The Annals of Statistics
1985, Vol. 13, No. 4, 1317–1334

A LARGE SAMPLE STUDY OF GENERALIZED MAXIMUM LIKELIHOOD ESTIMATORS FROM INCOMPLETE DATA VIA SELF-CONSISTENCY

BY WEI-YANN TSAI AND JOHN CROWLEY[1]

Brookhaven National Laboratory and Fred Hutchinson Cancer Research Center and University of Washington

Self-consistent estimators for estimating distribution functions from incomplete data are presented. In many cases these estimators are also generalized maximum likelihood estimators. In this paper we discuss the theoretical properties of such estimators: existence, uniform consistency, law of the iterated logarithm, and weak convergence. Applications to the product limit estimator for right-censored data and to the estimator proposed by Turnbull (1974, 1976) for doubly (right- and left-) censored data are also given.

1. Introduction. The field of nonparametric estimation as yet lacks a unifying principle such as maximum likelihood, which guarantees asymptotic normality and optimality in a large class of parametric problems. Generalized maximum likelihood, proposed by Kiefer and Wolfowitz (1956), at least provides an algorithm for the calculation of estimators, but no general theory exists which establishes the large sample properties of such estimators. An example of a generalized maximum likelihood estimator (GMLE) is the product limit estimator of a distribution function in the presence of right-censored data. This was implicit in the derivation by Kaplan and Meier (1958) and was made explicit by Johansen (1978). The product limit (PL) estimator is also self-consistent, as Efron, who defined the concept, showed (1967). This fact provides a clue to a general asymptotic theory for the GMLE: The self-consistency equation can, in this case, be cast in terms of the EM algorithm (Dempster, Laird, and Rubin, 1977), which converges to the GMLE. If this relationship holds more widely, an analogy with the Newton–Raphson algorithm for finding the MLE in parametric problems suggests that asymptotic results for generalized maximum likelihood estimators might be derived by considering an appropriate linearization of the self-consistency equation. In Section 2 we review generalized maximum likelihood, self-consistency, and the EM algorithm, and establish some relationships between these concepts. Section 3 contains results on the differential of statistical functions and an implicit function theorem, which will be used in the expansion of the self-consistency equations. The expansion is used in Section 4 to derive some large sample properties of self-consistent estimators. The product limit

Received April 1983; revised May 1985.
[1]This research supported in part by NIH grants R01-CA-18332 and R01-GM-28314.
AMS 1980 *subject classifications.* Primary 62E20; secondary 62G05.
Key words and phrases. Generalized maximum likelihood estimator, self-consistency, incomplete data, implicit function theorem, uniform consistency, law of the iterated logarithm, weak convergence, product limit estimator, censored data.

estimator is used as an example throughout, to fix ideas; no new results are obtained for this familiar estimator. However, for the GMLE of a distribution function in the presence of doubly censored data, proposed by Turnbull (1974, 1976), no proofs of large sample properties exist. We apply our general theory to this problem in Section 5.

2. Generalized maximum likelihood, self-consistency, and the EM algorithm. We start by defining the GMLE and giving some of its properties. Kiefer and Wolfowitz (1956) suggested that for a nondominated family of probability measure $\mathscr{P}$ one can define a generalized maximum likelihood estimator as follows: For P_1, P_2 in $\mathscr{P}$, let $f(X; P_1, P_2) = (dP_1/d(P_1 + P_2))(X)$, the Radon–Nikodym derivative of P_1 with respect to $P_1 + P_2$. If X represents the observed data vector, $\hat{P}$ is a GMLE if and only if

$$f(X; \hat{P}, P) \geq f(X; P, \hat{P}) \quad \text{for all } P \text{ in } \mathscr{P}. \tag{2.1}$$

A distribution function is said to be a GMLE if the probability measure, which induces the distribution function, is a GMLE.

The following three properties are direct consequences of the definition:

(i) If the family of probability measures $\mathscr{P}$ has a dominating measure, then the GMLE reduces to the usual MLE.

(ii) The empirical cumulative distribution function $F_x^n(\mathbf{t}) = n^{-1}\sum_{i=1}^n I(\mathbf{X}_i \leq \mathbf{t})$ from a sample of iid random vectors $\mathbf{X}_i$ is a GMLE. (Here I is the indicator function and $\mathbf{X}_i \leq \mathbf{t}$ means every coordinate of $\mathbf{X}_i$ is less than or equal to the corresponding coordinate of $\mathbf{t}$.)

(iii) If $\hat{P}$ gives positive probability to the observation X, then (2.1) is the same as

$$\hat{P}(X) \geq P(X) \quad \text{for all } P \text{ in } \mathscr{P} \tag{2.2}$$

[see Johansen (1978)].

The definition of (2.1) involves the usual problem of nonuniqueness of the choice of density [see Scholz (1980)]. Since the density in the definition (2.2) is uniquely defined, in the present situation, we will show a GMLE is self-consistent in the sense of (2.2).

When a set of data contains some observations that are not completely specified, the EM algorithm [Dempster, Laird, and Rubin (1977)] is often used to compute maximum likelihood estimators. The original setting of the EM algorithm was for parametric distributions, but in the literature there are also a few examples of its use in a nonparametric context, such as the self-consistent estimators proposed by Efron (1967), Turnbull (1974, 1976), and Laird (1978). Dempster, Laird, and Rubin (1977) did not formulate the EM algorithm explicitly for the infinite-dimensional case, but, as we now show, the formulation can easily be extended.

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be iid $p \times 1$ random vectors with distribution function F_x. Let $\mathscr{X}$ and $\mathscr{Y}$ be two sample spaces and let $\mathscr{M} = \mathscr{M}(\mathbf{X}_i)$ be a many-to-one mapping from $\mathscr{X}$ to $\mathscr{Y}$. The observed data $\mathbf{Y}_i$ ($q \times 1$ vector), $i = 1, \ldots, n$, comprise a realization from $\mathscr{Y}$, while the corresponding $\mathbf{X}_i$, $i = 1, \ldots, n$, are observed only

indirectly through $\mathbf{Y}_i$ via $\mathscr{M}$. From the principle of the EM algorithm, an estimator of F_x can be described as follows.

Let $F_x^{(0)}$ be an initial estimate of F_x, and let $F_x^{(m)}$ denote the current estimate of F_x after m steps of the algorithm. Then the E step of the algorithm estimates $I(\mathbf{X}_i \le \mathbf{t})$ by $E_{F_x^{(m)}}(I(\mathbf{X}_i \le \mathbf{t})|\mathbf{Y}_i = \mathbf{y}_i)$. The M step consists in finding the GMLE based on the estimated data, which gives

$$
\begin{aligned}
F_x^{(m+1)}(\mathbf{t}) &= n^{-1}\sum_{i=1}^{n} E_{F_x^{(m)}}(I(\mathbf{X}_i \le \mathbf{t})|\,\mathbf{Y}_i = \mathbf{y}_i) \\
&= E_{F_x^{(m)}}\left[\left\{n^{-1}\sum_{i=1}^{n} I(\mathbf{X}_i \le \mathbf{t})\right\}|\,\mathbf{Y}_1 = \mathbf{y}_1, \ldots, \mathbf{Y}_n = \mathbf{y}_n\right].
\end{aligned} \tag{2.3}
$$

Suppose $F_x^{(0)}, F_x^{(1)}, \ldots,$ converges to a distribution function $\hat{F}_x^n$, which cannot be changed by application of (2.3). Such a function would satisfy

$$
\hat{F}_x^n(\mathbf{t}) = E_{\hat{F}_x^n}\left[\left\{n^{-1}\sum_{i=1}^{n} I(\mathbf{X}_i \le \mathbf{t})\right\}|\,\mathbf{Y}_1 = \mathbf{y}_1, \ldots, \mathbf{Y}_n = \mathbf{y}_n\right] \tag{2.4}
$$

for all $\mathbf{t}$, which is just Efron's (1967) property of self-consistency of an estimator of F_x. Thus if in this context the EM algorithm converges, the convergence is to a self-consistent estimator. In the remainder of this section we give conditions for convergence and conditions under which the resulting self-consistent estimator is a GMLE.

