Asymptotic Properties of Nonparametric Estimation Based on Partly Interval-Censored Data

Jian Huang

Department of Statistics and Actuarial Science

University of Iowa, Iowa City, IA 52242

Email: jian@stat.uiowa.edu

Phone: (319) 335-0823   Fax: (319) 335-3017

Abstract: We study asymptotic properties of the nonparametric maximum likelihood estimator (NPMLE) of a distribution function based on partly interval-censored data in which the exact values of some failure times are observed in addition to interval-censored observations. It is shown that the NPMLE converges weakly to a mean zero Gaussian process whose covariance function is determined by a Fredholm integral equation. Simulations are conducted to demonstrate that the NPMLE based on all the observations substantially outperforms the empirical distribution function, using only the fully observed observations, in terms of the mean square error. It is also shown that the nonparametric bootstrap estimator of the distribution function is first order consistent, which provides asymptotic justification for the use of bootstrap to construct confidence bands for the unknown distribution function.

Keywords and phrases: Asymptotic normality, bootstrap, empirical process, interval censoring, nonparametric maximum likelihood estimation, self-consistency.

1. Introduction and description of the data

Suppose time to event random variables, or failure times, $T_1, \ldots, T_n$ are independent and identically distributed as F_0. If all the random variables are observable, then it is well known that the nonparametric maximum likelihood estimator (NPMLE) of F_0 is the empirical distribution function and it is asymptotically efficient. However, in many medical and reliability studies, observations are subject to censoring. The goal of this paper is to study asymptotic properties of the NPMLE of F_0 based on partly interval-censored data in which some of the failure times are observed, but some of the failure times are subject to interval censoring. More specifically, we consider the following two cases.

(i) "Case 1" partly interval-censored data. For some subjects, the exact failure times $T_1, \ldots, T_{n_1}$ are observed. But for the remaining subjects, only the information pertaining to their current status is available. That is, for the ith subject in this group, we only know whether or not failure has occurred at the examination time U_i; so the observed data is

$$(\delta_i, U_i), \quad i = n_1 + 1, \ldots, n,$$

where $\delta_i = 1$ if the unknown failure time $T_i \leq U_i$ and $\delta_i = 0$ otherwise. Note that this censorship model is different from the doubly-censored data studied by Chang and Yang (1987), Chang (1990) and Gu and Zhang (1993).

(ii) General partly interval-censored data. Again, some of the exact failure times are observed, but some of the failure times are interval-censored. Interval-censored data arises when a failure time T is not observable, but is only known to be bracketed between two examination times. We now describe the interval-censored data we consider.

Suppose there are m potential examination times $U_1 < U_2 < \cdots < U_m$. Let $U_0 = 0$ and $U_{m+1} = \infty$, and let $\delta_k = 1_{[U_{k-1} < T \leq U_k]}$, $\quad k = 1, \ldots, m, m+1$. That is, δ_k is an indicator function specifying which interval determined by $(U_1, U_2, \ldots, U_m)$ contains the unobservable failure time T. Thus δ_k takes values 0 or 1 and $\sum_{k=1}^{m+1} \delta_k = 1$. Let $\Delta =$

$(\delta_1, \ldots, \delta_{m+1})$. If $\delta_k = 1$, T is known to be bracketed between U_{k-1} and U_k and all the other δ's are zero. However, examination times may be censored. For example, if the ith failure time is bracketed in $(U_{m_i-1,i}, U_{m_i,i})$ where $U_{m_i-1,i}$ is the last examination time before the failure and $U_{m_i,i}$ is the first examination time after the failure has occurred, then it may no longer be necessary to conduct further examinations. So the *effective* observations are

$$(\Delta_i, U_{m_i-1,i}, U_{m_i,i}), \quad i = n_1 + 1, \ldots, n,$$

Notice that m_i may differ across subjects.

Partly interval-censored data arise frequently in practice. In estimating the distribution of onset age of chronic diseases, age at onset is known for some affected individuals in the study. However, for others, age at onset is unknown and only current age is available. Then age at onset is less than current age. For an unaffected and susceptible individual, age at onset is greater than current age. This gives rise to the "case 1" partly interval-censored data. Examples of such data can be found in Risch (1983) and Tang, Maestre, Tsai, Liu, Feng, Chung, Chung, Schofield, Stern, Tycko, and Mayeux (1995).

General partly interval-censored data arise often in follow-up studies. An example of such data is provided by the Framingham Heart Disease Study; see Odell, Anderson and D'Agostino (1992) for a description. In this study, times of the first occurrence of subcategory angina pectoris in coronary heart disease patients are of interest. For some patients, time of the first occurrence of subcategory angina pectoris is recorded exactly. But for others, time is recorded only between two clinical examinations. Another example of such data is provided by the study on incidence of proteinuria in insulin-dependent diabetic patients in Denmark; see Enevoldsen, Johnsen, Kreiner, Nerup and Deckrt (1986) for a detailed description.

Turnbull (1976) described a general scheme of incomplete failure time data and derived self-consistency equations for computing the maximum likelihood estimator of the survival function. However, theoretical properties of Turnbull's estimator have not been studied under an unknown continuous survival function except in some important special cases, such as right censoring. It does not seem likely that one can establish properties of the NPMLE under Turnbull's general censoring scheme. Note that both right-censored data and interval-censored data can be regarded as special cases of Turnbull's general censored data. The Kaplan-Meier estimator for right-censored data is asymptotically normal with $n^{1/2}$ rate of convergence (Breslow and Crowley, (1974)). However, the NPMLE with interval-censored data has only $n^{1/3}$-rate of convergence and its limiting distribution is the argmax of the standard two-sided Brownian motion minus parabola (Groeneboom (1991) and Groeneboom and Wellner, (1992)). Large sample properties of the NPMLE based on partly interval-censored data appear to be unknown.

In this paper, we show that the NPMLE based on either the "case 1" partly interval-censored data or the general partly interval-censored data converges weakly to a Gaussian process. We also show that the nonparametric bootstrap estimator of the distribution function is first order consistent, which provides asymptotic justification for the use of the bootstrap to construct confidence bands for the unknown distribution function.

