ASYMPTOTIC VARIANCE OF THE GMLE OF A SURVIVAL FUNCTION WITH INTERVAL-CENSORED DATA

By QIQING YU*
State University of New York at Binghamton, New York
LINXIONG LI**
University of New Orleans, Los Angeles
and
GEORGE Y.C. WONG***
Strang Cancer Preventive Center, New York

SUMMARY. Interval-censored data are generated by a random survival time \( X \) and a random censoring interval. We either observe the exact survival time or only know the survival time lies within the censoring interval. Turnbull (1976) proposes a self-consistent algorithm for obtaining the generalized maximum likelihood estimator (GMLE) of a survival function with interval-censored data. Yu, Li and Wong (1996) prove the strong consistency of the GMLE. In this paper, we establish the asymptotic normality of the GMLE and self-consistent estimators (SCE) and present a consistent estimator of the asymptotic variance of the GMLE and SCEs with interval-censored data.

1. Introduction

We consider the nonparametric estimation of distribution function \( F \) of a survival time \( X \) with incomplete observations due to interval censoring. Interval-censored (IC) data are bivariate observations \( (L_i, R_i) \), \( i = 1, ..., n \), where \( L_i \leq R_i \). If \( L_i = R_i \), then a survival time \( X_i = L_i = R_i \) is observed and we say it is an exact observation; if \( L_i < R_i \), then \( X_i \) is censored and a censoring interval \( [L_i, R_i] \) is observed instead.
Recent studies of interval censoring have focused on case 2 interval-censored (IC) data, which involve a time-to-event variable $X$ whose value is never observed but is known to lie in the time interval between two consecutive inspection times $U$ and $V$. Case 2 interval censoring arises naturally in a longitudinal follow-up study in which the event of interest cannot be easily observed (for instance, cancer recurrence, elevation of levels of a biomarker without any noticeable symptoms).

In this paper, we consider IC data which consist of both case 2 IC data and exact observations. We call such data mixed IC data. An example of such data from the Framingham Heart Study was presented by Odell et al. (1992).

To formulate a model for such data, let $(Y, Z)$ be a random censoring vector having distribution function $G(y, z)$, and let $T \geq 0$ be a random variable having distribution function $G_T(t)$. Assume that $0 \leq Y < Z \leq \tau_r$ with probability one, where $\tau_r = \sup\{t; P(\min(X, T) \leq t) < 1\}$. We say a data point is from a case 2 interval censorship (C2) model (Groeneboom and Wellner, 1992) if the corresponding observation is $(Y, Z, 1(X \leq Y), 1(X \leq Z))$, where $1(\cdot)$ is the indicator function; and a data point is from a right censorship (RC) model if the corresponding observation is $(\min(X, T), 1(X \leq T))$. Thus an observation obtained from the C2 model is always censored and that from the RC model is either right-censored or exact. IC data can be viewed as a mixture of data from a C2 model and a RC model. We introduce a random variable, $D$, to distinguish failure times coming from the two models:

$$D = \begin{cases} 1 & \text{if the observation is from the RC model,} \\ 0 & \text{if the observation is from the C2 model.} \end{cases}$$

Let $P\{D = 1\} = \pi$. We assume $0 < \pi \leq 1$ and $D, X, T$ and $(Y, Z)$ are independent. Formally, the observable random interval $\{L, R\}$ can be expressed as follows.

$$\{L, R\} = \begin{cases} (0, Y] & \text{if } D = 0 \text{ and } X \leq Y, \\ (Y, Z] & \text{if } D = 0 \text{ and } Y < X \leq Z, \\ (Z, \infty) & \text{if } D = 0 \text{ and } X > Z, \\ (T, \infty) & \text{if } D = 1 \text{ and } X > T, \\ [X, X] & \text{if } D = 1 \text{ and } X \leq T. \end{cases} \quad (1.1)$$

We call such a model an IC model.

