begin{aligned} \bar{P}_{j-1} &= Q^{(n)}(W_{j-1}) - \frac{1}{n} \sum_{i=1}^n \frac{I\{\delta_i = 3, V_i \geq W_j\}}{1 - \bar{P}_i} \\ &\quad + \frac{\bar{P}_{j-1}}{n} \sum_{i=1}^n \left( \frac{I\{\delta_i = 2, V_i \leq W_{j-1}\}}{\bar{P}_i} + \frac{I\{\delta_i = 3, V_i \geq W_j\}}{1 - \bar{P}_i} \right) \\ &= \frac{\bar{P}_{j-1}}{n} \sum_{i=1}^n \left( \frac{I\{\delta_i = 2, V_i \leq W_{j-1}\}}{\bar{P}_i} + \frac{I\{\delta_i = 3, V_i \geq W_j\}}{1 - \bar{P}_i} \right) \end{aligned}$$

imply

$$1 = \frac{1}{n} \sum_{i=1}^n \left( \frac{I\{\delta_i = 2, V_i \leq W_{j-1}\}}{\bar{P}_i} + \frac{I\{\delta_i = 3, V_i \geq W_j\}}{1 - \bar{P}_i} \right)$$

and, thus,

$$1 = \frac{\beta_{j-1}}{n\bar{P}_{j-1}} + \frac{1}{n} \sum_{i=1}^n \left( \frac{I\{\delta_i = 2, V_i \leq W_{j-2}\}}{\bar{P}_i} + \frac{I\{\delta_i = 3, V_i \geq W_j\}}{1 - \bar{P}_i} \right).$$

Therefore, (2.9) follows from (3.21)–(3.26).

The uniqueness of  $\bar{S}_X^{(n)}$  simply follows from observing that in fact we group the observations with zero  $\alpha$  with the closest  $\alpha$ -type observation on the left and that the problem becomes equivalent to the NPMLE problem with only

$\alpha$ -type points. One should note that in (2.1), we only need to consider the case of  $\gamma_A = 0$ . From (3.8), (3.11) and Theorem 2, this  $\hat{S}_X^{(n)}$  is unique.  $\square$

PROOF OF COROLLARY 4. Let  $\hat{S}_X^{(n)}$  be the NPMLE in  $\Omega_n$ . By Corollary 3, we only need to show  $\hat{S}_X^{(n)} \in \Omega_n^+$ , that is we need to show that  $\hat{S}_X^{(n)}$  has no jumps at each  $W_k$ ,  $W_i < W_k < W_j$ . From Theorem 2 and (3.9), it suffices to show that for each  $W_k$ , we have  $\Delta_k > 0$ . The proof follows from (3.11).  $\square$

PROOF OF COROLLARY 5. (i) Let  $\hat{S}_X^{(n)}$  be the NPMLE in  $\Omega_n$ . If  $\hat{S}_X^{(n)}$  has a jump at a non- $\alpha$ -type point  $W_k$ , by Theorem 2, (3.9) and (3.11) we have  $\Delta_k = 0$  and

$$(3.27) \quad n\Delta_{k+1} = \frac{\gamma_k}{1 - \hat{P}_k} - \frac{\beta_k}{\hat{P}_k} \geq 0,$$

$$(3.28) \quad n\Delta_{k-1} = \frac{\beta_{k-1}}{\hat{P}_{k-1}} - \frac{\gamma_{k-1}}{1 - \hat{P}_{k-1}} \geq 0.$$

Since  $\alpha_k = 0$ , by no ties assumption, we must have  $\beta_k + \gamma_k = 1$ . Hence, (3.27) implies  $\beta_k = 0, \gamma_k = 1$ . Clearly, (3.28) implies  $\gamma_{k-1} = 0$ . If  $\alpha_{k-1} = 1$ , then by (3.8), we must have  $\Delta_{k-1} > 0$ . This requires  $\beta_{k-1} = 1$  in (3.28)—contradiction. Hence, we must have  $\alpha_{k-1} = 0, \beta_{k-1} = 1$ . Hence,  $W_k$  is a 2-3-type point.