2. Definition and uniqueness of the NPMLE

Assume that the failure time and examination times are independent and that the distribution of the examination times is independent of the distribution of the failure time. The likelihood function for general partly interval-censored data is proportional to

$$L_n(F) = \prod_{i=1}^{n_1} dF(T_i) \prod_{i=n_1+1}^{n} [F(U_{m_i,i}) - F(U_{m_i-1,i})], \tag{2.1}$$

where $dF(t) = F(t) - F(t-)$ is the mass that F puts at t. The NPMLE $\widehat{F}_n$ is then the maximizer of $L_n(F)$ in the class of distribution functions. In the special case of "case 1" partly interval-censored data, the likelihood function simplifies to

$$L_n(F) = \prod_{i=1}^{n_1} dF(T_i) \prod_{i=n_1+1}^{n} F(U_i)^{\delta_i}(1 - F(U_i))^{1-\delta_i}.$$

For any finite sample size n, $\widehat{F}_n$ is determined only at the observed failure times $T_i, 1 \le i \le n_1$, and at the examination times $(U_{m_i-1,i}, U_{m_i,i}), 1 \le i \le n_2$, where $n_2 = n - n_1$ (only these values of F enter the likelihood function). Turnbull (1976) showed that $\widehat{F}_n$ is a discrete distribution function which puts positive mass only at the observed failure times and examination times. To see that $\widehat{F}_n$ is uniquely determined at the observation times, we divide the censored data part into three types: (a) left censoring time denoted by L_j, so the corresponding failure time T_j satisfies $0 < T_j \le L_j, 1 \le j \le n_{21}$; (b) interval censoring time (U_j, V_j), so the corresponding failure time T_j satisfies $U_j < T_j \le V_j, 1 \le j \le n_{22}$; and (c) right censoring time R_j, the corresponding failure time T_j satisfies $T_j > R_j, 1 \le j \le n_{23}$. The size of censored data n_2 is $n_{21} + n_{22} + n_{23}$. The log-likelihood function given in (2.1) can be rewritten as

$$\log L_n(F) = \sum_{i=1}^{n_1} \log(F(T_{(i)}) - F(T_{(i-1)})) + \sum_{j=1}^{n_{21}} \log F(L_j) + \sum_{j=1}^{n_{22}} \log(F(V_j) - F(U_j)) + \sum_{j=1}^{n_{23}} \log(1 - F(R_j)),$$

where $T_{(i)}$'s are the ordered values of $T_1, \ldots, T_{n_1}$ and $T_{(0)} = 0$. Clearly, we can let $F(T_{(0)}) = F(0) = 0$. Let $\mathcal{D} = (Y_{(1)}, \ldots, Y_{(m)})$ be the collection of the exact failure times and the three types of censoring times with $Y_{(1)} \le \cdots \le Y_{(m)}$. Without loss of generality, we can assume that $Y_{(1)}$ corresponds to a value in the T_i's, L_j's, or V_j's. Otherwise, if $Y_{(1)}$ corresponds to a U_j or a R_j, for the distribution function F to maximize $\log L_n$, we must have $F(Y_{(1)}) = 0$. Thus we can take $Y_{(1)}$ out of $\mathcal{D}$. Similarly, we can assume that the largest value $Y_{(m)}$ in D corresponds to a U_j or R_j, because if $Y_{(m)}$ corresponds to a T_i, L_j, or V_j, we must have $F(Y_{(m)}) = 1$ and can delete this $Y_{(m)}$ from $\mathcal{D}$. Let $s_j = F(Y_{(j)})$. Then the log-likelihood can be represented as a function of $\mathbf{s} \equiv (s_1, \ldots, s_m)$. The problem of finding $\widehat{F}_n$ becomes that of maximizing $\log L_n(\mathbf{s})$ over the convex set $\mathcal{S} = \{(s_1, \ldots, s_m) : 0 < s_1 \le \cdots \le s_m < 1\}$. It can be verified that $\log L$ is concave in $\mathcal{S}$. By an argument analogous to Proposition 1.3 of Gronenboom and Wellner (1992) for the uniqueness of the NPMLE with interval-censored data, the solution to the present maximization problem is unique. Therefore, the NPMLE $\widehat{F}_n$ is uniquely determined at the observation times.

In the above, we assumed that $n_{2k} \geq 1, k = 1, 2, 3$. For a fixed n_2, any one or two of n_{21}, n_{22} and n_{23} might be zero. For example, if $n_{12} = 0$, then there is no left-censored data, and the corresponding term in $\log L_n$ does not exist. The uniqueness of $\widehat{F}_n$ is this simpler case can be established similarly.

3. Asymptotic distribution of the NPMLE

Suppose that the distribution function F_0 of the failure time T is continuous. Then it can be shown using the method of Van der Vaart and Wellner (1992) that

$$\sup |\widehat{F}_n(t) - F_0(t)| \rightarrow_{a.s.} 0 \quad \text{as} \quad n_1 \rightarrow \infty.$$

These authors considered the consistency of the NPMLE when part of the data is observed and part is from a mixture density with the unknown distribution as the mixing distribution. Their proof can be adapted to the present situation. If it is further assumed that the distribution function of the examination times is continuous, then consistency holds as $\min\{n_1, n_2\} \rightarrow \infty$. When $n_1 = 0$, the uniform consistency of $\widehat{F}_n$ is proved in Groeneboom and Wellner (1992). We omit the proof of consistency and concentrate on the asymptotic distribution of $\widehat{F}_n$.

The first key assumption is that the number of exact observations is not negligible in the following sense.

Assumption (A1): $n_1/n \rightarrow \alpha_1, \quad \text{as } n \rightarrow \infty \quad \text{with} \quad 0 < \alpha_1.$

This assumption is crucial for us to obtain the $n^{1/2}$-rate of convergence for $\widehat{F}_n$. If $n_1/n \rightarrow 0$ as $n \rightarrow \infty$, then the rate of convergence of $\widehat{F}_n$ would be slower than $n^{1/2}$. In particular if $n_1 = 0$, the rate of convergence is $n^{1/3}$, see Groeneboom and Wellner (1992).

Let $D[0, \infty)$ be the class of bounded right continuous functions with left limits on $[0, \infty)$, equipped with the supremum norm. Convergence in distribution (denoted as $\Rightarrow_D$)

below is according to the definition of Hoffmann-Jørgensen; see for example, Van der Vaart and Wellner (1996) for a description.

We state the results on $\widehat{F}_n$ with "case 1" interval-censored data and with general partly interval-censored data separately, since in the former case, the conditions are weaker.

Let $\alpha_2 = 1 - \alpha_1$. The following are needed in Theorem 3.1. Define a linear operator $\dot{S}_0 : D[0,\infty) \to D[0,\infty)$ by

$$\dot{S}_0 = \alpha_1 I + \alpha_2 K,$$

where I is the identity operator and where, for any $h \in D[0,\infty)$, K is defined by

$$Kh(t) = E\left\{ \frac{F_0(U_1 \wedge t) - F_0(U_1)F_0(t)}{F_0(U_1)(1 - F_0(U_1))} h(U_1) \right\}.$$

Let $\xi_1(x;t)$ as a function of t be the solution to the integral equation $\dot{S}_0 h(t) = 1_{[x \le t]} - F_0(t)$, and let $\xi_2(\delta, u; t)$ as a function of t be the solution to the integral equation

$$\dot{S}_0 h(t) = \delta \frac{F_0(u \wedge t)}{F_0(t)} + (1 - \delta) \frac{F_0(u) - F_0(v \wedge t)}{1 - F_0(u)}.$$

Theorem 3.1. *("Case 1" partly interval-censored data.) Suppose that (A1) holds and that F_0 is continuous. Then*

(i)

$$n^{1/2}(\widehat{F}_n - F_0) \Rightarrow_D Z_1,$$

where Z_1 is a Gaussian process in $D[0,\infty)$ with mean zero and a variance that achieves the information lower bound for the estimation of F_0.