Under the RC model, the product limit estimator (Kaplan and Meier, 1958) of $F$ is the generalized maximum likelihood estimator (GMLE), which maximizes the generalized likelihood function, and has been studied by many authors. Under the C2 model and under certain smoothness conditions, Groeneboom and Wellner (1992) show that the GMLE of $F$ is strongly consistent and conjecture that the convergence rate of the GMLE is $(n \ln n)^{1/3}$. With IC data, Peto (1973) proposes the GMLE $\hat{F}$ of $F$. Turnbull (1976) proposes a self-consistent algorithm for obtaining the GMLE of $F$. It is known that the GMLE is a self-consistent estimator (SCE) (see, e.g., Gu and Zhang, 1993). Yu, Li and Wong (1996)
establish the strong consistency of the GMLE and SCEs under the following assumption.

Assumption 1. $F$ is arbitrary and $(Y, Z, T)$ takes on finitely many values.

However, the asymptotic distribution of the GMLE based on IC data has not been discussed. Due to lack of consistent estimators of the asymptotic variance of the GMLE, the application of the GMLE of $\hat{F}$ with IC data has been limited. A common practice in medical research with IC data is to treat the right endpoint $r_i$ as an exact observation if $r_i$ is not $\infty$ (see, e.g., Samuelsen and Kongerud, 1994). Then the data set is treated as a right-censored data set and thus standard statistical tools can be used.

In Section 2, we introduce notations and necessary background. In Section 3, we prove the asymptotic normality of the GMLE and SCEs and present a consistent estimator of the asymptotic variance. The proof of a technical lemma in Section 3 is provided in the appendix.

2. Notations and Definitions

We shall establish our results in two steps. First obtain the asymptotic results under the following stronger assumption.

Assumption 2. $(L, R)$ takes on finitely many distinct values $(l_1, r_1), \ldots, (l_g, r_g)$, with probabilities $p_1, \ldots, p_g$, where $p_i > 0$ and $\sum_{i=1}^{g} p_i = 1$.

Then, at the second step we extend the result to the case where a weaker assumption, Assumption 1, is assumed.

Suppose that the IC data consist of $N_1$ pairs of $(l_1, r_1)$, $N_2$ pairs of $(l_2, r_2), \ldots,$ and $N_g$ pairs of $(l_g, r_g)$, where $N_i \geq 0$ and $N_1 + \cdots + N_g = n$. Let $I_1, \ldots, I_n$ denote the observed intervals. Define innermost intervals $A_j, j = 1, 2, \ldots, m$ induced by $I_1, \ldots, I_n$ to be all the disjoint intervals which are non-empty intersections of these $I_i$’s such that for all possible $i$ and $j$, $A_j \cap I_i = \emptyset$ or $A_j$. Let the endpoints of the innermost intervals be $a_j$ and $b_j$, $j = 1, \ldots, m$, where $a_1 \leq b_1 \leq a_2 \leq b_2 \leq \cdots \leq a_m \leq b_m$. Let $\delta_{ij} = 1(A_j \subset I_i)$. The following example illustrates the procedure of finding innermost intervals.

Example 2.1. Suppose that there are five observed intervals $(1, 4], (2, 2], (2, 6], [5, 5],$ and $(1, 6]$. Then there are two exact observations, $I_1 = [2, 2]$ and $I_2 = [5, 5]$, and three censored intervals, $I_3 = (1, 4], I_4 = (2, 6]$ and $I_5 = (1, 6]$. Furthermore, there are three innermost intervals, $A_1 = [2, 2], A_2 = [5, 5],$ and $A_3 = (2, 4]$.

Peto (1973) shows that the GMLE of $F$ assigns weight, say $s_1, \ldots, s_m$, to the corresponding innermost intervals $A_1, \ldots, A_m$ only. The generalized likelihood
function $L^*$ (Kiefer and Wolfowitz, 1956) can be simplified as

$$L^* = L^*(s_1, ..., s_m) = \prod_{i=1}^{n} \sum_j \delta_{ij} s_{j}^i,$$  \hspace{1cm} \ldots (2.1)

where $s^i = (s_1, ..., s_{m-1}) \in D_s$ is the transpose of $s$, $D_s = \{s; s_i \geq 0, s_1 + \cdots + s_{m-1} \leq 1\}$ and $s_m = 1 - s_1 - \cdots - s_{m-1}$. Turnbull (1976) proposes a self-consistent algorithm for obtaining the GMLE via an iterative procedure as follows. At step 1, let $s_j^{(1)} = 1/m$ for $j = 1, ..., m$. At step $h$,

$$s_j^{(h)} = \frac{\sum_{i=1}^{n} \frac{1}{n} \delta_{ij} s_{j}^{(h-1)}}{\sum_{k=1}^{m} \delta_{ik} s_{k}^{(h-1)}}, \hspace{1cm} j = 1, ..., m, \hspace{0.5cm} h \geq 2.$$