(ii) The proof is similar to that of Corollary 4.  $\square$

**4. Proof of Theorem 6.**

PROOF OF THEOREM 6(i). Using the notation of Wu (1983) for the EM algorithm, the corresponding  $Q(F|F_0)$  for doubly censored data is given by

$$(4.1) \quad Q(S|S_0) = \sum_{i=1}^n E_{S_0}\{\log P_S\{X = X_i\}|V_i = v_i, \delta_i = \delta_i\},$$

where  $S_0$  (satisfying the conditions of Theorem 1) and  $S$  are the survival functions with  $f_0(x) = P_{S_0}\{X = x\}$  and  $f(x) = P_S\{X = x\}$ , respectively, and  $(v_i, \delta_i), i = 1, \dots, n$ , are observed points. Assume that  $S_0$  and  $S$  only put mass at the observed points  $v_1 \leq v_2 \leq \dots \leq v_n$  and the interval  $(v_n, \infty)$ . Then:

- (i)  $E_{S_0}\{\log P_S\{X = X_i\}|V_i = v_i, \delta_i = 1\} = \log f(v_i);$
- (ii)  $E_{S_0}\{\log P_S\{X = X_i\}|V_i = v_i, \delta_i = 2\} = \frac{\sum_{v_j > v_i} f_0(v_j) \log f(v_j)}{S_0(v_i)};$
- (iii)  $E_{S_0}\{\log P_S\{X = X_i\}|V_i = v_i, \delta_i = 3\} = \frac{\sum_{v_j \leq v_i} f_0(v_j) \log f(v_j)}{S_0(v_i)}.$

Hence, for all the distinct points  $w_1, \dots, w_m$  of the set  $\{v_1, \dots, v_n\}$  with  $w_1 < \dots < w_m$ ,

$$\begin{aligned}
 Q(S|S_0) &= \sum_{i=1}^n \left\{ \{\delta_i = 1\} \log f(v_i) + I\{\delta_i = 2\} \frac{\sum_{v_j > v_i} f_0(v_j) \log f(v_j)}{S_0(v_i)} \right. \\
 &\quad \left. + I\{\delta_i = 3\} \frac{\sum_{v_j \leq v_i} f_0(v_j) \log f(v_j)}{1 - S_0(v_i)} \right\} \\
 (4.2) \quad &= \sum_{j=1}^{n+1} \left\{ I\{\delta_j = 1\} + f_0(v_j) \sum_{i=1}^n \frac{I\{\delta_i = 2, v_i < v_j\}}{S_0(v_i)} \right. \\
 &\quad \left. + f_0(v_j) \sum_{i=1}^n \frac{I\{\delta_i = 3, v_i \geq v_j\}}{1 - S_0(v_i)} \right\} \log f(v_j) \\
 &= \sum_{k=1}^{m+1} \left\{ \alpha_k + \pi_k f_0(w_k) \sum_{i=1}^n \frac{I\{\delta_i = 2, v_i < w_k\}}{S_0(v_i)} \right. \\
 &\quad \left. + \pi_k f_0(w_k) \sum_{i=1}^n \frac{I\{\delta_i = 3, v_i \geq w_k\}}{1 - S_0(v_i)} \right\} \log f(w_k),
 \end{aligned}$$

where  $\pi_k = \alpha_k + \beta_k + \gamma_k$  for  $k = 1, \dots, m$  and  $\alpha_{m+1} = 0, \pi_{m+1} = 1$ .

To find the  $M$ -step  $S_1$ , that is, the  $S_1$  which maximizes  $Q(S|S_0)$ , we need to find an optimal point for the system

$$\begin{aligned}
 (4.3) \quad &\max \sum_{k=1}^{m+1} \lambda_k \log q_k, \\
 &q_1 + \dots + q_{m+1} = 1,
 \end{aligned}$$

where  $Q(S|S_0) = \sum_{k=1}^{m+1} \lambda_k \log q_k$  with  $q_k = f(w_k)$  and

$$\begin{aligned}
 (4.4) \quad \lambda_k &= \alpha_k + \pi_k f_0(w_k) \sum_{i=1}^n \frac{I\{\delta_i = 2, v_i < w_k\}}{S_0(v_i)} \\
 &\quad + \pi_k f_0(w_k) \sum_{i=1}^n \frac{I\{\delta_i = 3, v_i \geq w_k\}}{1 - S_0(v_i)}.
 \end{aligned}$$

From

$$(4.5) \quad \partial^2 Q = \begin{bmatrix} -\frac{\lambda_1}{q_1^2} & & & \\ & \ddots & & \\ & & & -\frac{\lambda_{m+1}}{q_{m+1}^2} \end{bmatrix} \quad \text{and} \quad \lambda_k \geq 0,$$

we know that  $Q$  is concave. Let  $\Phi = \sum_{k=1}^{m+1} \lambda_k \log q_k + \lambda(q_1 + \dots + q_{m+1} - 1)$ , where  $\lambda$  is the Lagrangian multiplier. For now, we assume that all  $\lambda_k > 0$ .