(ii) The covariance function of Z_1 is given by

$$Cov(Z_1(s), Z_1(t)) = \alpha_1 Cov[\xi_1(T_1; s), \xi_1(T_1; t)] + \alpha_2 Cov[\xi_2(\delta_1, U_1; s), \xi_2(\delta_1, U_1; t)],$$

where the first covariance is calculated with respect to the distribution of T_1, the second is with respect to the distribution of (δ_1, U_1).

We now give sufficient conditions under which the NPMLE $\widehat{F}_n$ in the general case converges in distribution to a Gaussian process.

Assumption (A2): The distribution function F_0 is continuous and strictly increasing.

Assumption (A3): The joint distribution function $G(u_1, \ldots, u_m)$ of $(U_1, \ldots, U_m)$ is continuous. Furthermore, there exists a positive number $\eta_0 > 0$ such that

$$P(U_{k+1} - U_k \geq \eta_0) = 1, k = 1, \ldots, m-1.$$

Assumption (A3)*: $(U_1, \ldots, U_m)$ is a vector of discrete random variables with finitely many support points.

Assumption (A3) assumes that there is a positive separation time between any two adjacent examination times. We conjecture that Theorem 3.2 below continues to hold without this assumption, but have not been able to prove this. The point of difficulty is discussed in Remark 6.1 of section 4.

Before stating Theorem 3.2, we first define several expressions needed to describe the covariance structure of the limiting Gaussian process of the NPMLE for the case of $m = 2$. These expressions can be generalized to the case of general m. Denote (U_1, U_2) by (U, V). Define a linear operator $\dot{S}_0 : D[0, \infty) \to D[0, \infty)$ by

$$\dot{S}_0 = \alpha_1 I + \alpha_2 K, \tag{3.1}$$

where I is the identity operator and where for any $h \in D[0, \infty)$, K is defined by

$$Kh(t) = E\left\{ \frac{F_0(U \wedge t)}{F_0(U)} h(U) + \frac{F_0(V \wedge t) - F_0(U \wedge t)}{F_0(V) - F_0(U)} (h(V) - h(U)) - \frac{F_0(t) - F_0(V \wedge t)}{1 - F_0(V)} h(V) \right\}.$$

Let $\psi_1(x; t)$ as a function of t be the solution to the integral equation $\dot{S}_0 h(t) = 1_{[x \leq t]} - F_0(t)$, and let $\psi_2(\delta_1, \delta_2, u, v; t)$ as a function of t be the solution to the integral equation

$$\dot{S}_0 h(t) = \delta_1 \frac{F_0(u \wedge t)}{F_0(u)} + \delta_2 \frac{F_0(v \wedge t) - F_0(u \wedge t)}{F_0(v) - F_0(u)} + (1 - \delta_1 - \delta_2) \frac{F_0(t) - F_0(v \wedge t)}{1 - F_0(v)}.$$

THEOREM 3.2. *(General partly interval-censored data.) Suppose that: (a) Conditions (A1) and (A2) hold; (b) either (A3) or (A3)* holds. Then*

(i)

$$n^{1/2}(\widehat{F}_n - F_0) \Rightarrow_D Z_2,$$

where Z_2 is a Gaussian process in $D[0,\infty)$ with mean zero and a variance that achieves the information lower bound for the estimation of F_0.

(ii) When $m=2$, the covariance function of Z_2 is given by

$$Cov(Z_2(s), Z_2(t)) = \alpha_1 Cov[\psi_1(T_1; s), \psi_1(T_1; t)] + \alpha_2 Cov[\psi_2(\delta_1, \delta_2, U, V; s), \psi_2(\delta_1, \delta_2, U, V; t)],$$

where the first covariance is calculated with respect to the distribution of T_1, the second is with respect to the distribution of $(\delta_1, \delta_2, U, V)$.

Notice that the conditions of Theorem 3.1 are weaker than those of Theorem 3.2. In particular, there is no restriction on the distribution of the examination time.

The covariance functions of Z_1 and Z_2 in the above two theorems are not expressible in closed forms. They are determined by two Fredholm integral equations which do not appear to have explicit solutions. Therefore, Theorems 3.1 and 3.2 can not be directly used to construct pointwise confidence limits or confidence bands for F_0.

One way to estimate the covariance of $\widehat{F}_n$ is to use the inverse of the observed information matrix. The observed information matrix is computed as the negative second derivative of the log-likelihood with respect to the values of $\widehat{F}_n$ at its jump points (note that some examination times may not be jump points of $\widehat{F}_n$). Following the discussion in Section 5 of Vardi and Zhang (1992), this method provides a consistent estimator of the covariance of $\widehat{F}_n$, which follows from the continuity of the inverse of the score operator in a neighborhood of F_0 proved in Section 6. See also Murphy (1995) for a discussion of the use of this method to estimate the covariance of the NPMLE in the frailty model. Thus pointwise confidence limits for F_0 can be obtained from this covariance estimator. However, knowing the covariance of $\widehat{F}_n$ is not enough for constructing confidence bands for F_0. One approach for constructing confidence bands is to use the bootstrap. Theorems 3.1 and 3.2 provide a starting point to verify that the nonparametric bootstrap described below is first order consistent.

Let P_{n_1} be the empirical measure of the exact observations $T_1, \ldots, T_{n_1}$, let P_{n_2} be the empirical measure of the interval-censored observations $(U_{m_i-1,i}, U_{m_i,i}), 1 \le i \le n_2$, let

$T_i^*, 1 \le i \le n_1$, be a bootstrap sample from P_{n_1}, and let $(U^*_{m_i-1,i}, U^*_{m_i,i}), 1 \le i \le n_2$, be a bootstrap sample from P_{n_2}, drawn independently of $T_i^*, 1 \le i \le n_1$. Define the bootstrap NPMLE to be the $\widehat{F}_n^*$ which maximizes

$$L_n^*(F) = \prod_{i=1}^{n_1} dF(T_i^*) \prod_{i=n_1+1}^{n} [F(U^*_{m_i,i}) - F(U^*_{m_i-1,i})]$$

over the class of distribution functions.

THEOREM 3.3. *(i) Under the conditions of Theorem 3.1,*

$$n^{1/2}(\widehat{F}_n^* - \widehat{F}_n) \Rightarrow_D Z_1 \quad \textit{in probability,}$$

where Z_1 is the limiting process given in Theorem 3.1.

(ii) Under the conditions of Theorem 3.2,

$$n^{1/2}(\widehat{F}_n^* - \widehat{F}_n) \Rightarrow_D Z_2 \quad \textit{in probability,}$$

where Z_2 is the limiting process given in Theorem 3.2.

Theorem 3.3 justifies the use of bootstrap to construct confidence bands for F_0.

