Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Short communication

The nonparametric maximum likelihood estimator for middle-censored data

Pao-Sheng Shen
Department of Statistics, Tunghai University, Taichung 40704, Taiwan

A R T I C L E  I N F O

Article history:
Received 6 July 2010
Received in revised form
10 February 2011
Accepted 11 February 2011
Available online 18 February 2011

Keywords:
Middle censoring
Mixed interval censoring
Self-consistent

A B S T R A C T

In this note, we consider data subjected to middle censoring where the variable of interest becomes unobservable when it falls within an interval of censorship. We demonstrate that the nonparametric maximum likelihood estimator (NPMLE) of distribution function can be obtained by using Turnbull's (1976) EM algorithm or self-consistent estimating equation (Jammalamadaka and Mangalam, 2003) with an initial estimator which puts mass only on the innermost intervals. The consistency of the NPMLE can be established based on the asymptotic properties of self-consistent estimators (SCE) with mixed interval-censored data (Yu et al., 2000, 2001).

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Middle censoring occurs when a data point falls inside a random censoring interval whereby it becomes unobservable. For some individuals the exact values are available while for others the corresponding intervals of censorship are observed. We mention two situations where middle censoring occurs. (i) In a follow-up study, if the childhood learning center where the observations are being taken, is closed for a period, due to an external emergency such as the outbreak of severe acute respiratory syndrome (SARS). (ii) In a clinical trial, where the clinic where the observations are being taken, is closed for a period, due to an external emergency such as the outbreak of war or a strike. For situations (i) and (ii) where during a fixed time interval (this fixed interval is indeed, a random interval (denoted by \((U, V)\)) relative to individual's lifetime) the observation was not possible. If some children (or patients) develop a skill (or disease) of interest during this time, we are not able to observe the exact age \(T\) of these children (or patients) at the time of skill (or disease) development, rather only the information that the event of interest occurred during a certain time interval \((U, V)\). At first glance, middle censoring, where a random middle part is missing, appears as complementary to the idea of double censoring in which the middle part is what is actually observed. However, a careful reflection and analysis shows them to be quite different ideas; see Jammalamadaka and Mangalam (2003) for details.

Let \(T_i, i=1,\ldots, n\), be a sequence of i.i.d. random variables with distribution function \(F_0\). Independent of \(T_i\)'s, let \((U_i, V_i)\), \(i=1,\ldots, n\), be i.i.d. extended real-valued random variables with joint distribution function \(K_0(x, y) = P(U_i \leq x, V_i \leq y)\) such that \(P(U_i < V_i) = 1\). The \(i\)th observation is said to be censored if \(T_i \in (U_i, V_i)\). We observe \((X_i, \delta_i), i=1,\ldots, n\), where \(X_i=T_i\) if \(\delta_i=I_{T_i \in (U_i, V_i)}\) and \(X_i=(U_i, V_i)\) if \(\delta_i=0\). For indentifiability of \(F_0\), we assume that

\[
A_0(x) = K_0(x, -\infty) - K_0(x, -x) = P(U_i < x < V_i) < 1,
\]

on any interval \([a,b], a \leq b\) for which \(F_0(b) > F_0(a^-)\). Observe that if \(A_0(x)=1\) on any interval where \(F_0\) has a positive mass, then censoring occurs with probability 1 on such an interval. As a consequence, there will be no observations on the
interval and that prevents us from distinguishing any two distributions which are identical outside \([a,b]\) but differing only on \([a,b]\).

In many censoring situations, if we were to try to estimate the distribution function via the EM algorithm the resulting equation takes the form

\[
\hat{F}_S(t) = E_{F_n[X]}[E_n|X]
\]  