Then  $Q$  is strictly concave. By Theorem 2.15 of Zangwill [(1967), page 43], we know that a solution of the system

$$(4.6) \quad \begin{aligned} \frac{\partial \Phi}{\partial q_k} &= \frac{\lambda_k}{q_k} + \lambda = 0, & k = 1, \dots, m + 1, \\ q_1 + \dots + q_{m+1} &= 1 \end{aligned}$$

will give a unique  $S_1$ . We easily find that the solution of (4.6) is given by

$$(4.7) \quad \hat{q}_k = \frac{\lambda_k}{\Lambda} = \frac{\lambda_k}{n}, \quad k = 1, \dots, m + 1,$$

because for  $N_j = \sum_{i=1}^n I\{\delta_i = j\}$ ,  $j = 1, 2, 3$ , we have

$$\begin{aligned} \Lambda &= \sum_{k=1}^{m+1} \lambda_k \\ &= \sum_{k=1}^{m+1} \left\{ \alpha_k + \pi_k f_0(z_k) \sum_{i=1}^n \frac{I\{\delta_i = 2, w_i < z_k\}}{1 - F_0(w_i)} \right. \\ &\quad \left. + \pi_k f_0(z_k) \sum_{i=1}^n \frac{I\{\delta_i = 3, w_i \geq z_k\}}{F_0(w_i)} \right\} \\ &= \sum_{i=1}^n \sum_{k=1}^{m+1} \left\{ \frac{\alpha_k}{n} + \pi_k f_0(z_k) \frac{I\{\delta_i = 2, w_i < z_k\}}{1 - F_0(w_i)} \right. \\ &\quad \left. + \pi_k f_0(z_k) \frac{I\{\delta_i = 3, w_i \geq z_k\}}{F_0(w_i)} \right\} \\ &= \sum_{i=1}^n \left\{ \frac{N_1}{n} + I\{\delta_i = 2\} + I\{\delta_i = 3\} \right\} = N_1 + N_2 + N_3 = n. \end{aligned}$$

Now, we show that  $Q(S|S_0)$  has a unique maximum. First, we note that  $q_k$  does not contribute anything to  $Q$  if  $\lambda_k = 0$ . If we assign  $q_k = 0$  when  $\lambda_k = 0$  and repeat the calculation above, we obtain a solution  $\hat{S}$  of (4.3) with  $\hat{q}_k = \lambda_k/n$ . Suppose there is a different solution  $\bar{S}$  for (4.3) which assigns some  $q_j \neq 0$  when  $\lambda_j$  is the only  $\lambda_i$ 's with value 0. Then there exists a constant  $0 < c < 1$  such that  $\bar{S}$  is the solution for the system

$$(4.8) \quad \begin{aligned} \max \sum_{k=1}^{m+1} \lambda_k \log q_k, \\ q_1 + \dots + q_{j-1} + q_{j+1} + \dots + q_{m+1} = c. \end{aligned}$$

By a calculation similar to the above one, an optimal point of (4.8) is given by  $\bar{q}_k = c\lambda_k/n$ . Clearly, we have

$$\sum_{k=1}^{m+1} \lambda_k \log \hat{q}_k \geq \sum_{k=1}^{m+1} \lambda_k \log \bar{q}_k = Q(\bar{S}|S_0).$$

This contradicts the assumption that  $\bar{S}$  is a solution of (4.3). Hence, the maximum of  $Q(S|S_0)$  must have  $q_k = 0$  if  $\lambda_k = 0$ . Therefore, (4.3) is equivalent to the system

$$(4.9) \quad \begin{aligned} \max \sum_{i=1}^l \lambda_{k_i} \log q_{k_i}, \\ q_{k_1} + \cdots + q_{k_l} = 1, \end{aligned}$$

where  $\lambda_{k_1}, \dots, \lambda_{k_l}$  are all those positive  $\lambda_k$ 's and  $q_{k_i}$  is the  $q_k$  corresponding to  $\lambda_{k_i}$ . Let  $Q_1 = \sum_{i=1}^l \lambda_{k_i} \log q_{k_i}$ . Then  $\partial^2 Q_1$  is negative definite; hence,  $Q_1$  is strictly concave and the maximum of  $Q_1$  is unique. Therefore,  $Q(S|S_0)$  has a unique maximum given by (4.7).