In Theorems 3.1 and 3.2, it is assumed that F_0 is continuous. It is sometimes of interest to treat the failure times as having a discrete distribution F_0 with finitely many known support points, say $(\tau_1, \ldots, \tau_d)$. Then the derivation of the asymptotic distribution of the maximum likelihood estimator becomes a standard finite dimensional parametric problem. Classical distributional theory on maximum likelihood estimators applies. Specifically, let $p_j = P(T \le \tau_j)$ and let $\hat{p}_j$ be the corresponding maximum likelihood estimator, $j = 1, \ldots, d$. Let $\mathbf{p} = (p_1, \ldots, p_d)'$ and $\hat{\mathbf{p}} = (\hat{p}_1, \ldots, \hat{p}_d)'$. Suppose (A1) holds. Then with either the "case 1" or the general partly interval censored data, and with no restriction on the nature of the distribution function G,

$$n^{1/2}(\hat{\mathbf{p}} - \mathbf{p}) \to_d N(0, \Sigma),$$

where Σ is the Cramér-Rao lower bound, consistently estimated by the inverse of the observed Fisher information matrix.

4. Self-consistency equations

Turnbull (1976) showed that the nonparametric maximum likelihood estimator satisfies the self-consistency equations which are exactly the score equations defined appropriately. In this section, we give a different derivation of the self-consistency equations. The main purpose is to write the self-consistency equations in terms of the present notation. This will be useful in Section 6 where the proofs of Theorems 3.1, 3.2 and 3.3 are based on the equations given below.

If all the failure times could be observed, the self-consistency estimator $\widehat{F}_n$ is simply the empirical distribution function

$$F_n(t) = \frac{n_1}{n} F_{n_1}(t) + \frac{n_2}{n} F_{n_2}(t),$$

where F_{n_1} is the empirical distribution function of the observable $T_1, \ldots, T_{n_1}$, and F_{n_2} is the empirical distribution function of the unobservable $T_{n_1+1}, \ldots, T_n$. When the data is subject to censoring, $\widehat{F}_n$ can be obtained by taking the conditional expectation of F_n given the observed data under the probability measure induced by $\widehat{F}_n$ itself (Efron, 1967). That is,

$$\begin{aligned}
\widehat{F}_n(t) &= E_{\widehat{F}_n}[F_n(t) \mid T_i, U_{m_j-1,j}, U_{m_j,j}, i = 1, \ldots, n_1, j = n_1+1, \ldots, n] \qquad (4.1)\\
&= \frac{n_1}{n} F_{n_1}(t) + \frac{n_2}{n} E_{\widehat{F}_n}[F_{n_2}(t) \mid U_{m_j-1,j}, U_{m_j,j}, j = n_1+1, \ldots, n]\\
&= \frac{n_1}{n} F_{n_1}(t) + \frac{1}{n} \sum_{j=1}^{n_2} \left\{ \frac{\widehat{F}_n(U_{m_j,j} \wedge t) - \widehat{F}_n(U_{m_j-1,j} \wedge t)}{\widehat{F}_n(U_{m_j,j}) - \widehat{F}_n(U_{m_j-1,j})} \right\}.
\end{aligned}$$

For "case 1" partly interval-censored data, the self-consistency equation simplifies to

$$\widehat{F}_n(t) = \frac{n_1}{n} F_{n_1}(t) + \frac{1}{n} \sum_{j=1}^{n_2} \left\{ \delta_j \frac{\widehat{F}_n(U_j \wedge t)}{\widehat{F}_n(U_j)} + (1-\delta_j) \frac{\widehat{F}_n(t) - \widehat{F}_n(U_j \wedge t)}{1 - \widehat{F}_n(U_j)} \right\}. \qquad (4.2)$$

Equation (4.1) or (4.2) immediately give an iterative algorithm to compute $\widehat{F}_n$, which can also be viewed as an EM algorithm (Dempster, Laird and Rubin, (1977)). This algorithm is easy to implement. A faster algorithm is the hybrid algorithm proposed by Wellner and Zhan (1997) which combines the EM algorithm and the iterative convex minorant algorithm. For a detailed description of the iterative convex minorant algorithm, see Groeneboom and Wellner (1992).

5. Efficiency gain by using censored observations

A referee raises the question of how much efficiency is gained by the NPMLE using all the observations over the empirical distribution function using only the fully observed observations, and how the gain in efficiency depends on α_1, the proportion of the fully observed observations. It appears difficult to analytically quantify the improvement in the efficiency of the NPMLE and how it depends on α_1, because the covariance function of the NPMLE is highly implicit. Therefore, I carried out two sets of simulations with "case 1" and "case 2" partly-interval censored data.

In the first set of simulations with "case 1" partly interval-censored observations, four different distributions for F_0 are used. These distributions are: uniform$[0, 1]$; exponential with mean 0.5; Weibull distribution with shape parameter 1.4 and scale parameter 0.55 (with hazard function $1.4 \times 0.55^{-1.4} t^{0.4} = 3.23\, t^{0.4}$); and Weibull distribution with shape parameter 0.7 and scale parameter 0.40 (with hazard function $0.7 \times 0.4^{-0.7} t^{-0.3} = 1.33\, t^{-0.3}$). The first Weibull distribution has an increasing hazard function, while the second Weibull distribution has a decreasing hazard function. The parameters are set so that all four distributions have mean 0.5. The distribution of the examination time is uniform $[0, 1]$ in all four cases.

In the second set of simulations with "case 2" partly interval-censored data, the four generating distributions are: uniform$[0, 4]$; exponential with mean 2; Weibull distribution with shape parameter 1.5 and scale parameter 2.25 (with hazard function $1.5 \times 2.25^{-1.5} t^{0.5} = 0.44\, t^{0.5}$); and Weibull distribution with shape parameter 0.7 and scale parameter 1.65 (with hazard function $0.7 \times 1.65^{-0.7} t^{-0.3} = 0.49\, t^{-0.3}$). Again, the two Weibull distributions have increasing and decreasing hazards, respectively. The parameters are specified so that all the distributions have mean 2. The distribution of the first examination time U_1 is uniform $[0, 1]$; and after the realization of U_1 is generated, the second examination time U_2 is obtained by adding a random number generated from uniform $[10^{-6}, 1]$ to this realization of U_1. The number 10^{-6} (instead of 0) is used here so that the conditions of Theorem 3.2 are satisfied. However, simulations (not shown here) suggest that the numerical difference in the NPMLE and the mean square error between using uniform $[10^{-6}, 1]$ and uniform $[0, 1]$ is negligible.

The total sample size $n = n_1 + n_2$ is 100, and the number of replications is 500 for every simulation model. Five different combinations of the values of (n_1, n_2) are used in the simulations with the ratio α_1 (defined to be n_1/n) equal to $0.1, 0.3, 0.5, 0.7$ and 0.9. The mean square errors (MSE) multiplied by 10^4 of the empirical distribution function F_n using only the fully observed data and the NPMLE $\widehat{F}_n$ along with their ratios are presented in Tables 1 and 2. The mean square error is defined to be the average of the squared differences between the estimator and the true distribution at the observations.