He shows that, as $h \to \infty$, $s_j^{(h)}$ converges to the GMLE, $\hat{s}_j$, which maximizes $L^*$ and satisfies the system of self-consistent equations

$$\hat{s}_j = \frac{\sum_{i=1}^{n} \frac{1}{n} \delta_{ij} \hat{s}_j}{\sum_{k=1}^{m} \delta_{ik} \hat{s}_k}, \hspace{1cm} j = 1, ..., m. \hspace{1cm} \ldots (2.2)$$

A solution $s = \hat{s}$ to (2.2) is called a self-consistent estimator of $s$ if $\hat{s} \in D_s$. An estimate $\hat{F}(t)$ of $F(t)$ can be uniquely defined for $t \in [b_i, a_{i+1})$ by $\hat{F}(b_i) = \hat{F}(a_{i+1}) = \hat{s}_1 + \cdots + \hat{s}_i$, but is not uniquely defined for $t$ being in an open innermost interval (Peto, 1973; Turnbull, 1976). To avoid this ambiguity we define

$$\hat{F}(t) = \begin{cases} \hat{s}_1 + \cdots + \hat{s}_i & \text{if } t \in (b_i, a_{i+1}], \\ \hat{s}_i + \cdots + \hat{s}_{i-1} + \frac{t-a_{i-1}}{a_i-a_{i-1}} \hat{s}_i & \text{if } t \in (a_i, b_i], a_i > 0 \text{ and } b_i \neq \infty, \\ \hat{s}_i & \text{if } 0 = a_1 \leq t \leq b_1, \\ 1 - \hat{s}_m & \text{if } t \geq a_m \text{ and } a_m < b_m = \infty, \end{cases} \hspace{1cm} \ldots (2.3)$$

where $(a, b]$ is an empty set if $a = b$. Under the RC model, the definition of $\hat{F}$ given by (2.3) reduces to the PLE (Gill, 1983) for $t \geq a_m$ when $b_m = \infty$, and possesses the uniform strong consistency. If we define $\hat{F}(t) = 0$ for $t \in [0, b_1]$ when $a_1 = 0$, or $\hat{F}(t) = 1$ for $t \in [a_m, \infty)$ when $b_m = \infty$, the uniform strong consistency of $\hat{F}(t)$ does not hold (see Yu and Li, 1994 or Stute and Wang, 1993).

Let $A_j^i$, $j = 1, ..., m$, be the innermost intervals induced by the $g$ intervals $\{l_i, r_i\}, i = 1, 2, \ldots, g$. To distinguish $A_j^{i}$ from the innermost intervals $A_j^i$ induced by the observed $I_i$’s, we call $A_j^{i}$ population innermost intervals. Let $s^*_{ij}$ be the weight assigned by the distribution function $F$ to $A_j^{i}$, i.e., $s^*_{ij} = P\{X \in A_j^{i}\}$. Note that $m = m$ or $A_j = A_j^i$ may not be true, since $N_i$ may be zero for some $i$. However, for every $i$, $\lim_{n \to \infty} N_i/n = p_i > 0$ almost surely and we
are investigating the asymptotic properties of \( \hat{F} \), thus, by taking large sample size \( n \), we can without loss of generality (WLOG) assume that \( N_i > 0 \) for all \( i = 1, 2, \ldots, g \). As a consequence, we have \( m_o = m \) and \( A_j = A^o_j \). WLOG, we can assume \( I_i = \{l_i, r_i\} \), \( i = 1, \ldots, g \). Then (2.3) can be changed to the following equivalent form,

\[
\hat{s}_j = \sum_{i=1}^{g} \frac{N_i}{n} \frac{\delta_{ij} \hat{s}_j}{\sum_{k=1}^{m} \delta_{ik} \hat{s}_k}, \quad j = 1, \ldots, m, \quad \ldots (2.4)
\]