(1.2)

as described by Tsai and Crowley (1985), where \(E_n\) is the empirical distribution function and \(X\) denotes the observed data. This equation was first introduced and referred to as self-consistency equation by Efron (1967). A solution \(\hat{F}_S\) of (1.2) is called a self-consistent estimator (SCE) of \(F_0\). In different types of censoring, the relationship between nonparametric maximum likelihood estimator (NPMLE) and SCE has been studied by various authors. In the case of right censoring the product-limit estimator (Kaplan and Meier, 1958) is the NPMLE, and Efron (1967) showed that it is also self-consistent. Furthermore, correct Turnbull's characterization. Here, we consider the case when there is no truncation. Let \(A_i\) denotes the class of distribution functions such that \(\bigcap_{\iota=1}^n A_i = \emptyset\) and \(\bigcup_{\iota=1}^n A_i = \mathbb{R}\) if \(L_i = R_i\) and \(A_i = A_{L_i,R_i} = (L_i,R_i)\) if \(L_i < R_i\). For \(i = 1, \ldots, n\), let \(L_i = R_i = t_i\) if \(\delta_i = 1\) and \((L_i,R_i) = (U_i,V_i)\) if \(\delta_i = 0\). Thus, we observe the middle-censored data \(A_1, \ldots, A_n\). Following Turnbull (1976), Friedman (1994) and Alioum and Commenges (1996), we consider nonparametric estimation of \(F_0\) called a self-consistent estimator (SCE) of \(F_0\) (see Jammalamadaka and Mangalam, 2003) \(\hat{F}_S\) satisfies the following equation:

\[
\hat{F}_S(t) = n^{-1} \sum_{i=1}^n \left\{ \delta_i I_{[t_i,t]} + (1-\delta_i) I_{[0,v_i]} + (1-\delta_i) I_{[t_i,u_i]} \right\}.
\]  

(1.3)

Jammalamadaka and Mangalam (2003) showed that the NPMLE satisfies the self-consistency equation (1.3). They also pointed out that an SCE provides only a local maximum of the likelihood equation and may not be an NPMLE. Furthermore, they showed that the NPMLE will have all its mass on the uncensored observations except when it so happens that a censored interval contains no uncensored observation. The consistency of the SCE \(\hat{F}_S\) was established by Jammalamadaka and Mangalam (2003) for the special case when either \(U_i\) or \(V_i\) is degenerate. Jammalamadaka and Iyer (2004) proposed an approximation to the distribution function \(F_0\) (denoted by \(\hat{F}_S\)) for which a modified self-consistent estimator \(\tilde{F}_S\) was obtained. They established the asymptotic properties of \(\hat{F}_S\) and provided an upper bound of the difference between \(\hat{F}_S\) and \(\tilde{F}_S\). Mangalam et al. (2008) showed that under condition (A) every censoring interval contains at least one uncensored observation, i.e. \(\delta_i = 0\) implies that there exists \(j\) such that \(\delta_j = 1\) and \(X_j \in (U_i,V_i)\), the solution of Eq. (1.3) will be unique, and as a consequence \(\hat{F}_n\) will be equal to an NPMLE. Mangalam et al. (2008) proposed a technique for obtaining the NPMLE by dividing the original problem into subproblems.

In this note, we aim to establish connections between the middle-censoring and interval censoring by investigating the self-consistency algorithm. In Section 2, we shall demonstrate that the NPMLE of \(F_0\) can be obtained by using the EM algorithm of Turnbull (1976) or the self-consistent estimating equation (Jammalamadaka and Mangalam, 2003) with an initial estimator which puts mass only on the innermost intervals. Furthermore, we establish the consistency and asymptotic normality of the NPMLE by using the results of Yu et al. (2000, 2001).

2. The NPMLE

2.1. Self-consistency

Turnbull (1976) characterized the NPMLE in the presence of interval censoring and truncation. Frydman (1994) later corrected Turnbull’s characterization. Here, we consider the case when there is no truncation. Let \(A_i = A_{L_i,R_i} = (L_i,R_i)\) if \(L_i = R_i\), and \(A_i = A_{L_i,R_i} = (L_i,R_i)\) if \(L_i < R_i\). For \(i = 1, \ldots, n\), let \(L_i = R_i = t_i\) if \(\delta_i = 1\) and \((L_i,R_i) = (U_i,V_i)\) if \(\delta_i = 0\). Thus, we observe the middle-censored data \(A_1, \ldots, A_n\). Following Turnbull (1976), Frydman (1994) and Alioum and Commenges (1996), we consider nonparametric estimation of \(F_0\) using the independent observation \(A_i\)'s. Since Turnbull’s EM algorithm can be used to tackle the case when \(A_i\) is a single point set, the connection between middle censoring and interval censoring can be established. Based on the notations defined above, the likelihood is proportional to

\[
L^*(F_0) = \prod_{i=1}^n P_{F_0}(A_i),
\]  