It is seen from Tables 1 and 2 that the reduction in mean square error of the NPMLE $\widehat{F}_n$ using all the observations is substantial for $\alpha_1 = 0.1$ to 0.7. For these α_1 values, the relative efficiency of $\widehat{F}_n$ with respect to the empirical distribution function F_n (measured by the ratio of the mean square error) ranges from approximately 1.2 to about 8.2. Even in the case when $\alpha_1 = 0.9$, there is a small but appreciable reduction in the mean square error in the NPMLE. Therefore, the results presented in Tables 1 and 2 demonstrate the advantage of using the censored observations in addition to the fully observed observations, at least for the models and sample sizes used in simulations. The efficiency gain increases as α_1 decreases, as expected. It is also seen from Tables 1 and 2 that the increase in the mean square error of the NPMLE is small to moderate when n_1 gets smaller for the fixed total sample size n. This suggests that the performance of the NPMLE is stable for a wide range of the values of α_1 and a fixed n.

Table 1. MSE $\times 10^4$: Case 1 partly interval censoring

	(n_1, n_2)	(10, 90)	(30, 70)	(50, 50)	(70, 30)	(90, 10)
Uniform	F_n	116.27	30.95	19.48	12.58	9.12
$[0, 1]$	NPMLE	14.16	14.07	12.08	10.44	8.60
	Ratio	8.21	2.20	1.61	1.20	1.06
Exponential	F_n	98.16	31.73	19.33	11.73	8.19
mean=0.5	NPMLE	16.24	14.12	13.07	10.80	7.88
	Ratio	6.04	2.25	1.48	1.09	1.04
Weibull	F_n	120.19	31.28	17.14	11.81	10.10
shape= 1.4	NPMLE	15.89	12.47	11.65	9.51	9.80
scale= 0.55	Ratio	7.56	2.51	1.47	1.24	1.03
Weibull	F_n	108.70	30.67	18.01	12.73	8.97
shape= 0.7	NPMLE	18.50	15.24	13.32	10.83	8.77
scale= 0.40	Ratio	5.88	2.01	1.35	1.18	1.02

Table 2. MSE $\times 10^4$: "Case 2" partly interval censoring

	(n_1, n_2)	(10, 90)	(30, 70)	(50, 50)	(70, 30)	(90, 10)
Uniform	F_n	109.51	31.20	18.87	11.77	10.21
$[0, 4]$	NPMLE	14.10	11.53	10.97	9.47	9.41
	Ratio	7.77	2.71	1.72	1.24	1.09
Exponential	F_n	105.47	29.63	17.55	14.35	9.90
mean= 2	NPMLE	15.94	14.49	12.09	11.46	9.59
	Ratio	6.61	2.04	1.45	1.25	1.03
Weibull	F_n	94.09	31.71	19.25	11.62	9.98
shape= 1.5	NPMLE	12.56	11.82	11.07	9.35	9.16
scale= 2.25	Ratio	7.49	2.68	1.74	1.24	1.09
Weibull	F_n	109.98	30.34	19.51	14.86	10.27
shape= 0.7	NPMLE	18.95	15.62	14.08	12.08	9.85
scale= 1.65	Ratio	5.70	1.94	1.38	1.23	1.04

6. Proofs

In this section, we prove Theorems 3.2 and 3.3 for the case $m = 2$. It can be verified that the argument works for general m; the only complication is notational. In proving Theorem 3.2, we use the infinite-dimensional M-estimator theorem of Van der Vaart (1995) which is briefly described below. The structures of the two proofs are similar, the main difference is in showing continuous invertibility of $\dot{S}_0$ defined in Theorem 3.1 or 3.2. The proof of Theorem 3.3 resembles the proof of Wellner and Zhan (1996), Theorem 3.1, on the asymptotic distribution of the bootstrap infinite-dimensional M-estimators (Wellner and Zhan call them Z-estimators). Since Theorem 3.1 of Wellner and Zhan (1996) is stated for a single random sample and there are two independent samples in the present situation, we provide a sketch of the proof of Theorem 3.3, including consistency of the bootstrap estimator. It should be noted that the consistency of the bootstrap estimator is the first step towards asymptotic normality. For example, consistency is an important condition in Theorem 3.1 of Wellner and Zhan (1996).

The proof of Theorem 3.1 is omitted, because it is similar to that of of Theorem 3.2. A referee points out that the proofs in Gu and Zhang (1992) for the asymptotic normality of the NPMLE of F_0 based on double censored data can be applied to Theorem 3.1 because of its close connection with the "case 1" partly interval-censored data.

6.1. Proof of Theorem 3.2. We only prove the theorem when $m = 2$. To simplify the notation, let $U = U_1$ (the first examination time) and let $V = U_2$ (the second examination time). Recall that $\delta_1 = 1_{[T\le U]}$, $\delta_2 = 1_{[U<T\le V]}$ and $\delta_3 = 1 - \delta_1 - \delta_2$. Set

$$\phi_F(\delta_1, \delta_2, u, v; t) = \delta_1 \frac{F(u \wedge t)}{F(u)} + \delta_2 \frac{F(v \wedge t) - F(u \wedge t)}{F(v) - F(u)} + \delta_3 \frac{F(t) - F(v \wedge t)}{1 - F(v)}, \qquad (6.1)$$

$$S_n(F)(t) = F(t) - \frac{n_1}{n} F_{n_1}(t) - \frac{n_2}{n} P_{n_2} \phi_F(\delta_1, \delta_2, u, v; t). \qquad (6.2)$$

Recall that P_{n_2} is the empirical measure of $(\delta_{1i}, \delta_{2i}, U_i, V_i), i = n_1 + 1, \ldots, n$. The self-consistency equation (4.1) for $\widehat{F}_n$ can be rewritten as

$$S_n(\widehat{F}_n) = 0.$$

The limiting version of $S_n(F)$ is:

$$S(F)(t) = F(t) - \alpha_1 F_0(t) - \alpha_2 P\phi_F(\delta_1, \delta_2, U, V; t), \tag{6.3}$$

where P is the probability distribution of $(\delta_{11}, \delta_{21}, U_1, V_1)$ and $\alpha_1 = 1 - \alpha_2$. Notice that $S(F_0) \equiv 0$.