Replacing \( N_i / n \) in (2.4) by \( p_i \), we obtain

\[
s_j = \sum_{i=1}^{g} \frac{p_i}{m} \frac{\delta_{ij} s_j}{\sum_{k=1}^{m} \delta_{ik} s_k}, \quad j = 1, \ldots, m. \quad \ldots (2.5)
\]

The following theorem is proved by Yu, Li and Wong (1996).

**Theorem 1.** Under the IC model and Assumption 2, \( s^o = (s^o_1, \ldots, s^o_m) \) is the unique solution of the equations (2.5), and for all SCE \( \hat{s} \), \( \lim_{n \to \infty} \hat{s}_i = s^o_i \) almost surely, \( i = 1, 2, \ldots, m \).

### 3. Main Result

In this section, we develop the asymptotic normality of the GMLE and SCEs \( \hat{s} \) and \( \hat{F}(t) \) under the IC model. We first consider Assumption 2. By the end of this section, we will relax it to Assumption 1.

#### 3.1. Results under Assumption 2

Note that each value \( x_i \in [\tau_l, \tau_r] \) of \( X \) constitutes a (population) innermost interval, where \( \tau_l = \sup \{x : F(x) = 0\} \). Suppose that \( A^o_1 < \cdots < A^o_m \) are the population innermost intervals. For the later development, we need to assume

\[
s^o_i > 0 \text{ for } i = 1, \ldots, m, \text{ and } \sum_{i=1}^{m} s^o_i = 1 \quad \ldots (3.1)
\]

Thus we assume that there exist at least \( m - 1 \) points \( x_1, x_2, \ldots, x_{m-1} \in [\tau_l, \tau_r] \) such that \( x_1 < x_2 < \cdots < x_{m-1} \) and \( A^o_i = \{x_i\} \). Under Assumption 2, assumption (3.1) is equivalent to

**Assumption 3.** \( \textbf{P}\{X \in (l, r]\} > 0 \) for all \( l < r \), where \( l \) and \( r \) are realizations of \( L \) and \( R \), respectively.
Let \( \hat{p} = (\hat{p}_1, ..., \hat{p}_{g-1}) \), then \((N_1, ..., N_g)\) follows a multinomial distribution \( M(n; p_1, ..., p_g) \). Denote \( \Sigma_{\hat{p}} \) the \((g-1) \times (g-1)\) covariance matrix of \( \hat{p} \), that is,

\[
\Sigma_{\hat{p}} = \begin{pmatrix} \alpha_{ij} \end{pmatrix}_{(g-1) \times (g-1)}; \quad \alpha_{ij} = \begin{cases} p_i(1-p_i)/n & \text{if } i = j, \\ -p_ip_j/n & \text{otherwise}. \end{cases}
\]

Then, it follows from Rao (1973, p.382) that

\[
\Sigma_{\hat{p}}^{-1/2}(\hat{p} - \mathbf{p}) \xrightarrow{D} N(\mathbf{0}, \mathbf{3}) \quad \ldots(3.2)
\]

where \( \mathbf{O} \) is a \((g-1) \times 1\) zero vector and \( \mathbf{3} \) is a \((g-1) \times (g-1)\) identity matrix.

To find the asymptotic distribution of \( \hat{s} \), we use the result (3.2) and the connection between \( \mathbf{p} \) and \( \mathbf{s} \). Theorem 1 implies that \( \mathbf{s} \) can be implicitly expressed as a function of \( \mathbf{p} \). For convenience we use \((s_1^1, ..., s_m^1)\) instead of \((s_1^o, ..., s_m^o)\) for the rest of the paper. Since \( s_j > 0 \) for all \( j \), (2.5) becomes

\[
1 = g \sum_{i=1}^{g} p_i \delta_{ij} \frac{\delta_{ik}s_k}{m \sum_{k=1}^{g} \delta_{ik}s_k}, \quad j = 1, ..., m. \quad \ldots(3.3)
\]