(2.1)

where \(P_{F_0}(A_i)\) denotes the probability that is assigned to the interval by \(F_0\). We define an NPMLE as \(\hat{F}_M = \text{argmax}_{F \in \mathcal{F}} L^*(F_0)\), where \(\mathcal{F}\) denotes the class of distribution functions such that \(P_{F_0}(\bigcup_{i=1}^n A_i) = 1\). For censored data, using graph theory, Gentleman and Vandal (2001) presented methods for finding the NPMLE of \(F_0\). For censored and truncated data, Hudgens (2005) employed a graph theoretical approach to describe the support set of the NPMLE of \(F_0\).

Define innermost intervals \(H_j, j = 1, \ldots, J\), induced by \(A_1, \ldots, A_n\) to be all the disjoint intervals which are non-empty intersections of these \(A_i\)'s (e.g. \(A_2 = A_1 \cap A_2\) is an intersection of \(A_i\)'s) such that \(A_i \cap H_j = \emptyset\) or \(H_j\) for all \(i\) and \(j\). Let the endpoints of the innermost intervals be \(q_i\) and \(p_j, j = 1, \ldots, J\), where

\[0 \leq q_1 \leq p_1 \leq q_2 \leq p_2 \leq \cdots \leq q_J \leq p_J < \infty.\]
Peto (1973) showed that the NPMLE of $F_0$ assigns weight, say $s_1, \ldots, s_J$, to the corresponding innermost intervals $H_1, \ldots, H_J$ only. Thus, it suffices to maximize

$$L^*(s) = \prod_{i=1}^n \sum_{j=1}^J z_{ij} s_j,$$

(2.2)

where $z_{ij} = I[H_j \subset A_i]$ and $I[\cdot]$ is the usual indicator function. The goal is to maximize likelihood (2.2) subject to the constraints

$$\sum_{j=1}^J s_j = 1$$

(2.3)

and

$$s_j \geq 0 \quad (j = 1, \ldots, J).$$

(2.4)

Thus, we can limit our search to the space given by constraints (2.3) and (2.4). We shall use $\Omega$ to denote the parameter space that is given by constraints (2.3) and (2.4), i.e.

$$\Omega = \left\{ s \in \mathbb{R}^J : \sum_{j=1}^J s_j = 1; s_j \geq 0 \text{ for } j = 1, \ldots, J \right\}.$$

To find the maximum likelihood estimate of the vector $s$, we use an EM algorithm as follows.

E-step: Let the expected value of $z_{ij}$ be denoted by $\mu_{ij}(s)$. Then under $s$,

$$\mu_{ij}(s) = \frac{z_{ij} s_j}{\sum_{k=1}^J z_{ik} s_k}.$$

(2.5)

M-step: In the maximization step, we treat the expected values as observed. The overall proportion of failures in the interval $H_j$ is

$$\pi_j(s) = \frac{\sum_{i=1}^n \mu_{ij}(s)}{\sum_{i=1}^n \sum_{k=1}^J \mu_{ik}(s)} = n^{-1} \sum_{i=1}^n \mu_{ij}(s).$$

(2.6)

The EM algorithm iterates between Eqs. (2.5) and (2.6) after selecting initial estimates $s_j^{(0)} > 0$ such that $\sum_{j=1}^J s_j^{(0)} = 1$, i.e., computes $\mu_{ij}(s^{(0)})$, updates $s$ by $\mu_{ij}(s^{(1)})$, and repeats until convergence. The resulting self-consistent estimate of $s$, which is a solution of simultaneous equation $s_j = \pi_j(s)$ ($j = 1, \ldots, J$), is exactly the Turnbull’s (1976) self-consistency algorithm as follows:

$$s_j^{(b)} = \left\{ \frac{d_j(s^{(b-1)})}{n} \right\} s_j^{(b-1)} \quad (1 \leq j \leq J),$$

(2.7)

where

$$d_j(s^{(b-1)}) = \sum_{i=1}^n \left( z_{ij} / \sum_{k=1}^J z_{ik} s_k^{(b-1)} \right).$$

Let $s_j$ $(j = 1, \ldots, J)$ denote the estimators obtained from (2.7). Using standard convex optimization techniques, Gentleman and Geyer (1994, see p. 619) provided easily verifiable conditions, i.e. Kuhn–Tucker conditions, for the self-consistent estimator $s_j$ to be a maximum likelihood estimator.