For any $f \in D[0, \infty)$, denote the supremum norm by $||f|| = \sup_{0 \le t < \infty} |f(t)|$. Following the general theorem of Van der Vaart (1992) on asymptotics of infinite-dimensional M-estimators, suppose we can prove that:

(i) $n^{1/2}(S_n - S)(F_0) \Rightarrow_D Z_0$, where Z_0 is a tight random map in $D[0, \infty)$;

(ii) $||n^{1/2}(S_n - S)(\widehat{F}_n) - n^{1/2}(S_n - S)(F_0)|| = o_p(1 + \|\widehat{F}_n - F_0\|)$;

(iii) There exists a continuously invertible linear map $\dot{S}_0$ such that

$$||S(F) - S(F_0) - \dot{S}_0(F - F_0)|| = o(\|F - F_0\|) \quad \text{as } ||F - F_0|| \to 0.$$

Then

$$n^{1/2}(\widehat{F}_n - F_0) = \dot{S}_0^{-1} n^{1/2}(S_n - S)(F_0) + o_p(1) \Rightarrow_D -\dot{S}_0^{-1} Z_0.$$

We first prove (ii) and (iii), and then explain that (i) follows analogous arguments to those in proving (ii).

Proof of (ii). We have

$$\begin{aligned}(S_n - S)(\widehat{F}_n) - (S_n - S)(F_0) &= -\frac{n_2}{n}(P_{n_2} - P)\{\phi_{\widehat{F}_n}(\delta_1, \delta_2, u, v; t) - \phi_{F_0}(\delta_1, \delta_2, u, v; t)\} \\ &= -\frac{n_2}{n}(II_{1n}(t) + II_{2n}(t) + II_{3n}(t)),\end{aligned}$$

where

$$II_{1n}(t) = (P_{n_2} - P)\delta_1 \left\{ \frac{\widehat{F}_n(u \wedge t)}{\widehat{F}_n(u)} - \frac{F_0(u \wedge t)}{F_0(u)} \right\},$$

$$II_{2n}(t) = (P_{n_2} - P)\delta_2 \left\{ \frac{\widehat{F}_n(v \wedge t) - \widehat{F}_n(u \wedge t)}{\widehat{F}_n(v) - \widehat{F}_n(u)} - \frac{F_0(v \wedge t) - F_0(u \wedge t)}{F_0(v) - F_0(u)} \right\},$$

and

$$II_{3n}(t) = (P_{n_2} - P)\delta_3 \left\{ \frac{\widehat{F}_n(t) - \widehat{F}_n(v \wedge t)}{1 - \widehat{F}_n(v)} - \frac{F_0(t) - F_0(v \wedge t)}{1 - F_0(v)} \right\}.$$

That the terms II_{1n} and II_{3n} are $o_p(n^{-1/2})$ uniformly in t follows from Pollard (1989), Theorem 4.4, based on the following facts: (a) $F(t)/F(u)$ has total variation bounded by 2 on $u > t$ for any t and any F; (b) the class of uniformly bounded variation functions is Donsker; see for example, Van der Vaart and Wellner (1996), Theorem 2.7.5, page 159.

For $II_{2n}(t)$, the assumptions (A2) and (A3) and the uniform convergence of $\widehat{F}_n$ ensure that $\widehat{F}_n(v) - \widehat{F}_n(u) \geq \lambda_0$ for some $\lambda_0 > 0$ with probability one for all n sufficiently large. Thus it is clear that for any t and F, the functions

$$\frac{\widehat{F}_n(v \wedge t) - \widehat{F}_n(u \wedge t)}{F(v) - F(u)} = 1_{[v \leq t]} + \frac{\widehat{F}_n(t) - \widehat{F}_n(u)}{\widehat{F}_n(v) - \widehat{F}_n(u)} 1_{[u<t<v]}$$

are of bounded uniform sectional variation and hence are in a Donsker class. (A bivariate function $f(x, y)$ on $[0, \infty) \times [0, \infty)$ is said to be of bounded uniform sectional variation if the variations of all sections and of the function itself, are uniformly bounded.) See, for example, Van der Laan (1996), Example 1.2. So again, it can be shown similar to $II_{1n}(t)$ that

$$\sup_{0 \leq t < \infty} |II_{2n}(t)| = o_p(n^{-1/2}).$$

This completes the proof of (ii).

Remark 6.1. For the proof of Theorem 3.2, assumption (A3) is needed only to show that the class of functions

$$\left\{ \frac{F(t) - F(u)}{F(v) - F(u)} 1_{[u<t<v]} : \ t \in [0, \infty), F \text{ is a distribution function} \right\}$$

is a Donsker class. At present, we are not able to verify that this class is Donsker without assumption (A3).

Proof of (iii). Note that in proving (iii), we only need to consider F satisfying $||F - F_0|| = o(1)$. For such a distribution function F, let Q_F be the product of probability measures induced by F and G. (Recall that G is the joint distribution of (U, V)). Define a linear operator A_F

$$A_F h(t', u, v) = E_F(h|\delta_1, \delta_2, u, v), \quad h \in L_2(F),$$

where the conditional expectation is taken under Q_F. Observe that $A_F h$ is a function of (t', u, v) since $\delta_1 = 1_{[t' \leq u]}$ and $\delta_2 = 1_{[u < t' \leq v]}$. ($t'$ will be a dummy variable in the integrals below and it can also be thought of as the unobservable failure time.) Let $h_t(x) = 1_{[0,t]}(x)$. By the definition of ϕ_F given in (6.1), we have

$$A_F h_t(t', u, v) = \phi_F(\delta_1, \delta_2, u, v; t).$$

This also follows directly from the fact that the score for the observed data is equal to the conditional expectation of the score for the complete data given the observed data. The operator A_F maps $L_2(F)$ functions of t' to $L_2(Q_F)$ functions of (x, u, v). Its adjoint A_F^* maps $L_2(Q_F)$ functions to $L_2(F)$ functions and can be expressed as

$$A_F^* b(t') = E_F(b|T = t') = \int b(t', u, v) dG(u, v)$$

for any $b \in L_2(Q_F)$ (Bickel, Klaassen, Ritov and Wellner (1993), pages 271-272, or Groeneboom and Wellner (1992), pages 8 and 9). In particular, we have

$$A_F^* A_F h(t') = \int A_F h(t', u, v) dG(u, v).$$

Furthermore, by Fubini's theorem, we have

$$\int \phi_F(\delta_1, \delta_2, u, v; t) dQ_F = \int A_F h_t(t', u, v) dG(u, v) dF(t') = F(t)$$

and

$$\int (A_F h_t - A_{F_0} h_t)(t', u, v) dG(u, v) d(F - F_0)(t') = o(\|F - F_0\|).$$

These two equations can be verified straightforwardly based on the identity

$$\int \phi_F(\delta_1, \delta_2, u, v; t) dF(t') = F(t)$$

for any (u, v). Therefore,

$$
\begin{aligned}
&S(F)(t) - S(F_0)(t) \\
&= F(t) - \alpha_1 F_0(t) - \alpha_2 \int \phi_F(\delta_1, \delta_2, u, v; t) dG(u, v) dF_0(t') \\
&= \alpha_1 (F(t) - F_0(t)) + \alpha_2 \int A_F h_t(t', u, v) dG(u, v) d(F - F_0)(t') \\
&= \alpha_1 (F(t) - F_0(t)) + \alpha_2 \int A_{F_0} h_t(t', u, v) dG(u, v) d(F - F_0)(t') + o(\|F - F_0\|) \\
&= \alpha_1 (F(t) - F_0(t)) + \alpha_2 \int A^*_{F_0} A_{F_0} h_t(t') d(F - F_0)(t') + o(\|F - F_0\|) \\
&= \alpha_1 \int (I + \frac{\alpha_2}{\alpha_1} A^*_{F_0} A_{F_0}) h_t(t') d(F - F_0)(t') + o(\|F - F_0\|) \\
&= \dot{S}_0 (F - F_0)(t) + o(\|F - F_0\|),
\end{aligned}
$$

where $\dot{S}_0$ is defined by (3.1).