THEOREM 2.1. *If the initial estimator $F_x^{(0)}$ in the algorithm* (2.1) *is a step function with mass at the observed points $\mathbf{Y}_i$ (and possibly elsewhere), then the algorithm converges.*

PROOF. Under these conditions the algorithm (2.1) is exactly the EM algorithm for incomplete multinomial data. Since the multinomial is a member of the exponential family, condition (10) of Wu (1983) is satisfied. Therefore, Theorem 2 of Wu (1983) implies the convergence of this algorithm.

REMARK 2.1. It seems likely that convergence can be established for more general initial estimators $F_x^{(0)}$, but the conditions of Theorem 2.1 suffice for our purposes here.

Let $\mathscr{P}_x$ be a family of probability measures on the sample space $\mathscr{X}$. Let $\mathscr{P}_y$ be the corresponding family on the sample space $\mathscr{Y}$. For each mapping $\mathscr{M}$: $\mathscr{X} \to \mathscr{Y}$, there is a corresponding mapping $\mathscr{M}^*$: $\mathscr{P}_x \to \mathscr{P}_y$ such that for any random vectors $\mathbf{X}$ and $\mathbf{Y} = \mathscr{M}(\mathbf{X})$, $\mathscr{M}^*(P_x)$ is a probability measure of $\mathbf{Y}$ where P_x is a probability measure of $\mathbf{X}$.

Let $\mathscr{M}^*(\mathscr{P}_x) = \{P_y | P_y = \mathscr{M}^*(P_x), P_x \in \mathscr{P}_x\}$ and let P_y^n be the empirical probability measure induced by the empirical distribution $F_y^n(\mathbf{t}) = n^{-1}\sum_{i=1}^{n}(\mathbf{Y}_i \le \mathbf{t})$ of

the observed values of $\mathbf{Y}$. Then we have the following theorem:

THEOREM 2.2. *If $P_y^n \in \mathscr{M}^*(\mathscr{P}_x)$ then*
(i) *Any GMLE of F_x based on iid $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ is self-consistent, and*
(ii) *There exists a self-consistent estimator of F_x based on $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ which is a GMLE.*

PROOF. (i) Define the likelihood function L by $L(P_y) = P_y(\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$; $P_y \in \mathscr{M}^*(\mathscr{P}_x)$. We want to maximize L on $\mathscr{M}^*(\mathscr{P}_x)$. The global maximum is attained on $\mathscr{P}_y$ by $P_y = P_y^n$. If P_y^n happens to be in $\mathscr{M}^*(\mathscr{P}_x)$ then the set of GMLEs of P_x is just the set $\mathscr{M}^{*-1}(P_y^n)$; hence any $\hat{P}_x^n$ which satisfies $\mathscr{M}^*(\hat{P}_x^n) = P_y^n$ is a GMLE.

For any $Q_x \in \mathscr{P}_x$ and $Q_y = \mathscr{M}^*(Q_x)$ we have

$$\begin{aligned} Q_x\{\mathbf{X} \le \mathbf{t}\} &= E_{Q_x}(I(\mathbf{X} \le \mathbf{t})) \\ &= E_{Q_y}\big(E_{Q_x}(I(\mathbf{X} \le \mathbf{t})|\mathbf{Y})\big) \\ &= \int E_{Q_x}(I(\mathbf{X} \le \mathbf{t})|\mathbf{Y} = \mu) Q_y(d\mu). \end{aligned}$$

If Q_y happens to be P_y^n (the empirical measure), then we have $Q_x = \hat{P}_x^n$ and

$$\begin{aligned} \hat{P}_x^n\{\mathbf{X} \le \mathbf{t}\} &= \int E_{\hat{P}_x^n}(I(\mathbf{X} \le \mathbf{t})|\mathbf{Y} = \mu) P_y^n(d\mu) \\ &= \frac{1}{n}\sum_{i=1}^n E_{\hat{P}_x^n}(I(\mathbf{X}_i \le \mathbf{t})|\mathbf{Y} = \mathbf{Y}_i) \\ &= E_{\hat{P}_x^n}\left(\frac{1}{n}\sum_{i=1}^n I(\mathbf{X}_i \le \mathbf{t})|\mathbf{Y} = \mathbf{Y}_1, \ldots, \mathbf{Y}_n\right), \end{aligned}$$

which is the definition of self-consistency.

(ii) Since $P_y^n \in \mathscr{M}^*(\mathscr{P}_x)$, there exists a measure $\hat{P}_x^n$ such that $P_y^n = \mathscr{M}^*(\hat{P}_x^n)$, which is then self-consistent by the above argument. Since P_y^n is also a global maximum in P_y,

$$L\big(\mathscr{M}^*(Q_x)\big) \le L\big(P_y^n\big) = L\big(\mathscr{M}^*\big(\hat{P}_x^n\big)\big) \quad \text{for all } Q_x \text{ in } \mathscr{P}_x,$$

i.e., $\hat{P}_x^n$ is a GMLE.

We have thus established that any convergence of the EM algorithm is to a self-consistent estimator and have given a condition for convergence. If this estimator of F_x maps into the empirical distribution function F_y^n, it is also a GMLE. There are of course other situations in which the GMLE is self-consistent, and to which our general theory for self-consistent estimators will apply.

EXAMPLE. The product limit estimator. Let $X_1, \ldots, X_n$ be nonnegative random variables having distribution function F_x. As is common, we work with the survival function $S_x(t) = 1 - F_x(t) = P(X_i > t)$. Let $Z_1, \ldots, Z_n$ be censoring times, iid random variables independent of $X_1, \ldots, X_n$ and with common distri-

bution function (the random censorship model). The observations $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ consist of the n pairs $(Y_1, \Delta_1), \ldots, (Y_n, \Delta_n)$, where $Y_i = \min(X_i, Z_i)$ and $\Delta_i = I(Y_i = X_i)$ is an indicator of uncensored observations. Kaplan and Meier (1958) suggested an estimator $\hat{S}_x^n$ of S_x, defined by

$$\hat{S}_x^n(t) = \prod_{Y_i \le t} \left(1 - \frac{1}{\sum_{j=1}^n I(Y_j \ge Y_i)}\right)^{\Delta_i},$$

which was shown to be a GMLE by Johansen (1978). Efron (1967) showed that $\hat{S}_x^n$ is the unique solution to

$$\begin{aligned} S(t) &= n^{-1} E_S \sum_{i=1}^n (I(X_i > t) | \mathbf{Y}_i) \\ &= n^{-1} \sum_{i=1}^n I(Y_i > t) + n^{-1} \sum_{Y_i \le t} \frac{(1 - \Delta_i) S(t)}{S(Y_i)}, \end{aligned}$$

which is a special case of (2.4). If the largest observation $Y_{(n)}$ is censored then $\hat{S}_x^n(t)$ is defined arbitrarily [between $\hat{S}_x^n(Y_{(n)})$ and 0] for $t > Y_{(n)}$. Similarly, a product limit estimator $\hat{S}_z^n(t)$ of the survival function of the Z_is can be defined.