Taking partial derivatives \( \frac{\partial}{\partial \mathbf{p}_h} \) on both sides of the above equations yields

\[
0 = \sum_{i=1}^{g} \frac{\partial p_i}{\partial \mathbf{p}_h} \frac{\delta_{ij}}{m \sum_{k=1}^{g} \delta_{ik}s_k} - \sum_{i=1}^{g} p_i \frac{\delta_{ij} m \sum_{l=1}^{m} \frac{\partial s_l}{\partial \mathbf{p}_h}}{(m \sum_{k=1}^{g} \delta_{ik}s_k)^2}
\]

\[
= \sum_{k=1}^{m} \frac{\delta_{sj}}{\delta_{hk}s_k} - \sum_{k=1}^{m} \frac{\delta_{sj}}{\delta_{gk}s_k} - \sum_{i=1}^{g} p_i \frac{\delta_{ij} m \sum_{l=1}^{m} \left[ \delta_{il} \frac{\partial s_l}{\partial \mathbf{p}_h} - \delta_{im} \frac{\partial s_l}{\partial \mathbf{p}_h} \right]}{(m \sum_{k=1}^{g} \delta_{ik}s_k)^2}
\]

\[
= \sum_{k=1}^{m} \frac{\delta_{sj}}{\delta_{hk}s_k} - \sum_{k=1}^{m} \frac{\delta_{sj}}{\delta_{gk}s_k} - \sum_{i=1}^{g} p_i \frac{\delta_{ij} m \sum_{l=1}^{m-1} \left[ \delta_{il} - \delta_{im}/(m-1) \right] \frac{\partial s_l}{\partial \mathbf{p}_h}}{(m \sum_{k=1}^{g} \delta_{ik}s_k)^2},
\]

\( j = 1, ..., m-1, \quad h = 1, ..., g-1 \). Thus,

\[
\sum_{l=1}^{m-1} \left\{ \sum_{i=1}^{g} p_i \frac{\delta_{ij} \left[ \delta_{il} - \delta_{im}/(m-1) \right]}{(m \sum_{k=1}^{g} \delta_{ik}s_k)^2} \right\} \frac{\partial s_l}{\partial \mathbf{p}_h} = \sum_{k=1}^{m} \frac{\delta_{sj}}{\delta_{hk}s_k} - \sum_{k=1}^{m} \frac{\delta_{sj}}{\delta_{gk}s_k}.
\]

\[
\ldots(3.4)
\]
Rewrite the system of linear equations in (3.4) as

$$C \left( \frac{\partial s}{\partial p_1}, \ldots, \frac{\partial s}{\partial p_{g-1}} \right) = \left( w_1, \ldots, w_{g-1} \right), \ldots (3.5)$$

where

$$C = \left( c_{jl} \right)_{(m-1) \times (m-1)}$$

with

$$c_{jl} = \sum_{i=1}^{g} p_i \frac{\delta_{ij} \delta_{il} - \delta_{im}}{(\sum_{k=1}^{m} \delta_{ik} s_k)^2}, \quad \frac{\partial s}{\partial p_h} = \left( \frac{\partial s_1}{\partial p_h}, \ldots, \frac{\partial s_{g-1}}{\partial p_h} \right),$$

and

$$w_h = \left( w_{1h} \ldots w_{(m-1)h} \right), \quad w_{jh} = \delta_{hj} \frac{\delta_{gj} - \delta_{gj}}{(\sum_{k=1}^{m} \delta_{gk} s_k)} + \delta_{gh}/(m-1), \quad h = 1, \ldots, g-1, \quad j = 1, \ldots, m-1.$$