We need to show that $\dot{S}_0$ is continuously invertible. Let H be the class of functions of uniformly bounded variation. Then $I + (\alpha_1/\alpha_2) A^*_{F_0} A_{F_0} : H \to H$ and $\dot{S}_0 : H \to H$ and clearly $h_t(t') = 1_{[0,t]}(t') \in H$. From the proof of Theorem 3.3 of Van der Vaart (1994), to show that $\dot{S}_0$ is continuously invertible it suffices to show that $I + (\alpha_1/\alpha_2) A^*_{F_0} A_{F_0}$ is one to one and that $A^*_{F_0} A_{F_0}$ is compact (hence $I + (\alpha_1/\alpha_2) A^*_{F_0} A_{F_0}$ is continuously invertible). Consider two cases corresponding to assumption (b) of Theorem 3.2.

Case 1. G is continuous. Then the operator $I + (\alpha_1/\alpha_2) A^*_{F_0} A_{F_0}$ is one to one follows from Lemma 3.2 of Van der Vaart (1994), by the fact that $A^*_{F_0} A_{F_0}$ is a self-adjoint, positive definite operator on $L_2(F_0)$ and hence the eigenvalues of $A^*_{F_0} A_{F_0}$ are not less than 1. To see that $A^*_{F_0} A_{F_0}$ is compact, let h_m be a sequence in H. We have

$$
\begin{aligned}
&A^*_{F_0} A_{F_0} h_m(t') \\
&= \int \left\{ 1_{[t' \le u]} \frac{\int_0^u h_m dF_0}{F_0(u)} + 1_{[u < t' \le v]} \frac{\int_0^v h_m dF_0 - \int_0^u h_m dF_0}{F_0(v) - F_0(u)} + 1_{[t' > v]} \frac{\int h_m dF_0 - \int_0^v h_m dF_0}{1 - F_0(v)} \right\} dG(u, v).
\end{aligned}
$$

By Helly's selection theorem, we can find a subsequence $h_{m'}$ of h_m that converges at every continuity point (since h_m has uniformly bounded variation). By the dominated convergence theorem and the continuity of F_0 and G, $A^*_{F_0} A_{F_0} h_{m'}(t')$ converges for every t'.

Case 2. G is discrete with finitely many mass points. Then it is clear that $A^*_{F_0}A_{F_0}$ reduces to a finite dimensional self-adjoint operator. Thus it follows that it is compact and that $I + (\alpha_1/\alpha_2)A^*_{F_0}A_{F_0}$ is one to one. This completes the proof of (iii).

Proof of (i). Since

$$S_n(F_0)(t) - S(F_0)(t) = \frac{n_1}{n}(F_0(t) - F_{n_1}(t)) - \frac{n_2}{n}(P_{n_2} - P)\phi_{F_0}(\delta_1, \delta_2, u, v; t),$$

we only need to show

$$n^{1/2}(P_{n_2} - P)\phi_{F_0}(\delta_1, \delta_2, u, v; t)$$

converges in distribution in $D[0, \infty)$. This follows since the class of functions $\phi_{F_0}(\delta_1, \delta_2, u, v; t), 0 \leq t < \infty$, is a Donsker class using arguments similar to those in the proof of (ii).

Furthermore, it follows from the general result of Van der Vaart (1995) that the NPMLE $\widehat{F}_n$ is regular and asymptotically efficient.

We now identify the covariance function of Z_1. The proofs of (i), (ii) and (iii) imply that

$$\dot{S}_0(\widehat{F}_n - F_0)(t) = -S_n(F_0)(t) + o_p(n^{-1/2}).$$

Since $\dot{S}_0$ is continuously invertible, we have

$$(\widehat{F}_n - F_0)(t) = -\dot{S}_0^{-1} S_n(F_0)(t) + o_p(n^{-1/2}).$$

Let $a(x;t) = 1_{[x \leq t]} - F_0(t)$. By Fubini's theorem we have

$$\dot{S}_0^{-1} S_n(F_0)(t) = \frac{n_1}{n}(P_{n_1} - P_1)[\dot{S}_0^{-1} a(\cdot;t)] + \frac{n_2}{n}(P_{n_2} - P_2)[\dot{S}_0^{-1}\phi_{F_0}(\cdot;t)].$$

Therefore, the covariance function of Z_1 is given as in Theorem 3.2. This completes the proof of Theorem 3.2. □

6.2. Proof of Theorem 3.3. We only prove part (ii) of this theorem. The proof of (i) is similar and is omitted. As in (4.1),

$$\widehat{F}^*_n(t) = \frac{n_1}{n} F^*_{n_1}(t) + \frac{n_2}{n} P^*_{n_2}\phi_{\widehat{F}^*_n}(t), \tag{6.4}$$

where ϕ_F is defined by equation (6.1). We first show that

$$\sup_{0\le t<\infty} |\widehat{F}_n^*(t) - F_0(t)| \to 0, \quad \text{as } n \to \infty \tag{6.5}$$

almost surely with respect to the underlying probability measure and the bootstrap measure. By the bootstrapping Law of Large Numbers (Giné and Zinn (1990), Theorem 2.6), we can find a subsequence n' of n such that $F^*_{n'_1}$ converges uniformly to F_0 almost surely and $P^*_{n'_2}$ converges uniformly to P_2 almost surely. On the other hand, by Helly's selection theorem, there exists a subsequence of $\widehat{F}_n^*$ that converges vaguely to a subdistribution function F^*. Take a common subsequence, and take limit on both sides of equation (6.4) to get

$$F^*(t) = \alpha_1 F_0(t) + \alpha_2 P_2 \phi_{F^*}(t).$$

As in the proof of Theorem 3.2, this equation can be written as

$$\int (I + \frac{\alpha_2}{\alpha_1} A^*_{F^*} A_{F^*}) h_t(t') d(F^* - F_0)(t') = 0.$$

Therefore, using the same argument as in the proof of Theorem 3.2, this implies $F^*(t) = F_0(t)$ for almost all t. By continuity of F_0, $F^*(t) = F_0(t)$ for all t. This finishes the proof of (6.5).