In this case, the space $\mathscr{X} = \{(X, Z) | X \ge 0, Z \ge 0\} = R^+ \times R^+$, the space $\mathscr{Y} = \{(Y, \Delta) | Y \ge 0, \Delta = 0 \text{ or } 1\} = R^+ \times \{0, 1\}$, and the mapping $\mathscr{M}$: $\mathscr{X} \to \mathscr{Y}$ is defined by $\mathscr{M}(X) = (\min(X, Z), \Delta = I(X \le z))$. The corresponding survival function spaces are $\mathscr{S}_x = \{S | S(s, t) = S(s, 0) S(0, t) = S_x(0) S_z(t)$, where S_x and S_z are survival functions in $R^+\}$ and $\mathscr{S}_y = \{S_y | S_y$ is a survival function with domain in $R^+ \times \{0, 1\}\}$; the corresponding mapping $\mathscr{M}^*$: $\mathscr{S}_x \to \mathscr{S}_y$ is defined for a survival function $S_z(t)$ by $\mathscr{M}^*(\mathscr{S}) =$

$$\begin{aligned} S_y(t, \delta) &= P(Y > t, \Delta \ge \delta) \\ &= \begin{cases} S_x(t) S_z(t) & \text{if } \delta = 0, \\ -\int_t^\infty S_z(u)\, dS_x(u) & \text{if } \delta = 1. \end{cases} \end{aligned}$$

It can readily be verified that $\mathscr{M}^*$ maps $\hat{S}_x^n(s, t) = \hat{S}_x^n(s) \hat{S}_z^n(t)$ into the empirical survival function of $(Y_1, \Delta_1), \ldots, (Y_n, \Delta_n)$, so that the condition of Theorem 2.2 is satisfied.

We now proceed to present some large sample theory for self-consistent estimators.

3. The differential of statistical functions. Before we describe the main theorems, we give some background. Very often functions of interest in statistics can be expressed as an operator $T(F_y)$ of the underlying population distribution function, in which case a natural sample analogue estimator is provided by $T(F_y^n)$, where F_y^n is the empirical distribution function of the sample. The functional representation of statistics was first studied in detail by von Mises (1947), was extended by Filippova (1962), and was applied by Boos (1979) to L estimators. In the problem of "robust" estimation, Hampel (1974) introduced the

"influence curve" of measuring robustness by exploiting Frechet differentiation. Later, Reid (1981) studied functions of the product limit estimator by using such influence curves.

Let $\mathscr{F}_q$ be the set of q-dimensional distribution functions and $\mathscr{D}_q$ be the linear space generated by differences $F_2 - F_1$ of members of $\mathscr{F}_q$. Let $\mathscr{D}_q$ be equipped with a norm $\|\cdot\|_{\mathscr{D}_q}$ and let B_1 be a normed vector space. Consider a mapping T of $\mathscr{F}_q$ into B_1.

DEFINITION 3.1. The mapping T is said to be differentiable at a point $F_1 \in \mathscr{F}_q$, if there is a linear mapping $T'(F_1; \cdot)$ of $\mathscr{D}_q$ into B_1 that satisfies

$$T(F_2) - T(F_1) = T'(F_1; F_2 - F_1) + o(\|F_2 - F_1\|\mathscr{D}_q) \quad \text{as } \|F_2 - F_1\|_{\mathscr{D}_q} \to 0. \tag{3.1}$$

In such a case, $T'(F_1; \cdot)$ is called the differential of T at F_1.

If T is differentiable with respect to the sup-norm $\|h\|_\infty = \sup_t |h(\mathbf{t})|$, then we can study the estimator $T(F_y^n)$ by using the Kolmogorov–Smirnov distance $\|F_y^n - F_y\|_\infty$ and (3.1). If F_y is a continuous function then $n^{1/2}\|F_y^n - F_y\|_\infty = O_p(1)$, which follows from Kiefer's (1961) inequality

$$P\left(\left\{n^{1/2}\|F_y^n - F_y\|_\infty > \lambda\right\}\right) \le c \exp\left(-(2 - \varepsilon)\lambda^2\right) \tag{3.2}$$

for each ε, $\lambda > 0$, where c is a universal constant depending only on ε and the dimension q of $\mathbf{Y}$. We obtain from (3.1), (3.2), and the linearity of T' that

$$n^{1/2}\left(T\left(F_y^n\right) - T\left(F_y\right)\right) = T'\left(F_y; n^{1/2}\left(F_y^n - F_y\right)\right) + o_p(1). \tag{3.3}$$

Note that $n^{1/2}(F_y^n - F_y) \to_D Y$, where Y is a Brownian sheet with $E(Y) = 0$ and $\mathrm{Cov}(Y(\mathbf{s}), Y(\mathbf{t})) = F_y(\mathbf{s} \wedge \mathbf{t}) - F_y(\mathbf{s})F_y(\mathbf{t})$ and where "$\to_D$" means convergence in distribution and $\mathbf{s} \wedge \mathbf{t} = (\min(s_1, t_1), \ldots, \min(s_q, t_q))$. If we assume that $T'(F_y; \cdot)$ is continuous, then from (3.3) and the invariance principle [Breiman (1968)], we obtain that $n^{1/2}(T(F_y^n) - T(F_y)) \to_D T'(F_y; X)$. Further, by using the fact that a continuous linear operator is also a bounded operator and by the LIL of F_y^n, it follows that if F_y is a continuous function then

$$\begin{aligned}
\left\|T\left(F_y^n\right) - T\left(F_y\right)\right\|_{B_1} &= o\left(\|F_y^n - F_y\|_\infty\right) + \left\|T'\left(F_y; F_y^n - F_y\right)\right\|_{B_1} \\
&\le o\left(\|F_y^n - F_y\|_\infty\right) + \|T'\|_{B_1}\|F_y^n - F_y\|_\infty \\
&= O\left(\|F_y^n - F_y\|_\infty\right) \\
&= O\left(\left((\log\log n)/n\right)^{1/2}\right) \quad \text{a.s. as } n \to \infty,
\end{aligned}$$

where $\|T'\|_{B_1}$ is the norm of T'. Since $T'(F, \cdot)$ is continuous if and only if T is continuous and differentiable at F by (3.1) [also see Brown and Page (1976), page 265], it would be reasonable to assume T is continuously differentiable at F.

In some situations the function $G = T(F_y)$ cannot be expressed explicitly in terms of the distribution function F_y, but only implicitly through an equation $H(F_y, G) = 0$. In this case, a naive estimator $\hat{G}^n$ of G could be the solution of $H(F_y^n, G) = 0$. As we now point out, the implicit function theorem gives sufficient

conditions for the existence, uniqueness and differentiability of local solutions $G = T(F_y)$.

DEFINITION 3.2. Let B_i, $i = 1,2,3$ be normed vector spaces and let H be a mapping from $B_1 \times B_2$ to B_3. The mapping H is said to be partially differentiable with respect to the first variable at the point (F_1, F_2), iff $g(F) = H(F, F_2)$ is differentiable at F_1. In this case we write $g'(F_1; \cdot) = H_1'(F_1, F_2; \cdot)$ and call H_1' the partial differential of H with respect to the first variable. Similar definitions can be made for the partial differential with respect to the second variable.

PROPOSITION 3.1 (Implicit function theorem). *Let H be a continuously differentiable mapping of a nonempty open set A of $B_1 \times B_2$ into B_3, where B_2 and B_3 are complete normed spaces. Suppose that, for some point $(F_1, F_2) \in A$, $H(F_1, F_2) = 0$, and the partial differential $H_2'(F_1, F_2; \cdot)$ has an inverse. Then there is a positive real number b and a continuously differentiable mapping T of the open ball $N_b = \{G \in B_1: \|G - F_1\|_{B_1} < b\}$ into B_2 such that*

(a) $(G, T(G)) \in A$ *for all* $G \in N_b$,

(b) $F_2 = T(F_1)$ *and* $H(G, T(G)) = 0$ *for all* $G \in N_b$ *and,*

(c) T *is continuously differentiable and,*

$$T'(F_1; \cdot) = -\left(H_2'(F_1, T(F_1); H_1'(F_1, T(F_1); \cdot))\right)^{-1}.$$

See Brown and Page (1976, pages 291–293) for a proof of this proposition.