Based on the estimators $s_j$’s, an estimator $\hat{F}_M(t)$ of $F_0(t)$ can be uniquely defined for $t \in [p_i, q_{i+1})$ by $\hat{F}_M(p_j) = \hat{F}_M(q_{j+1}) = \hat{s}_1 + \cdots + \hat{s}_j$, but is not uniquely defined for $t$ being in an open innermost interval $(q_j, p_j)$ with $q_j < p_j$. To avoid ambiguity we define $\hat{F}_M(t) = \hat{s}_1 + \cdots + \hat{s}_{j-1} + \hat{s}_j (t-q_j) / (p_j-q_j)$ if $t \in (q_j, p_j)$ and $0 < q_j < p_j < \infty$. Next, we shall show that the estimator $\hat{F}_M$ satisfies self-consistent equation (1.3), i.e.

$$F(t) = n^{-1} \sum_{i=1}^n \left\{ \delta_{i[H_i \leq t]} + \delta_{i[H_i > t]} \right\} \frac{F(t)-F(L_i)}{F(R_i)-F(L_i)}.$$

(2.8)

**Theorem 1.** The NPMLE $\hat{F}_M$ satisfies Eq. (2.8).

**Proof.** First, notice that for each $(q_j, p_j)$ $(j = 1, \ldots, J)$ either $q_j=p_j$ or $q_j < p_j$ and there is no uncensored observation in $(q_j, p_j)$ if $q_j < p_j$. Furthermore, $d_j(s)$ can be written as

$$d_j(s) = \sum_{i=1}^n \left( \frac{z_{ij}}{\sum_{k=1}^J z_{ik} s_k} \right) + \sum_{i=1}^n \left( 1-\delta_i \right) \left( \frac{z_{ij}}{\sum_{k=1}^J z_{ik} s_k} \right).$$

(2.9)
Consider an initial estimator $\hat{F}^{(0)}$, which puts mass only on $(q_j, p_j)$ $(j = 1, \ldots, J)$. Let $\hat{F}^{(1)}$ denote the first step estimator. Without changing the innermost intervals and likelihood function, we can transform data by moving all right censored points between $p_{j-1}$ and $q_j$ to $p_j$. Similarly, move all left censored points between $p_{j-1}$ and $q_j$ to $q_j$ (see Li et al., 1997). Hence, we have $\sum_{i=1}^{n} (1 - \delta_{ij}) I_{[q_j, R_i \cap (p_{j-1}, q_j )]} = 0$. Since $\sum_{i=1}^{n} \delta_{ij} I_{[q_j, R_i \cap (p_{j-1}, q_j )]} = 0$, we have

$$\hat{F}^{(1)}(p_j - ) - \hat{F}^{(1)}(p_{j-1}) = n^{-1} \sum_{i=1}^{n} (1 - \delta_{ij}) I_{[q_j, R_i \cap (p_{j-1}, q_j )]} + n^{-1} \sum_{i=1}^{n} (1 - \delta_{ij}) I_{[p_{j-1}, q_j \cap (L_i, R_i )]} \frac{\hat{F}^{(0)}(p_j - ) - \hat{F}^{(0)}(p_{j-1})}{\hat{F}^{(0)}(R_i - ) - \hat{F}^{(0)}(L_i )} = 0.$$ 

Hence, $\hat{F}^{(1)}$ also puts mass only on $(q_j, p_j)$ $(j = 1, \ldots, J)$. Next, we consider the following two cases:

**Case 1:** $q_j = p_j$. When $q_j = p_j$, since there is no uncensored observations in $(q_j, p_j)$, we have

$$\hat{F}_M(p_j) - \hat{F}_M(p_j - ) = \hat{F}_M(p_j - ) - \hat{F}_M(p_{j-1}) = n^{-1} \sum_{i=1}^{n} (1 - \delta_{ij}) \left( \frac{1}{J} \sum_{k=1}^{J} \hat{z}_j k \right) \hat{z}_j,$$ 

(2.9)

Now

$$\hat{F}^{(1)}(p_j) - \hat{F}^{(1)}(p_j - ) = n^{-1} \sum_{i=1}^{n} (1 - \delta_{ij}) \frac{\hat{F}^{(1)}(p_j - ) - \hat{F}^{(1)}(p_{j-1})}{\hat{F}^{(1)}(R_i - ) - \hat{F}^{(1)}(L_i )},$$

which can be written as

$$d\hat{F}^{(1)}(p_j) = n^{-1} \sum_{i=1}^{n} (1 - \delta_{ij}) \frac{\hat{z}_j k \hat{z}_j / C_0}{d\hat{F}^{(0)}(p_j - )} d\hat{F}^{(0)}(p_j).$$

Equation (2.10) is equivalent to Equation (2.9).

**Case 2:** $q_j < p_j$. When $q_j < p_j$, since there is no uncensored observations in $(q_j, p_j)$, we have

$$\hat{F}_M(p_j) - \hat{F}_M(q_j) = n^{-1} \sum_{i=1}^{n} (1 - \delta_{ij}) \left( \frac{1}{J} \sum_{k=1}^{J} \hat{z}_j k \right) \hat{z}_j,$$ 

(2.11)

Now

$$\hat{F}^{(1)}(p_j - ) - \hat{F}^{(1)}(q_j) = n^{-1} \sum_{i=1}^{n} (1 - \delta_{ij}) I_{[q_j, R_i \cap (p_{j-1}, q_j )]} + n^{-1} \sum_{i=1}^{n} (1 - \delta_{ij}) I_{[p_{j-1}, q_j \cap (L_i, R_i )]} \frac{\hat{F}^{(0)}(p_j - ) - \hat{F}^{(0)}(q_j)}{\hat{F}^{(0)}(R_i - ) - \hat{F}^{(0)}(L_i )}.$$ 

First, we have $\sum_{i=1}^{n} (1 - \delta_{ij}) I_{[q_j, R_i \cap (p_{j-1}, q_j )]} = 0$. Next, note that given an interval $(L_i, R_i)$ and $\delta_i = 0$, we either have $(q_j, p_j) \subseteq (L_i, R_i)$ or $(q_j, p_j) \cap (L_i, R_i) = 0$. Hence, we have $\sum_{i=1}^{n} (1 - \delta_{ij}) I_{[q_j, R_i \cap (p_{j-1}, q_j )]} = 0$. It follows that

$$\hat{F}^{(1)}(p_j - ) - \hat{F}^{(1)}(q_j) = n^{-1} \sum_{i=1}^{n} (1 - \delta_{ij}) I_{[q_j, R_i \cap (p_{j-1}, q_j )]} \frac{\hat{F}^{(0)}(p_j - ) - \hat{F}^{(0)}(q_j)}{\hat{F}^{(0)}(R_i - ) - \hat{F}^{(0)}(L_i )}.$$ 

Equation (2.12) is equivalent to Equation (2.11). The proof is completed. □

This conclusion is the same as in Theorem 1 of Jammalamadaka and Mangalam (2003). However, our proof is based on the EM algorithm of Turnbull (1976). It is obvious that if condition (A) holds then $q_j = p_j$ for all $j$, and the solution $\hat{z}_j$’s will have all its mass on the uncensored observations. Furthermore, if we start with an initial estimator which puts weight 1/J on $q_j = p_j$ for uncensored observations and on $(q_j + p_j)/2$ for censored observations, we can obtain an NPMLE by using Eq. (2.8). However, similar to interval-censored data, the self-consistent NPMLE of $F_0$ is not uniquely defined for $x \in (q_j, p_j)$ if $q_j < p_j$. An SCE with an initial estimator which puts weight on intervals other than $(q_j, p_j)$ can lead to a less efficient estimator (not NPMLE).