Now we are ready to prove Theorem 3.3. Combining (4.1) and (6.4), we obtain

$$\widehat{F}_n^*(t) - \widehat{F}_n(t) = \frac{n_1}{n}[F^*_{n_1}(t) - F_{n_1}(t)] + \frac{n_2}{n}[P^*_{n_2}\phi_{\widehat{F}_n^*}(t) - P_{n_2}\phi_{\widehat{F}_n}(t)]. \tag{6.6}$$

First consider

$$\begin{aligned} &P^*_{n_2}\phi_{\widehat{F}_n^*}(t) - P_{n_2}\phi_{\widehat{F}_n}(t)] \qquad (6.7)\\ &= (P^*_{n_2} - P_{n_2})\phi_{F_0}(t) + P[\phi_{\widehat{F}_n^*}(t) - \phi_{\widehat{F}_n}(t)] \\ &\quad + (P^*_{n_2} - P_{n_2})[\phi_{\widehat{F}_n^*}(t) - \phi_{F_0}(t)] + (P_{n_2} - P)[\phi_{\widehat{F}_n^*}(t) - \phi_{\widehat{F}_n}(t)]. \end{aligned}$$

For the second term on the right side of (6.7), after some straightforward calculation, it can be verified that

$$P[\phi_{\widehat{F}_n^*}(t)] = -K(\widehat{F}_n^* - F_0)(t) + \widehat{F}_n^*(t) + o(\|\widehat{F}_n^* - F_0\|),$$

and

$$P[\phi_{\widehat{F}_n}(t)] = -K(\widehat{F}_n - F_0)(t) + \widehat{F}_n(t) + o(\|\widehat{F}_n - F_0\|),$$

where K is the integral operator defined in Theorem 3.2. These two equations combined with $\|\widehat{F}_n - F_0\| = O_p(n^{-1/2})$ give

$$P[\phi_{\widehat{F}_n^*} - \phi_{\widehat{F}_n}](t) = -K(\widehat{F}_n^* - \widehat{F}_n)(t) + \widehat{F}_n^*(t) - \widehat{F}_n(t) + o(\|\widehat{F}_n^* - \widehat{F}_n\|) + o_p(n^{-1/2}). \quad (6.8)$$

Furthermore, by (6.5) and the asymptotic equicontinuity of bootstrapping empirical measures (Giné and Zinn (1990), Theorem 2.4), we have, for the third and fourth term on the right side of (6.7),

$$(P_{n_2}^* - P_{n_2})[\phi_{\widehat{F}_n^*}(t) - \phi_{F_0}(t)] = o_p(n^{-1/2})$$

and

$$(P_{n_2} - P)[\phi_{\widehat{F}_n^*}(t) - \phi_{\widehat{F}_n}(t)] = o_p(n^{-1/2})$$

uniformly in t. Thus by (6.6), (6.7) and (6.8), and noting that $\dot{S}_0 = \alpha_1 I + \alpha_2 K$, we have

$$\dot{S}_0(\widehat{F}_n^* - \widehat{F}_n)(t) = \alpha_1(F_{n_1}^* - F_{n_1})(t) + \alpha_2(P_{n_2}^* - P_{n_2})\phi_{F_0}(t) + o(\|\widehat{F}_n^* - \widehat{F}_n\|) + o_p(n^{-1/2}).$$

So the theorem follows from the continuous invertibility of $\dot{S}_0$ proved in the proof of Theorem 3.2 and the result on weak convergence of bootstrapping empirical measures of Giné and Zinn (1990). □

ACKNOWLEDGEMENTS. I thank Jon A. Wellner for many helpful comments on this paper and discussions on related topics. I also thank two anonymous referees, the Chair-Editor and my colleague Dale Zimmerman for very constructive comments that lead to improvement of this article. This research was supported in part by a 1996 Old Gold Summer Fellowship from the University of Iowa.

REFERENCES

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins University Press, Baltimore.

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.

Dempster, A. P., Laird, N., and Rubin, D. B.(1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B* **39**, 1-38.

Efron, B. (1967). The two sample problem with censored data. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,* 831-853, University of California Press, Berkeley.

Enevoldsen, A. K., Borch-Johnson, K., Kreiner,S., Nerup, J. and Deckert, T. (1987). Declining incidence of persistent proteinuria in type I (insulin-dependent) diabetic patient in Denmark. *Diabetes* **36**, 205-209.

Giné, E. and Zinn, J. (1990). Bootstrapping general empirical measures. *Ann. Probability* **7**, 1-26.

Groeneboom, P. (1991). Nonparametric maximum likelihood estimators for interval censoring and deconvolution. *Technical Report 378*, Department of Statistics, Stanford University.

Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* DMV Seminar Band 19, Birkhäuser, Basel.

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

Huang, J. (1997). Asymptotic properties of the NPMLE of a distribution function based on ranked set samples self-consistent estimation based on ranked set samples. *Ann. Statist.* **25**, 1036-1049.

Kress, R. (1989), *Linear Integral Equations.* Springer-Verlag, Berlin.

Kantorovich, L. V. and Akilov, G . P. (1982). *Functional Analysis.* Pergamon Press Ltd, Oxford.

Kress, R. (1989). *Linear Integral Equations.* Springer-Verlag, Berlin.

Murphy, S. A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23**, 182-198.

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

Pollard, D. (1989). Asymptotics via empirical processes (with discussion). *Statistical Science* **4**, 341-366.

Risch, N. (1983) Estimating morbidity risks with variable age of onset: review of methods and a maximum likelihood approach. *Biometrics* **39**, 929-939.

Tang, M. X., Maestre, C., Tsai W. Y., Liu X. H., Feng, L., Chung W. Y., Chung M., Schofield, P., Stern, Y., Tycko, B. and Mayeux R. (1995). Relative risk of Alzheimer disease and age-at-onset distributions, based on APOE genotype among elderly African Americans, Caucasians, and Hispanics in New York City. *Am. J. Hum. Genet.* **49**, 76-87.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped censored and truncated data. *J. R. Statist. Soc. B* **38**, 290-295.

Vardi, Y. and Zhang, C.-H. (1992). Large sample study of empirical distributions in a random-multiplicative censoring model. *Ann. Statist.* **20**, 1022-1039.

Van der Laan, M. J. (1996). Efficient and inefficient estimation in semiparametric models. *CWI Tract* **44**. Centre for Mathematics and Computer Science, Amsterdam.

Van der Vaart, A. W. (1995). Efficiency of infinite dimensional M-estimators. *Statistica Neerlandica* **49**, 9-30.

Van der Vaart, A. W. (1994). Maximum likelihood estimation with partially censored data. *Ann. Statist.* **22**, 1896-1916.

Van der Vaart, A. W. and Wellner, J. A. (1992). Existence and consistency of maximum likelihood in upgraded mixture models. *J. Multivariate Analysis* **43**, 133-146.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer Verlag, New York.

Wellner, J. A. and Zhan, Y. (1996). Bootstrapping Z–Estimators. *Technical Report* **308**, Department of Statistics, University of Washington.

Wellner, J. A. and Zhan, Y. (1997). A hybrid algorithm for computation of the NPMLE from censored data. *J. Amer. Statist. Assoc.* **92**, 845-959.