If $H(F_y, G) = 0$, $H(F_y^n, \hat{G}^n) = 0$, and $\hat{G}^n$ is a consistent estimator of G, then by the implicit function theorem, there exists a continuously differentiable mapping T such that $G = T(F_y)$ and $\hat{G}^n = T(F_y^n)$ for large enough n; therefore, arguing as before, weak convergence and the LIL for $\hat{G}_n$ also hold.

4. Asymptotic properties of self-consistent estimators. We have seen that a self-consistent estimator satisfies

$$\begin{aligned}\hat{F}_x^n(\mathbf{t}) &= E_{\hat{F}_x^n}\left[\left\{n^{-1}\sum_{i=1}^{n} I(\mathbf{X}_i \le \mathbf{t})\right\} | \mathbf{Y}_1 = \mathbf{y}_1, \ldots, \mathbf{Y}_n = \mathbf{y}_n\right] \\ &= n^{-1}\sum_{i=1}^{n} E_{\hat{F}_x^n}\left(I(\mathbf{X}_i \le \mathbf{t}) | \mathbf{Y}_i = \mathbf{y}_i\right) \\ &= \int E_{\hat{F}_x^n}\left(I(\mathbf{X} \le \mathbf{t}) | \mathbf{Y} = \mu\right) dF_y^n(\mu).\end{aligned}$$

Similarly,

$$\begin{aligned}F_x(\mathbf{t}) &= E\left(I(\mathbf{X} \le \mathbf{t})\right) = E\left(E\left(I(\mathbf{X} \le \mathbf{t}) | \mathbf{Y}\right)\right) \\ &= \int E\left(I(\mathbf{X} \le \mathbf{t}) | \mathbf{Y} = \mu\right) dF_y(\mu).\end{aligned}$$

Write $H(F_y, G_x(\mathbf{t})) = -G_x(\mathbf{t}) + \int E_{G_x}(I(\mathbf{X} \le \mathbf{t}) | \mathbf{Y} = \mu)\, dF_y(\mu)$. Then we have the following lemma and theorem:

LEMMA 4.1. *If for every* ε, *there exists a left-continuous step function* $s(\mu, \mathbf{t})$ *such that* $\sup_\mu |s(\mu, \mathbf{t}) - E_{G_x}(I(\mathbf{X} \le \mathbf{t})|\mathbf{Y} = \mu)| \le \varepsilon$ *for every* $\mathbf{t}$, *then* $\|H(F_y^n, G_x) - H(F_y, G_x)\|_\infty \to 0$ *almost surely as* $n \to \infty$ *for every distribution function* G_x, *provided* $\|F_y^n - F_y\|_\infty \to 0$ *almost surely as* $n \to \infty$.

PROOF. Since $E_{G_x}(I(\mathbf{X} \le \mathbf{t})|\mathbf{Y} = \mu) \le 1$ for all μ and $\mathbf{t}$ then $S(\mu, \mathbf{t})$ is bounded for every μ and $\mathbf{t}$. The rest of proof is a multivariate generalization of the proof in Lemma 6.1 of Aalen (1976).

THEOREM 4.1. *Suppose the condition of Lemma* 4.1 *is satisfied. If* F_x *is the unique solution of* $H(F_y, G_x) = 0$ *and there exists an integer* m *such that for every sample of size* $n > m$, $H(F_y^n, G_x) = 0$ *has a solution* $\hat{F}_x^n$ (*which need not be unique*), *then* $\|\hat{F}_x^n - F_x\|_\infty \to 0$ *almost surely as* $n \to \infty$ *for every sequence of solutions* $\hat{F}_x^n$ *of* $H(F_y^n, G_x) = 0$.

PROOF. Since the solution to $H(F_y, G_x) = 0$ is unique, it follows that for any F_x^* not in the neighborhood $N_\varepsilon = \{G\colon \|G - F_x\|_\infty < \varepsilon\}$, $\|H(F_y, F_x^*)\|_\infty > 0$. By Lemma 4.1, $\|H(F_y^n, F_x^*) - H(F_y, F_x^*)\|_\infty \to 0$ almost surely as $n \to \infty$. Therefore, for almost all realizations, there exists an $m' > m$ (depending on the realization) such that for all $n > m'$, $\|H(F_y^n, F_x^*)\|_\infty > 0$. Thus any solution to $H(F_y^n, G_x) = 0$ is in N_ε. As ε is arbitrary, we have that $\|\hat{F}_x^n - F_x\|_\infty \to 0$ almost surely for any sequence of solutions of $H(F_y^n, G_x) = 0$.

REMARK 4.1. If there is a neighborhood of F_y for which $H(G_y, G_x) = 0$ has a solution for any G_y in the neighborhood, then there exists a sequence of solutions $\hat{F}_x^n$ of $H(F_y^n, G_x) = 0$ such that $\|\hat{F}_x^n - F_x\|_\infty \to 0$ almost surely as $n \to \infty$. This can be seen from the above argument.

REMARK 4.2. The uniqueness assumption of the solution to $H(F_y, G_x) = 0$ in Theorem 4.1 is associated with an identifiability condition. If there is more than one solution, then there are two or more different distribution functions which will produce the same F_y under the mapping $\mathcal{M}\colon \mathcal{X} \to \mathcal{Y}$. That means F_x cannot be identified through F_y for the observations $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$. We have implicitly assumed that $E(I(\mathbf{X}_i \le \mathbf{t})|Y_1, \ldots, Y_n)$ is a function of F_x and $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ only. Otherwise, there will not likely be a unique solution. For example, if in the censored data case the random variables X_i and Z_i are not independent, then $E(I(X_i \le t)|Y_i, \Delta_i)$ may be a function of the joint distribution of X_i and Z_i, and the solution is not unique. In this case the distribution of the observable random vector (Y, Δ) does not identify F_x uniquely (Tsiatis, 1978).

THEOREM 4.2. *The equation* $H(F_y^n, G_x) = 0$ *has a solution* $\hat{F}_x^n$.

PROOF. By Theorem 2.1 the EM algorithm (2.1) converges if the initial estimator is a step function, and the convergence is to a function satisfying

$$\hat{F}_x^n(\mathbf{t}) = \int E_{\hat{F}_x^n}(I(\mathbf{X} \le \mathbf{t})|\mathbf{Y} = \mu)\, dF_y^n(\mu).$$

Thus

$$0 = -\hat{F}_x^n + \int E_{\hat{F}_x^n}(I(\mathbf{X} \le \mathbf{t})|\mathbf{Y} = \mu)\, dF_y^n$$
$$= H(F_Y^n, \hat{F}_x^n).$$

EXAMPLE. Recall that the product limit estimator $\hat{S}_x^n$ of a survival function $S_x = 1 - F_x$ is based on data given by pairs (Y_i, Δ_i), $i = 1, \ldots, n$, where $Y_i = \min(X_i, Z_i)$ and $\Delta_i = I(Y_i = X_i)$. Efron (1967) showed that the product limit estimator is the unique solution [for $t < Y_{(n)}$] to the self-consistency equation, which in this case reduces to

$$S(t) = n^{-1} \sum_{i=1}^{n} I(Y_i > t) + n^{-1} \sum_{Y_i \le t} \frac{(1 - \Delta_i)S(t)}{S(Y_i)}$$
$$= S_u^n(t) + S_c^n(t) - \int_0^t \frac{S(t)}{S(y^-)}\, dS_c^n(y),$$

where

$$S_u^n(t) = n^{-1} \sum_{i=1}^{n} I(Y_i > t, \Delta_i = 1)$$

and

$$S_c^n(t) = n^{-1} \sum_{i=1}^{n} I(Y_i > t, \Delta_i = 0).$$

Let $S_u(t) = P(Y_i > t, \Delta_i = 1)$, $S_c(t) = P(Y_i > t, \Delta_i = 0)$, $S_y(t) = S_u(t) + S_c(t)$, and $S_Y^n(t) = S_u^n(t) + S_c^n(t)$. Define $H(S_u, S_c, S)(t)$ to be $-S(t) + S_u(t) + S_c(t) - \int_0^t (S(t)/S(y^-))\, dS_c(y)$. Then we have