Next, we will show that algorithm (2.5) gives the EM algorithm. Specifically, we need to show that  $S_1(w_k) = P_{S_1}\{X > w_k\}$ , that is  $f_1(w_k) = P_{S_1}\{X = w_k\} = \lambda_k/n$  is given by the algorithm (2.5) for  $x_k = S_0(w_k)$ . It suffices to show that

$$ny_k = \sum_{j=1}^{m+1} \lambda_j I\{w_j > w_k\}, \quad k = 1, \dots, m.$$

For any  $k$ , we have

$$\begin{aligned} & \sum_{j=1}^{m+1} \lambda_j I\{w_j > w_k\} \\ &= \sum_{j=1}^{m+1} I\{w_j > w_k\} \left\{ \alpha_j + \pi_j f_0(w_j) \sum_{i=1}^n \frac{I\{\delta_i = 2, v_i < w_j\}}{S_0(v_i)} \right. \\ & \quad \left. + \pi_j f_0(w_j) \sum_{i=1}^n \frac{I\{\delta_i = 3, v_i \geq w_j\}}{1 - S_0(v_i)} \right\} \\ &= \sum_{i=1}^n \sum_{j=1}^{m+1} I\{w_j > w_k\} \left\{ \frac{\alpha_j}{n} + \pi_j f_0(w_j) \frac{I\{\delta_i = 2, v_i < w_j\}}{S_0(v_i)} \right. \\ & \quad \left. + \pi_j f_0(w_j) \frac{I\{\delta_i = 3, v_i \geq w_j\}}{1 - S_0(v_i)} \right\} \\ &= \sum_{i=1}^n \left\{ Q_1^{(n)}(w_k) + \frac{I\{\delta_i = 2\}}{S_0(v_i)} \sum_{j=1}^n I\{v_j > v_i \vee w_k\} f_0(v_j) \right. \\ & \quad \left. + \frac{I\{\delta_i = 3\}}{1 - S_0(v_i)} \sum_{j=1}^n I\{w_k < v_j \leq v_i\} f_0(v_j) \right\} \\ &= \sum_{i=1}^n \left\{ Q_1^{(n)}(w_k) + \frac{I\{\delta_i = 2\}}{S_0(v_i)} \{I\{v_i \leq w_k\} S_0(w_k) + I\{v_i > w_k\} S_0(v_i)\} \right. \\ & \quad \left. + \frac{I\{\delta_i = 3\}}{1 - S_0(v_i)} I\{v_i > w_k\} \{S_0(w_k) - S_0(v_i)\} \right\} \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=1}^n \left\{ Q_1^{(n)}(w_k) + \frac{I\{\delta_i = 2\}}{S_0(v_i)} I\{v_i \leq w_k\} S_0(w_k) + I\{\delta_i = 2, v_i > w_k\} \right. \\
 &\quad \left. + I\{\delta_i = 3, v_i > w_k\} - \frac{I\{\delta_i = 3\}}{1 - S_0(v_i)} I\{v_i > w_k\} \{1 - S_0(w_k)\} \right\} \\
 &= nQ_1^{(n)}(w_k) + x_k \sum_{i=1}^n \frac{I\{\delta_i = 2, v_i \leq w_k\}}{x_i} + nQ_2^{(n)}(w_k) + nQ_3^{(n)}(w_k) \\
 &\quad - (1 - x_k) \sum_{i=1}^n \frac{I\{\delta_i = 3, v_i > w_k\}}{1 - x_i} \\
 &= nQ^{(n)}(w_k) + x_k \sum_{i=1}^n \frac{I\{\delta_i = 2, v_i \leq w_k\}}{x_i} \\
 &\quad - (1 - x_k) \sum_{i=1}^n \frac{I\{\delta_i = 3, v_i > w_k\}}{1 - x_i} \\
 &= ny_k. \qquad \square
 \end{aligned}$$

PROOF OF THEOREM 6(ii). From (4) of Wu (1983), we know that  $L(S_1) > L(S_0)$  unless  $S_0$  is a fixed point of  $\Psi$ , because  $Q(S_1|S_0) > Q(S_0|S_0)$  unless  $S_1 = S_0$ , that is,  $S_0$  is a fixed point of  $\Psi$ . We should notice that any fixed point of  $\Psi$  is a self-consistent estimator of  $S_X$ .