### 2.2. Large sample properties

In this section, we shall investigate large sample properties of $\hat{F}_M$. First, we introduce mixed interval censored (MIC) data. A data set is called a MIC data when it consists of both exact observations and case 2 interval censoring data (i.e. $L_i < R_i$). Mixed IC data arises in clinical follow-up studies where a tumor marker (e.g., CA-125 in ovarian cancer) is available, a patient whose marker value is consistently on the high (or low) end of normal range in repeated testing is usually under close surveillance for possible relapse. If such a patient should relapse, then the time to clinical relapse can often be accurately determined. However, if a patient is not under close surveillance, and would seek assistance only after some tangible symptoms have appeared, then time to relapse would be subject to case 2 interval censoring. For MIC data, several models have been proposed, and the asymptotic properties of the NPMLE have been investigated under the assumption that either the censoring variables take on finite many values (see Huang, 1999; Yu et al., 1998, 2000), or the censoring and
survival distributions are strictly increasing and continuous and they have “positive separation” (see Huang, 1999, Assumption (A3)). For MIC data, define \((Y_i, Z_i)\) as a pair of extended random censoring times \((\infty \text{ allowed})\) with \(P(Y_i < Z_i) = 1\), and \(T_i\) is independent of \((Y_i, Z_i)\). Yu et al. (2000, see (2.1)) considered a mixture interval censorship model to characterize MIC data as follows:

\[
(L_i, R_i) = \begin{cases} 
(T_i, T_i) & \text{if } T_i \notin (Y_i, Z_i) \text{ (exact)}, \\
(Y_i, \infty) & \text{if } T_i > Y_i \text{ and } Z_i = \infty \text{ (right censoring)}, \\
(-\infty, Z_i) & \text{if } T_i \leq Y_i \text{ and } Y_i = -\infty \text{ (left censoring)}, \\
(Y_i, Z_i) & \text{if } -\infty < Y_i < T_i \leq Z_i < \infty \quad \text{(strictly interval censored)}. 
\end{cases}
\]

(2.13)

For middle censored data, since neither right censoring nor left censoring occurs, model (2.13) is reduced to

\[
(L_i, R_i) = \begin{cases} 
(T_i, T_i) & \text{if } T_i \notin (Y_i, Z_i), \\
(Y_i, Z_i) & \text{if } T_i \in (Y_i, Z_i). 
\end{cases}
\]

(2.14)

Replacing \((Y_i, Z_i)\) and \((U_i, V_i)\) in (2.14) with \((U_i, V_i)\), we obtain the model for middle-censored data as follows:

\[
(L_i, R_i) = \begin{cases} 
(T_i, T_i) & \text{if } T_i \notin (U_i, V_i), \\
(U_i, V_i) & \text{if } T_i \in (U_i, V_i). 
\end{cases}
\]

Hence, although the sampling scheme of MIC data seems to be quite different in character from that of middle-censored data described in Section 1, the resulting observations \((L_i, R_i)\) would reduce to the observations from middle-censoring data when there is no left or right censoring.

**Theorem 2.** Let \(S_T\) denote the set of all support points of variables \(T_i\). Similarly, define \(S_V\) and \(S_U\) for \(V_i\) and \(U_i\), respectively. Suppose that \(S_T \cup S_V \subseteq S_U\). Then (i) The NPMLE \(\hat{F}_M\) is strongly consistent, i.e. \(\sup_{x \geq 0} |\hat{F}_M(x) - F_0(x)| = 0 \) a.s., and (ii) \(\sqrt{n}(\hat{F}_M(x) - F_0(x))\) converges in distribution to a normal variate.