LEMMA 4.2. *If the survival function S_x and the censoring distribution function have no common discontinuities, then for any $U < \infty$ such that $S_y(U) > 0$, the equation $H(S_u, S_c, S)(t) = 0$ has unique solution $S_x(t)$ for $t \le U$.*

PROOF. To see that $S_x(t)$ satisfies $H(S_u, S_c, S_x)(t) = 0$, write

$$S_x(t) = E(E(I(X > \mathbf{t})|\mathbf{Y}_1, \ldots, \mathbf{Y}_n))$$
$$= E(E(I(X_i > t)|Y_i, \Delta_i))$$
$$= -\int_0^\infty E(I(X_i > t)|Y_i = y, \Delta_i = 1)\, dS_u(y)$$
$$- \int_0^\infty E(I(X_i > t)|Y_i = y, \Delta_i = 0)\, dS_c(y)$$
$$= -\int_0^\infty I(X_i > t)\, dS_u(y) - \int_0^\infty E(I(X_i > t)|Y_i = y, \Delta_i = 0)\, dS_c(y)$$
$$- \int_0^t E(I(X_i > t)|Y_i = y, \Delta_i = 0)\, dS_c(y)$$
$$= S_u(t) + S_c(t) - \int_0^t \frac{S_x(t)}{S_x(y^-)}\, dS_c(y).$$

Uniqueness of the solution, for $t \le U$, follows from the fact that for any $\varepsilon > 0$, and step functions S_u^ε and S_c^ε having no common discontinuities such that

$\sup_{0 \le t \le U} |S_u^\varepsilon(t) - S_u(t)| < \varepsilon$ and $\sup_{0 \le t \le U} |S_c^\varepsilon(t) - S_c(t)| < \varepsilon$, the result of Efron (1967) implies that there is a unique solution of $H(S_u^\varepsilon, S_c^\varepsilon, S)(t) = 0$, for $t \le U$. It can also be proved that

$$\sup_{0 \le t \le U} |H(S_u^\varepsilon, S_c^\varepsilon, S)(t) - H(S_u, S_c, S)(t)| \to 0 \quad \text{as} \quad \sup_{0 \le t \le U} |S_u^\varepsilon(t) - S_u(t)| \to 0$$

and $\sup_{0 \le t \le U} |S_c^\varepsilon(t) - S_c(t)| \to 0$. Therefore, the unique solution of $H(S_u^\varepsilon, S_c^\varepsilon, S) = 0$ implies the unique solution of $H(S_u, S_c, S) = 0$. [For a direct proof see Tsai (1986).]

This gives us:

THEOREM 4.3. *Under the conditions of Lemma* 4.2, *the product limit estimator*

$$\hat{S}_x^n(t) = \prod_{Y_i \le t} \left(1 - \frac{1}{\sum_{j=1}^{n} I(Y_j \ge Y_i)} \right)^{\Delta_i}$$

converges almost surely to $S_x(t)$, *uniformly* (*for* $t \le U$) *as* $n \to \infty$.

PROOF. Apply Lemma 4.2 and Theorem 4.1.

REMARK 4.3. Theorem 4.3 can be extended to cover the case where the censoring times are fixed arbitrary constants provided the empirical subsurvival functions S_u^n and S_c^n converge uniformly to the functions $\bar{S}_u(t) = (1/n)\sum_{i=1}^n P(X_i > t, \Delta_i = 1)$ and $\bar{S}_c(t) = (1/n)\sum_{i=1}^n P(X_i > t, \Delta_i = 0)$, respectively. It also seems likely that the theorem can be extended to cover convergence on the entire half line if there is no $U < \infty$ such that $S_y(U) = 0$, but this is less important in practice. We return to the general theory, and present results on weak convergence and the law of the iterated logarithm.

THEOREM 4.4. *Let* $\hat{F}_x^n$ *be a solution of* $H(F_y^n, \hat{F}_x^n) = 0$. *If* H *satisfies the conditions of the implicit function theorem and Theorem* 4.1 (*or the Remark* 4.1), *then*

(a) $n^{1/2}(\hat{F}_x^n(\mathbf{t}) - F_x(\mathbf{t})) \to {}_D T^*(F_y, F_x; Y)$, *where* Y *is a Brownian sheet with*

$$E(Y) = \mathbf{0} \quad \text{and} \quad \operatorname{Cov}(Y(\mathbf{s}), Y(\mathbf{t})) = F_y(\mathbf{s} \wedge \mathbf{t}) - F_y(\mathbf{s})F_y(\mathbf{t}) \tag{4.2}$$

and

$$T^*(F_y, F_x; Y) = -\left(H_2'\left(F_y, F_x; H_1'(F_y, F_x; Y)\right)\right)^{-1}.$$

(b) *Assume also that* F_y *is a continuous function. Then* $\|\hat{F}_x^n - F_x\|_\infty = O(((\log\log n)/n)^{1/2})$ *almost surely as* $n \to \infty$.

PROOF. By applying the implicit function theorem and the fact that $\|\hat{F}_x^n - F_x\|_\infty \to 0$ almost surely as $n \to \infty$, which follows from Theorem 4.1, there exists a continuously differentiable mapping T, such that $\hat{F}_x^n = T(F_y^n)$, $F_x = T(F_y)$, and $T'(F_y; \cdot) = T^*(F_y, F_x; \cdot)$. Thus the results follow from the arguments in Section 3.

COROLLARY 4.1. *Under the same conditions of Theorem* 4.4,

$$n^{1/2}\big(H_2'(F_y, F_x; \hat{F}_x^n) - H_2'(F_y, F_x; F_x)\big) \to_D - H_1'(F_y, F_x; Y),$$

where Y *is defined in* (4.2).

PROOF. By the invariance principle and Theorem 4.4, we obtain

$$\begin{aligned} n^{1/2}\big(H_2'(F_y, F_x; \hat{F}_x^n) - H_2'(F_y, F_x; F_x)\big) &= H_2'\big(F_y, F_x; n^{1/2}(\hat{F}_x^n - F_x)\big) \\ &\to_D H_2'\big(F_y, F_x; T^*(F_y, F_x; Y)\big) \\ &= -H_1'(F_y, F_x; Y). \end{aligned}$$

REMARK 4.4. In order to guarantee some asymptotic properties in the parametric setting, some smooth conditions are required either on the likelihood function or on the normal equation. (4.1) serves as a normal equation for self-consistent estimators (or for GMLEs, if the GMLE is a self-consistent estimator). The smooth conditions in Theorem 4.4 can be viewed as an analog to the smooth conditions for the MLE, guaranteeing certain asymptotic results. And Corollary 4.1 can be viewed as an analog to the asymptotic relationship between the MLE and the score statistics (cf. Theorem 5f.2(ii) in Rao, 1974, page 365). For the product limit estimator $\hat{S}_x^n$ these results take the following form:

LEMMA 4.3. *Under the conditions of Lemma* 4.2, *the function*

$$H(S_u, S_c, S)(t) = -S(t) + S_u(t) + S_c(t) - \int_0^t \frac{S(t)}{S(y^-)}\, dS_c(y)$$

has the following properties:

(a) H *is continuously differentiable at* (S_u, S_c, S_x) *and*

$$H_1'(S_u, S_c, S_x; S_u^n - S_u)(t) = (S_u^n - S_u)(t),$$

$$H_2'(S_u, S_c, S_x; S_c^n - S_c)(t) = (S_c^n - S_c)(t) - \int_0^t \frac{S_x(t)}{S_x - (y^-)}\, d(S_c^n - S_c)(y),$$

and

$$\begin{aligned} H_3'(S_u, S_c, S_x; \hat{S}_x^n - S_x)(t) = &-(\hat{S}_x^n - S_x)(t) - \int_0^t \frac{(\hat{S}_x^n - S_x)(t)}{S_x(y^-)}\, dS_c(y) \\ &+ \int_0^t \frac{S_x(t)(\hat{S}_x^n - S_x)(y)}{S_x^2(y^-)}\, dS_c(y). \end{aligned}$$

(b) H_3' *has an inverse. Furthermore*

$$\begin{aligned} &-\big(H_3'(S_u, S_c, S_x; H_1'(S_u, S_c, S_x; S_u^n - S_u) + H_2'(S_u, S_c, S_x; S_c^n - S_c))\big)^{-1}(t) \\ &\quad = -S_x(t)\Big\{-\int_0^t 1/(S_u + S_c)(y^-)\, d(S_u^n - S_u)(y) \\ &\qquad + \int_0^t (S_u^n - S_u)(y^-)/(S_u + S_c)^2(y^-)\, dS_u(y) \\ &\qquad + \int_0^t (S_c^n - S_c)(y^-)/(S_u + S_c)^2(y^-)\, dS_u(y)\Big\}. \end{aligned}$$

PROOF. See the appendix.

By integration by parts and the relation $S_y = S_u + S_c$ we have

$$n^{1/2}\{H_3'(S_u, S_c, S_x; H_1'(S_u, S_c, S_x; S_u^n - S_u) + H_2'(S_u, S_c, S_x; S_c^n - S_c))\}^{-1}(t)$$
$$= -S_x(t)\left\{-n^{1/2}(S_u^n - S_u)(t)/S_y(t) - \int_0^t n^{1/2}\frac{S_u^n(x) - S_u(x)}{S_y^2(x)}\,dS_y(x-)\right.$$
$$\left. + \int_0^t n^{1/2}\frac{S_y^n(x) - S_y(x)}{S_y^2(x-)}\,dS_u(x^-)\right.$$

which is $S_x(t)$ times the terms $(A_n + B_n)$ in (7.9) of Breslow and Crowley (1974). Since Lemma 4.3 implies that H satisfies the conditions of the implicit function theorem, by Theorem 4.4 and Lemma 4.2, we have the following theorem:

THEOREM 4.5. *Under the conditions of Lemma* 4.2, *for any* U *such that* $S_y(U) > 0$,

(a) $n^{1/2}(\hat{S}_x^n - S_x)(t)$ *converges weakly to a Gaussian process* $Z(t)$ *for* t *in* $[0, U]$ *with* $E(Z) = 0$ *and* $\mathrm{Cov}(Z(s), Z(t)) = S_x(s)S_x(t)\int_0^{s\wedge t} S_y^{-2}\,dS_u$,

(b) $\sup_{0\le t\le U}|\hat{S}_x^n - S_x| = O(((\log\log n)/n)^{1/2})$ *almost surely as* $n \to \infty$.

Weak convergence of the product limit estimator has been studied by Breslow and Crowley (1974) under the random censorship model, by Meier (1977) under the fixed censorship model, and by Aalen (1976, 1978) under a competing risks model using counting process approach. The above authors either assume that both S_x and G are continuous in the random censorship model or S_x is continuous in the fixed censorship model to establish the weak convergence of $\hat{S}_x^n$. Under this continuity assumption and the random censorship model, Földes and Rejtö (1981) proved the uniform (sup norm) consistency of the product limit estimator with rate factor $O(((\log n)/n)^{1/2})$, while Burke, Csörgö, and Harvath (1981) proved uniform consistency with rate $O(((\log\log n)/n)^{1/2})$. Gill (1983) recently extended some of these results to the entire half-line.

5. The self-consistent estimator of doubly censored data. Let $X_1, \ldots, X_n$ be iid random variables, having $S_x(t) = P(X_i > t)$ as their common survival function. Let $(L_1, R_1), \ldots, (L_n, R_n)$, where $L_i \le R_i$ for all $i = 1, \ldots, n$, can either be iid random vectors with common survival function or be fixed constant censoring times, though we will cover only the former case here. The observations are the n pairs $(Y_1, D_1), \ldots, (Y_n, D_n)$, where $Y_i = \max(\min(X_i, R_i), L_i)$ and

$$D_i = \begin{cases} 0 & \text{if } Y_i = L_i \\ 1 & \text{if } Y_i = X_i \\ 2 & \text{if } Y_i = R_i. \end{cases}$$

Turnbull (1974, 1976) proposed a self-consistent estimator $\hat{S}_x^n$ of S_x that satisfies

$$\hat{S}_x^n(t) = S_u^n(t) + S_r^n(t) + S_l^n(t) + \sum_{Y_i\le t} I(D_i = 2)\hat{S}_x^n(t)/\hat{S}_x^n(Y_i) - \sum_{Y_i\le t} I(D_i = 0)(1 - \hat{S}_x^n(t))/(1 - \hat{S}_x^n(Y_i)), \tag{5.1}$$

where $S_u(t) = P(Y_i > t, D_i = 1)$, $S_l(t) = P(Y_i > t, D_i = 0)$, and $S_r(t) = P(Y_i > t, D_i = 2)$ and S_u^n, S_l^n, and S_r^n are the empirical subsurvival functions of S_u, S_l, and S_r, respectively. Turnbull studied the uniqueness, consistency, and weak convergence of $\hat{S}_x^n$ with grouped data. The consistency and weak convergence of $\hat{S}_x^n$ with ungrouped data are established in this section. In fact, the proofs of the following corollaries all essentially follow the lines of the proofs in the last section, so we omit the proofs here.

COROLLARY 5.1. *For any real numbers L and R with $0 < L < R < \infty$ such that $S_r(R) > 0$ and $S_l(L) < 1$, and $((L_1, R_1), \ldots, (L_n, R_n))$ iid with distribution function having no common discontinuities then*

$$S_x(t) = S_l(t) + S_u(t) + S_r(t) - \int_0^t S_x(t)/S_x(y)\, dS_r(y^-) + \int_t^\infty (1 - S_x(t))/(1 - S_x(y^-))\, dS_l(y).$$

Let $\mathbf{S} = (S_u, S_l, S_r, S_x)$ *and*

$$H^*(S_u, S_l, S_r, S_x)(t) = -S_x(t) + S_l(t) + S_r(t) + S_u(t) - \int_0^t S_x(t)/S_x(y)\, dS_r(y^-) + \int_t^\infty (1 - S_x(t))/(1 - S_x(y^-))\, dS_l(y).$$

COROLLARY 5.2. *Under the conditions of Corollary* 5.1, $H^*(S_u, S_l, S_r, G_x) = 0$ *has a unique solution* $G_x(t) = S_x(t)$ *for* $L \le t \le R$.

COROLLARY 5.3. *Under the conditions of Corollary* 5.1, $\sup_{L \le t \le R} |\hat{S}_x^n(t) - S_x(t)| \to 0$ *almost surely as* $n \to \infty$, *where* $\hat{S}_x^n$ *is defined in* (5.1).