Now, we consider the algorithm (2.5) with an initial point  $\mathbf{x}_1 \in \Omega_n$ . This is a special case of the EM algorithm. Hence, for any  $\mathbf{x} \in \Omega_n$ , we have  $L(\mathbf{y}) > L(\mathbf{x})$  unless  $\mathbf{x}$  is a fixed point of  $\Psi$  in  $\Omega_n$ . From the properties of  $\Psi$  in  $\Omega_n$  discussed in the proof of Theorem 2 and from Convergence Theorem A of Zangwill [(1967), page 91], all the limit points of  $\{\mathbf{x}_N\}$  are the fixed points of  $\Psi$  in  $\Omega_n$ .  $\square$

PROOF OF THEOREM 6(iii). From above and (3.2), we know that all limit points of  $\{\mathbf{x}_N\}$  are self-consistent estimators in  $\Omega_n^+$ . From the uniqueness of  $\bar{S}_X^{(n)}$  obtained in Corollary 3, there is only one self-estimator in  $\Omega_n^+$ . Hence, the limit of every convergent subsequent of  $\{\mathbf{x}_N\}$  is  $\bar{S}_X^{(n)}$ . Therefore, algorithm (2.5) converges to  $\bar{S}_X^{(n)}$  for any initial point  $\mathbf{x}_1 \in \Omega_n^+$ .  $\square$

APPENDIX

PROOF OF THEOREM 1. Consider the proof of Lemma A2 of Turnbull (1974). It suffices to show that the matrix  $\mathbf{J}$  is positive definite. In our notation, we have

$$x_k = \frac{\alpha_k}{P_{k-1} - P_k} \quad \text{and} \quad y_k = \frac{\beta_k}{P_k^2} + \frac{\gamma_k}{(1 - P_k)^2},$$



and  $z_k = x_{k+1} + x_k + y_k$  for  $1 \leq k \leq m$  and  $x_{m+1} = 0$ . Let  $\mathbf{u}$  be an arbitrary vector and note that in this notation,

$$\mathbf{u}^\top \mathbf{J} \mathbf{u} = \sum_{k=1}^m z_k u_k^2 - 2 \sum_{k=1}^m x_k u_k u_{k-1} = \sum_{k=1}^m x_k (u_k - u_{k-1})^2 + \sum_{k=1}^m y_k u_k^2,$$

with  $u_0 = 0$ . Now assume that this expression is zero for some  $\mathbf{u} \neq \mathbf{0}$ . Let  $j$  be the biggest number such that  $u_j \neq 0$ . Under our assumption,  $y_j$  must be zero, and so  $x_j \neq 0$ , and so  $u_{j-1} = u_j$ . Go on with this until you hit  $u_0$ , which is zero. This is a contradiction. Hence,  $\mathbf{u}^\top \mathbf{J} \mathbf{u} > 0$ .  $\square$

**Acknowledgments.** The authors would like to thank Peter Bickel for useful discussions, and would also like to thank the Associate Editor and two referees for their comments on the manuscript.

## REFERENCES

- CHANG, M. N. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *Ann. Statist.* **18** 391–404.
- CHANG, M. N. and YANG, G. L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *Ann. Statist.* **15** 1536–1547.
- EFRON, B. (1967). The two sample problem with censored data. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **4** 831–853. Univ. California Press, Berkeley.
- GEHAN, E. A. (1965). A generalized two-sample Wilcoxon test for doubly censored data. *Biometrika* **52** 650–653.
- GU, M. G. and ZHANG, C. H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.* **21** 611–624.
- MANTEL, N. (1967). Ranking procedures for arbitrarily restricted observations. *Biometrics* **23** 65–78.
- ORTEGA, J. M. and RHEINBOLDT, W. C. (1970). *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York.
- PEER, P. G., VAN DIJCK, J. A., HENDRIKS, J. H., HOLLAND, R. and VERBEEK, A. L. (1993). Age-dependent growth rate of primary breast cancer. *Cancer* **71** 3547–3551.
- PETO, R. (1973). Experimental survival curves for interval-censored data. *Appl. Statist.* **22** 86–91.
- REN, J. (1994). On self-consistent estimators and kernel density estimators with doubly censored data. *J. Statist. Plann. Inference*. To appear.
- REN, J. (1995). Generalized Cramér–von Mises tests of goodness of fit for doubly censored data. *Ann. Inst. Statist. Math.* **47** 525–549.
- REN, J. and ZHOU, M. (1993).  $L$ -estimators and  $M$ -estimators with doubly censored data. *J. Nonparametric Statist.* To appear.
- 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.
- TURNBULL, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.* **69** 169–173.
- WU, J. F. (1983). On the convergence properties of the EM algorithm. *Ann. Statist.* **11** 95–103.
- ZANGWILL, W. I. (1967). *Nonlinear Programming: A Unified Approach*. Wiley, New York.

DEPARTMENT OF STATISTICS  
UNIVERSITY OF CHICAGO  
CHICAGO, ILLINOIS 60637

DEPARTMENT OF MATHEMATICS AND STATISTICS  
UNIVERSITY OF NEBRASKA  
LINCOLN, NEBRASKA 68588-0323