Note that $C$ is independent of $h$. Rewrite

$$C = \left( \sum_{i=1}^{g} p_i \frac{\delta_{ij} \delta_{il} - \delta_{im}}{(\sum_{k=1}^{m} \delta_{ik} s_k)^2} \right)_{(m-1) \times (m-1)} = UDV,$$

where

$$U = \begin{pmatrix} \delta_{11} & \ldots & \delta_{g1} \\ \vdots & \ddots & \vdots \\ \delta_{1(m-1)} & \ldots & \delta_{g(m-1)} \end{pmatrix}, \quad D = \begin{pmatrix} \frac{p_1}{\sum_{k=1}^{m} \delta_{1k} s_k} & 0 & \ldots & 0 \\ 0 & \frac{\sum_{k=1}^{m} \delta_{2k} s_k^2}{(\sum_{k=1}^{m} \delta_{1k} s_k)^2} & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & \frac{p_g}{\sum_{k=1}^{m} \delta_{gk} s_k^2} \end{pmatrix},$$

and

$$V = \begin{pmatrix} \delta_{11} - \delta_{1m}/(m-1) & \ldots & \delta_{1(m-1)-\delta_{1m}/(m-1)} \\ \vdots & \ddots & \vdots \\ \delta_{g1} - \delta_{g(m-1)-\delta_{g(m-1)}/(m-1)} & \ldots & \delta_{g(m-1)} - \delta_{g(m-1)}/(m-1) \end{pmatrix}.$$

By relabeling the observed intervals $I_i$, we can assume that $I_j = A_j^o$ for $1 \leq j \leq m-1$. In other words, $\delta_{ij} = 1$ if $i = j$ and 0 otherwise for $1 \leq i, j \leq m-1$. Thus $U = (3, U_2)$ where $3$ is the $(m-1) \times (m-1)$ identity matrix and $U_2$ is a $(m-1) \times (g-m+1)$ matrix. Therefore, the rank of $U$ is of $m-1$. Similarly we can show that the rank of $V$ is also $m-1$ by observing that $\delta_{im} = 0$ for $1 \leq i \leq m-1$. The rank of the matrix $D$ is $g$ ($\geq m$) since all $p_i > 0$. Thus, it can be shown that $C$ is nonsingular (see Lemma A in the appendix). Moreover,
by letting $I_g = A_m$,

$$\begin{pmatrix} 1 & \cdots & 0 & \cdots \\ \sum_{k=1}^m \delta_{1k}s_k & \cdots & \delta_{1(m-1)}s_k \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \cdots & 1 & \cdots \\ \sum_{k=1}^m \delta_{(m-1)k}s_k & \cdots & \delta_{(m-1)(m-1)}s_k \end{pmatrix} \quad (3.6)$$

and thus it is of rank $m - 1$. It follows from (3.5) that $\frac{\partial s}{\partial p_h} = C^{-1}w_h$, $h = 1, \ldots, g - 1$. Let $\frac{\partial s}{\partial p} = (\frac{\partial s_l}{\partial p})_{(m-1) \times (g-1)}$. By the first order Taylor expansion, $\hat{s} - s \approx \frac{\partial s}{\partial p}(\hat{p} - p)$. The validity of the approximation can be verified by the fact that the second order partial derivatives $\frac{\partial^2 s}{\partial p_t \partial p_h}$ exist and are continuous. In fact, taking partial derivatives on both sides of (3.5) yields

$$U(\frac{\partial}{\partial p_t} D)V \frac{\partial s}{\partial p_h} + C \frac{\partial^2 s}{\partial p_t \partial p_h} = \frac{\partial}{\partial p_t} w_h, \; t, h = 1, \ldots, g - 1. \quad (3.7)$$

Thus

$$\frac{\partial^2 s}{\partial p_t \partial p_h} = C^{-1}(\frac{\partial}{\partial p_t} w_h - U(\frac{\partial}{\partial p_t} D)V \frac{\partial s}{\partial p_h}), \; t, h = 1, \ldots, g - 1,$$

where

$$\frac{\partial}{\partial p_t} w_h = \left( \frac{\delta_{ij}}{(\sum_{k=1}^m \delta_{ik}s_k)^2} - \frac{\delta_{ij}}{(\sum_{k=1}^m \delta_{ik}s_k)^2} \right)_{(m-1) \times 1}$$

and $\frac{\partial}{\partial p_t} D$ is a diagonal matrix with the $i$-th diagonal element

$$\frac{1(i = t)}{(\sum_{k=1}^m \delta_{ik}s_k)^2} - 2\frac{p_i(\sum_{l=1}^m \delta_{il} \frac{\partial s_l}{\partial p_i})}{(\sum_{k=1}^m \delta_{ik}s_k)^3}.$$

In view of (3.7), the second order partial derivatives of $s = s(p)$ exist. Thus, by (2.5), (3.6) and the well-known result of multivariate normal convergence theorem (see, for example, Rao, 1973, p.387), we have the following result.