COROLLARY 5.4. *Under the conditions of Corollary* 5.1, H^* *is continuously differentiable at* $\mathbf{S}$ *and* $H_1^{*\prime}(S; S_u^n - S_u) = S_u^n - S_u$,

$$H_2^{*\prime}(\mathbf{S}; S_l^n - S_l)(t) = (S_l^n - S_l)(t) + \int_t^\infty (1 - S_x(t))/(1 - S_x(y^-))\, d(S_l^n - S_l)(y),$$

$$H_3'(\mathbf{S}; S_r^n - S_r)(t) = (S_r^n - S_r)(t) + \int_0^t S_x(t)/S_x(y)\, d(S_r^n - S_r)(y),$$

$$H_4'(\mathbf{S}, S_x^n - S_x)(t) = -(S_x^n - S_x)(t) - \int_0^t (S_x^n - S_x)(t)/S_x(y)\, dS_r(y) + \int_0^t S_x(t)(S_x^n - S_x)(y)/S_x^2(y)\, dS_r(y) - \int_t^\infty (S_x^n - S_x)(t)/(1 - S_x(y))\, dS_l(y) + \int_t^\infty (1 - S_x(t))(S_x^n - S_x)(y)/(1 - S_x(y))^2\, dS_l(y).$$

COROLLARY 5.5. *If the conditions of Corollary* 5.1 *hold and* $H_4^{*\prime}$ *has an inverse mapping, then*

(a) $n^{1/2}(\hat{S}_x^n - S_x)(t)$ *converges weakly to a Gaussian process* Z *for* $L \le t \le R$,

(b) $\sup_{L \le t \le R} |\hat{S}_x(t) - S_x(t)| = O(((\log\log n)/n)^{1/2})$ *almost surely as* $n \to \infty$.

We have not been able to prove the existence of an inverse mapping of $H_4^{*\prime}$. Even if $(H_4^{*\prime})^{-1}$ exists, we do not have an explicit form for $(H_4^{*\prime})^{-1}$, therefore the covariance structure Z cannot be obtained directly. The following approach may prove helpful in such cases.

Let $Z = T^*(F_y, F_x; Y)$, where T^* is defined in (4.2), with covariance structure $V(\mathbf{s}, \mathbf{t}) = \mathrm{Cov}(Z(\mathbf{s}), Z(\mathbf{t}))$. By Corollary 4.1, we have $H_2^{*\prime}(F_y, F_x; Z) = -H_1^{*\prime}(F_y, F_x; Y)$. Since the covariance structure of Y is a function of F_x, the covariance structure of $H_1^{*\prime}(F_y, F_x; Y)$ is a function of F_x and F_y. Let $G(F_y, F_x)(\mathbf{s}, \mathbf{t}) = \mathrm{Cov}(H_1^{*\prime}(F_y, F_x; Y)(\mathbf{s}), H_1^{*\prime}(F_y, F_x; Y)(\mathbf{t}))$, which can be derived from (4.2), and H^*. Furthermore, the covariance structure of $H_2^{*\prime}(F_y, F_x; Z)$ is a function of F_y, F_x, and V, say $\psi(F_y, F_x, V)(\mathbf{s}, \mathbf{t})$. Thus we have an implicit functional equation for V,

$$
\begin{aligned}
\Psi(F_y, F_x, V)(\mathbf{s}, \mathbf{t}) &= \mathrm{Cov}\big(H_2^{*\prime}(F_y, F_x; Z)(\mathbf{s}), H_2^{*\prime}(F_y, F_x; Z)(\mathbf{t})\big) \\
&= \mathrm{Cov}\big(H_1^{*\prime}(F_y, F_x; Y)(\mathbf{s}), H_1^{*\prime}(F_y, F_x; Y)(\mathbf{t})\big) \qquad (5.2) \\
&= G(F_y, F_x)(\mathbf{s}, \mathbf{t}).
\end{aligned}
$$

The function V can be solved from (5.2). An estimator $\hat{V}$ of V would be the solution of $\Psi(F_y^n, \hat{F}_x^n; \hat{V}) = G(F_y^n, \hat{F}_x^n)$. The consistency of $\hat{V}$ can be established by showing that $\Psi(F_y^n, \hat{F}_x^n, V) \to \Psi(F_y, F_x, V)$, and $G(F_y^n, \hat{F}_x^n) \to G(F_y, F_x)$. The conditions needed to prove consistency in general are currently under investigation.

APPENDIX

PROOF OF LEMMA 4.3. By using the formula

$$
\int_0^t F(u^-)\, dG(u) = F(u^-)G(u)\Big|_0^t - \int_0^t G(u)\, dF(u^-),
$$

we find

$$
\begin{aligned}
\int_0^t \frac{1}{S_x(y^-)}\, d(S_c^n(y) - S_c(y)) &= \frac{1}{S_x(y^-)}(S_c^n(y) - S_c(y))\Big|_0^t \\
&\quad - \int_0^t (S_c^n(y) - S_c(y))\, d\frac{1}{S_x(y^-)}.
\end{aligned}
$$

Define

$$
\|f(t)\| = \sup_{0 \le t \le U} |f(t)|.
$$

Now

$$\left\| \int_0^t (S_c^n(y) - S_c(y))\, d\frac{1}{S_x(y^-)} \right\| \le \left\| \int_0^t \frac{(S_c^n(y) - S_c(y))}{S_x^2(y)}\, dS_x(y^-) \right\|,$$

and this gives

$$\begin{aligned}
&\left\| \int_0^t \hat{S}_x^n(t)/\hat{S}_x^n(y^-)\, dS_c^n(y) - \int_0^t S_x(t)/S_x(y^-)\, dS_c(y) \right\| \\
&\quad \le \left\| \int_0^t \left(\frac{\hat{S}_x^n(t)}{\hat{S}_x^n(y)} \right) - \left(\frac{S_x(t)}{S_x(y^-)} \right) dS_c^n(y) \right\| + \left\| \int_0^t \frac{S_x(t)}{S_x(y^-)}\, d(S_c^n(y) - S_c(y)) \right\| \\
&\quad \le \|\hat{S}_x^n - S_x\|\, \|1/\hat{S}_x^n\| + \|\hat{S}_x^n - S_x\|\, \|1/S_x\| + \|S_c^n - S_c\| \\
&\qquad + \|S_x\|\, |S_c^n(0) - S_c(0)| + \|S_x\|\, \|1/S_x\|\, \|S_c^n - S_c\|.
\end{aligned}$$

Since $S_y(U) > 0$ implies $S_x(U) > 0$, we have $\|1/S_x\| = 1/S_x(U) < \infty$, and $\|S_x\| = S_x(0) = 1$. Therefore $\|H(S_u^n, S_c^n, \hat{S}_x^n) - H(S_u, S_c, S_x)\| \to 0$ as $\max(\|S_u^n - S_u\|, \|S_c^n - S_c\|, \|\hat{S}_x^n - S_x\|) \to 0$, for every subsurvival function S_u^n, S_u, S_c^n, and S_c, and for every survival function $\hat{S}_x^n$, S_x. This proves H is a continuous mapping. Since H is linear in its first and second variables, the derivation of the first and second partial derivatives is straightforward, and

$$\begin{aligned}
&\|H(S_u, S_c, \hat{S}_x^n) - H(S_u, S_c, S_x) - H_3'(S_u, S_c, \hat{S}_x^n - S_x)\| \\
&\quad = \left\| (\hat{S}_x^n - S_x)(t) \int_0^t \left(\frac{1}{\hat{S}_x^n(y^-)} - \frac{1}{S_x(y^-)} \right) dS_c(y) \right. \\
&\qquad \left. + S_x(t) \int_0^t \frac{(\hat{S}_x^n - S_x)(y^-)}{S_x(y^-)} \left(\frac{1}{\hat{S}_x^n(y^-)} - \frac{1}{S_x(y^-)} \right) dS_c(y) \right\| \\
&\quad \le \|\hat{S}_x^n - S_x\|\, \|1/\hat{S}_x^n - 1/S_x\| + \|S_x\|\, \|\hat{S}_x^n - S_x\|\, \|1/S_x\|\, \|1\|/\hat{S}_x^n - 1/S_x\| \\
&\quad = o(\|\hat{S}_x^n - S_x\|).
\end{aligned}$$

Since $(S_u, S_c, S_x) \to H_1'(S_u, S_c, S_x; \cdot)$ and $(S_u, S_c, S_x) \to H_2'(S_u, S_c, S_x; \cdot)$ are continuous mappings and all partial derivatives exist, Theorem 7.4.3 of Brown and Page (1976, pages 284–285) implies H is differentiable. Therefore, (a) is proved.