**Proof.** Let \(Q\) denote the empirical version of the joint distribution function of \((L_i, R_i)\) \((i = 1, \ldots, n)\). It follows that Eq. (2.8) can be written as

\[
\begin{align*}
F(t) &= \int_{t \leq r} dQ_n(l, r) + \int_{t \leq l} \frac{F(t) - F(l)}{F(r) - F(l)} dQ_n(l, r),
\end{align*}
\]

(2.15)

where \(Q_n\) is the empirical version of \(Q\). In (2.15), if \(F(t) = F(r) = F(l)\), then we encounter \(\frac{0}{0}\) in the integrand. In this case, we define \(\frac{0}{0} = 1\). Notice that Eq. (2.15) is exactly the same as Eq. (2.3) of Yu et al. (2000) and Eq. (2.2) of Yu et al. (2001), which is a self-consistent equation of \(F_0\) for the model in Yu et al. (2000, 2001) with mixed interval-censored (MIC) data. By Theorems 2.1, 2.2 and 3.1 of Yu et al. (2001), the strong consistency and asymptotic normality of \(\hat{F}_M\) and \(\hat{F}_S\) are established. \(\Box\)

3. Discussion

We have demonstrated how middle-censored data relate to mixed interval-censored data. With some modification of the definition for intervals \((q_j, p_j)\)’s, we can obtain the NPMLE of distribution function by using EM algorithm of Turnbull (1976) or self-consistent estimating equation (Jammalamadaka and Mangalam, 2003) with a proper initial estimator. The consistency and asymptotic normality of the NPMLE can be established based on the asymptotic properties of self-consistent estimators (SCE) with mixed interval-censored data (Yu et al., 2000, 2001).

Acknowledgement

The author would like to thank the associate editor and the referees for their helpful and valuable comments and suggestions.

**References**

Alioum, A., Commenges, D., 1996. A proportional hazards model for arbitrarily censored and truncated data. Biometrics 52, 512–524.

Efron, B., 1967. The two-sample problem with censored data. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley.

Frydman, H., 1994. A note on nonparametric estimation of the distribution function from interval-censored and truncated data. J. Roy. Statist. Soc. Ser. B 56, 71–74.

Gentleman, R., Geyer, C., 1994. Maximum likelihood for interval censored data: consistency and computation. Biometrika 81, 618–623.

Gentleman, R., Vandal, A.C., 2001. Computational algorithm for censored-data problems using intersection graphs. J. Comput. Graphical Statist. 10, 403–421.

Huang, J., 1999. Asymptotic properties of nonparametric estimation based on partly interval-censored data. Statist. Sinica 9, 501–520.

Hudgens, M.G., 2003. On nonparametric maximum likelihood estimation with interval censoring and truncation. J. Roy. Statist. Soc. Ser. B 67 (Part 4), 573–587.
Jammalamadaka, S.R., Mangalam, V., 2003. Non-parametric estimation for middle-censored data. J. Nonparametr. Statist. 15, 253–265.
Jammalamadaka, S.R., Iyer, S.K., 2004. Approximate self consistency for middle-censored data. J. Statist. Plann. Inference 124, 75–86.
Kaplan, E.L., Meier, P., 1958. Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457–481.
Li, L., Watkins, T., Yu, Q.Q., 1997. An EM algorithm for smoothing the self-consistent estimator of survival functions with interval-censored data. Scand. J. Statist. 24, 531–542.
Mangalam, V., Nair, G.M., Zhao, Y., 2008. On computation of NPMLE for middle-censored data. Statist. Probab. Lett. 78, 1452–1458.
Mykland, P.A., Ren, J., 1996. Algorithms for computing self-consistent and maximum likelihood estimators with doubly censored data. Ann. Statist. 24, 1740–1764.
Peto, R., 1973. Experimental survival curves for interval-censored data. Appl. Statist. 22, 86–91.
Tsai, W.Y., Crowley, J., 1985. A large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency. Ann. Statist. 13, 1317–1334.
Turnbull, B.W., 1976. The empirical distribution with arbitrarily grouped, censored and truncated data. J. Roy. Statist. Soc. B 38, 290–295.
Yu, Q.Q., Li, L., Wong, G.Y.C., 1998. Asymptotic variance of the GMLE of a survival function with interval-censored data. Sankhyâ 60, 184–197.
Yu, Q.Q., Li, L., Wong, G.Y.C., 2000. On consistency of the self-consistent estimator of survival function with interval censored data. Scand. J. Statist. 27, 35–44.
Yu, Q.Q., Wong, G.Y.C., Li, L., 2001. Asymptotic properties of self-consistent estimators with mixed interval-censored data. Ann. Inst. Statist. Math. 53, 469–486.