**Theorem 2.** Under the IC model and Assumptions 2 and 3

$$\Sigma^{-1/2}_\hat{s}(\hat{s} - s) \xrightarrow{p} N(0, \Omega) \; \text{as} \; n \to \infty, \; \text{where} \; \Sigma_\hat{s} = (\frac{\partial s}{\partial p})^T \Sigma p (\frac{\partial s}{\partial p})^T,$$
**Theorem 3.** Under the assumptions of Theorem 2, we have
\[
\left( e_x \Sigma^\dagger \right)^{-1/2} (\hat{F}(x) - F(x)) \xrightarrow{D} N(0,1) \quad \text{as } n \to \infty,
\]
for \( x \in \mathcal{B} \), where \( \hat{F}(x) \) is given by (2.3) and \( \mathcal{B} = \{ x : F(x) \in (0,1), \ x \leq \tau \} \).

Verify that \( \hat{F}(x) = 0 \) for \( x < \tau \) (and thus \( x \notin \mathcal{B} \)). It is obvious that then the theorem is not true. An estimator \( \hat{\Sigma} \) can be obtained by replacing \( \mathbf{p} \) and \( \mathbf{s} \) in \( \hat{\Sigma} \) by \( \hat{\mathbf{p}} \) and \( \hat{\mathbf{s}} \), respectively. By the strong consistency of \( \hat{\mathbf{p}} \) and \( \hat{\mathbf{s}} \) in Theorem 1 it is easy to see that the estimator \( \hat{\Sigma} \) is strongly consistent. Then, as an immediate consequence of Theorem 3, we have the following theorem.

**Theorem 4.** Under the assumptions of Theorem 2, we have,
\[
\frac{\hat{F}(x) - F(x)}{\hat{\sigma}} \xrightarrow{D} N(0,1)
\]
as \( n \to \infty \) for \( x \in \mathcal{B} \), where \( \hat{\sigma}^2 = e_x \hat{\Sigma} e_x^t \).

3.2 Results under Assumption 1. In this subsection, we show that the results in section 3.1 are also valid under Assumptions 1 and 3. For simplicity, we only give the proof of the extension of Theorem 3.

**Theorem 5.** Under the IC model and Assumptions 1 and 3,
\[
\left( e_x \Sigma^\dagger \right)^{-1/2} (\hat{F}(x) - F(x)) \xrightarrow{D} N(0,1) \quad \text{as } n \to \infty, \quad \text{for } x \in \mathcal{B}. \quad \ldots \quad (3.8)
\]

**Proof.** For any arbitrary distribution function \( F \) and fixed \( x_0 \in [\tau_l, \tau_r] \), choose finitely many points, say \( \xi_1 < \cdots < \xi_{M/2+1} \), such that \( x_0 \), \( T \), \( Y \), \( Z \in \{ \xi_1, \ldots, \xi_{M/2+1} \} \) w.p.1 with \( \xi_1 = \tau_l, \xi_{M/2} = \tau_r \) and \( \xi_{M/2+1} = \infty \), where \( M \) is an even positive integer. Denote \( B_{2i-1} = \{ \xi_i \} \) and \( B_{2i} = (\xi_i, \xi_{i+1}) \), \( i = 1, \ldots, (M/2) \), and denote the midpoint of the set \( B_j \) by \( m_j \), where the midpoint of an infinite interval \((b, +\infty)\) is defined to be \( b+1 \). Let \( B_{k1}, \ldots, B_{km^*} \) (for some \( m^* \)) be all the distinct elements of \( \{ B_k : k = 1, \ldots, M \} \) such that \( s_i^* = \text{P}(X \in B_{ki}) > 0 \). Define a new distribution function \( F^* \) by \( F^*(t) = \sum_{i : m_i \leq t} s_i^* \), and define pseudo observations
\[
I_j = \begin{cases} 
I_j & \text{if } I_j \text{ is non-singleton,} \\
\{ m_i \} & \text{if } I_j \text{ is a singleton and } I_j \subset B_i.
\end{cases}
\]
Let \( \hat{s} = (\hat{s}_1, \ldots, \hat{s}_m) \) \( (\hat{F}) \) be an SCE of \( s \) \( (F) \) based on observations \( I_i, i = 1, \ldots, n \), i.e., \( \hat{s} \) satisfies (2.2). Let \( A_1 < \cdots < A_m \) be the innermost intervals induced by
the original data. As assumed, \( n \) is large, thus all innermost intervals, except perhaps \( A_{m} \), are singleton set due to Assumption 3. Let \( A_{i}^{*} = \{ k_{i} \} \), \( i = 1, \ldots, m^{*} - 1 \) and