(b) In order to prove that H_3' has an inverse mapping, we have to show that $H_3'(S_u, -S_c, S_x; g)(t) = 0$ has a unique solution $g(t) = 0$. But $H_3'(S_u, S_c, S_x; g) = 0$ implies the following integral equation:

$$0 = g(t) + \int_0^t \frac{S_x(t) g(y^-)}{S_x^2(y^-) \left(1 + \int_0^t \frac{1}{S_x(u)}\, dS_c(u) \right)}\, dS_c(y),$$

which is the Homogeneous Volterra Equation and is known to have exactly one solution $g(t) = 0$. Therefore H_3' has an inverse mapping [see proof of Theorem 6

of Hochstadt (1973, page 33)]. Since

$$0 = -S_x(t) + S_c(t) + S_u(t) - S_x(t)\int_0^t \frac{1}{S_x(y)}\, dS_c(y), \tag{A.1}$$

we have

$$\begin{aligned} 0 &= -dS_x(t) + dS_c(t) + dS_u(t) - (dS_x(t))\int_0^t \frac{1}{S_x(y)}(dS_c(y) - dS_c(t)) \\ &= \frac{dS_x(t)}{S_x(t)}\left(S_x(t) + \int_0^t \frac{S_x(t)}{S_x(y)}\, dS_c(y)\right) + dS_u(t). \end{aligned}$$

Thus

$$\frac{1}{S_x}\, dS_x = \frac{1}{S_u + S_c}\, dS_u, \tag{A.2}$$

which has been established for the random censoring model. To prove the rest of (b), we have

$$\begin{aligned} &-H_3'\left(S_u, S_c, S_x; -S_x(t)\int_0^t \frac{(S_c^n - S_c)(y^-)}{(S_u + S_c)^2(y^-)}\, dS_u(y)\right) \\ &= -S_x(t)\int_0^t \frac{(S_c^n - S_c)(y^-)}{(S_u + S_c)^2(y^-)}\, dS_u(y) - \int_0^t\int_0^t \frac{S_x(t)(S_c^n - S_c)(s^-)}{S_x(y^-)(S_u + S_c)^2(s^-)}\, dS_u(s)\, dS_c(y) \\ &\quad + \int_0^t\int_0^y \frac{S_x(t)(S_c^n - S_c)(s^-)}{S_x(y^-)(S_u + S_c)^2(s^-)}\, dS_u(s)\, dS_c(y) \\ &= -S_x(t)\int_0^t \frac{(S_c^n - S_c)(y^-)}{(S_u + S_c)^2(y^-)}\, dS_u(y) \\ &\quad - \int_0^t \frac{S_x(t)(S_c^n - S_c)(y^-)}{(S_u + S_c)^2(y^-)S_x(y^-)}\int_0^y \frac{S_x(y^-)}{S_x(s^-)}\, dS_c(s)\, dS_u(y) \\ &= -\int_0^t \frac{S_x(t)(S_c^n - S_c)(y^-)}{S_x(y^-)(S_u + S_c)(y^-)}\, dS_u(y) \quad \text{by (A.1)} \\ &= -\int_0^t \frac{S_x(t)(S_c^n - S_c)(y^-)}{S_x^2(y^-)}\, dS_x(y) \quad \text{by (A.2)} \\ &= H_2'(S_u, S_c, S_x; S_c^n - S_c). \end{aligned}$$

Similarly,

$$\begin{aligned} &-H_3'\left(S_u, S_c, S_x; S(x)\left[\int_0^t \frac{(S_u^n - S_u)(y^-)}{(S_u + S_c)^2(y^-)}\, dS_u(y)\right.\right. \\ &\qquad\qquad \left.\left. + \int_0^t \frac{1}{(S_u + S_c)(y^-)}\, d(S_u^n - S_u)(y)\right]\right) \\ &\qquad = H_1'(S_u, S_c, S_x; S_u^n - S_u). \end{aligned}$$

Acknowledgment. The authors would like to thank the referees for their careful reading of this paper. We would also like to thank Søren Johansen and Richard Gill for many helpful comments and suggestions.

Note Added in Proof. Richard Gill has pointed out a technical difficulty with Lemma 4.3(a). A Correction Note will appear in a future issue.

REFERENCES

AALEN, O. (1976). Nonparametric inference in connection with multiple decrement models. *Scand. J. Statist.* **3** 15–27.

AALEN, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726.

BOOS, D. D. (1979). A differential for *L*-statistics. *Ann. Statist.* **7** 955–959.

BREIMAN, L. (1968). *Probability.* Addison-Wesley, Reading, Mass.

BRESLOW, N. E. and CROWLEY, J. (1974). A large-sample study of the life table and product limit estimates under random censorship. *Ann. Statist.* **2** 437–453.

BROWN, A. L. and PAGE, A. (1970). *Elements of Functional Analysis.* Van Nostrand Reinhold, New York.

BURKE, M. D., CSÖRGÖ, S. and HARVATH, L. (1981). A strong approximation for some biometric estimates under random censorship. *Z. Warsch. verw. Gebiete.* **56** 87–112.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* **39** 1–38.

EFRON, B. (1967). The two sample problem with censored data. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **4**, 831–883.

FILIPPOVA, A. A. (1962). von Mises' theorem on the asymptotic behavior of functionals of empirical distribution functions. *Theor. Probab. Appl.* **7** 25–57.

FÖLDES, A. and REJTÖ, L. (1981). Strong uniform consistency for nonparametric survival curve estimators from randomly censored data. *Ann. Statist.* **9** 122–129.

GILL, R. (1983). Convergence of the product limit estimator on the entire half line. *Ann. Statist.* **11** 49–58.

HAMPEL, F. R. (1974). The inluenfce curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393.

HOCHSTADT, H. (1973). *Integral Equations.* Wiley, New York.

JOHANSEN, S. (1978). The product limit estimator as maximum likelihood estimator. *Scand. J. Statist.* **5** 195–199.

KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

KIEFER, J. (1961). On large deviations of the empirical D.F. of vector chance variables and a law of interated logarithm. *Pacific J. Math.* **11** 649–660.

KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887–906.

LAIRD, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *J. Amer. Statist. Assoc.* **73** 805–815.

MEIER, P. (1977). Estimation of a distribution function from incomplete observations. *Perspectives in Probability and Statistics* (J. Gani, ed.) 67–87. Academic Press, New York.

RAO, L. R. (1973). *Linear Statistical Inference and its Applications.* 2nd ed. Wiley, New York.

REID, N. (1981). Influence functions for censored data. *Ann. Statist.* **9** 78–92.

SCHOLZ, F. W. (1980). Towards a unified definition of maximum likelihood. *Canad. J. Statist.* **8** 193–203.

TSAI, W. Y. (1986). Estimation of survival curves from dependent censorship models via a generalized self-consistent property with nonparametric Bayesian estimation application. To appear in *Ann. Statist.*

TSIATIS, A. A. (1978). 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 survivalship function with doubly censored data. *J. Amer. Statist. Assoc.* **69** 169–173.

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

VON MISES, R. (1947). On the asymptotic distributions of differentiable statistical functions. *Ann. Math. Statist.* **18** 309–348.

WU, C.-F. (1983). On the convergence properties of the EM algorithm. *Ann. Statist.* **11** 95–103.

BROOKHAVEN NATIONAL LABORATORY
UPTON, NEW YORK 11973

FRED HUTCHINSON CANCER RESEARCH CENTER
ZD-08
1124 COLUMBIA STREET
SEATTLE, WASHINGTON 98104