\[
A_{m^{*}}^{*} = \begin{cases} \{ k_{m^{*}} \} & \text{if } \sup \{ x : x \in B_{k_{m^{*}}} \} < +\infty, \\ B_{k_{m^{*}}} & \text{otherwise}. \end{cases}
\]

Then the definition of \( \{ \xi_{i} \} \) implies that

\[
I_{i} \cap A_{*}^{j} = \text{either } \emptyset \text{ or } A_{*}^{j} \text{ or } I_{i}. \tag{3.9}
\]

WLOG, we can assume that \( I_{*}^{1}, \ldots, I_{*}^{g} \) are all the distinct elements among \( I_{1}^{*}, \ldots, I_{n}^{*} \). Verify that \( A_{*}^{1}, \ldots, A_{*}^{m^{*}} \) are all the innermost intervals induced by \( I_{1}^{*}, \ldots, I_{n}^{*} \). Let \( \tilde{s} \) be such that

\[
\tilde{s}_{j} = \sum_{h : A_{h} \subset B_{kj}} \tilde{s}_{h}, \quad j = 1, \ldots, m^{*}. \tag{3.10}
\]

It follows from (3.9) and (3.10) that \( \tilde{s} \) satisfies equations

\[
\tilde{s}_{j} = \sum_{i} \frac{N_{s}^{*} \delta_{ij}^{s} \tilde{s}_{j}}{n} \sum_{k=1}^{m^{*}} \delta_{ik}^{*} \tilde{s}_{k}, \quad j = 1, \ldots, m^{*},
\]

where the summation is over distinct \( I_{i}^{*} \)'s, \( \delta_{ij}^{*} = 1(A_{j}^{*} \subset I_{i}^{*}) \) and \( N_{s}^{*} = \sum_{j=1}^{n} 1(I_{j}^{*} = I_{*}^{j}) \). Verify that

\[
N_{i}^{*} = \begin{cases} N_{j} & \text{if } I_{i}^{*} \text{ is a non-singleton and } I_{i}^{*} = I_{j}, \\ \sum_{j=1}^{g^{*}} N_{j} 1(I_{j} \subset B_{k_{h}}, \ I_{j} \text{ is a singleton}) & \text{if } I_{i}^{*} = \{ k_{m} \}, \end{cases}
\]

for \( i = 1, \ldots, g^{*} \). Thus \( \tilde{s} \) is an SCE of \( s^{*} = (s_{1}^{*}, \ldots, s_{m^{*}}^{*}) \) based on the pseudo observations \( I_{i}^{*}, i = 1, \ldots, n \). It is important to note that 1. \( F^{*} \) takes on finitely many (\( \leq m^{*} \)) values,

2. \( s_{j}^{*} = dF^{*}(A_{j}^{*}) = dF(B_{k_{j}}) > 0 \), and

3. pseudo observations \( I_{1}^{*}, \ldots, I_{n}^{*} \), generated by \( F^{*}, G_{T} \) and \( G \), are i.i.d. since original observations \( I_{1}, \ldots, I_{n} \), generated by \( F, G_{T} \) and \( G \), are i.i.d.

Thus Theorems 1, 2 and 3 apply to \( \tilde{s} \). It is readily seen from the definitions of \( \{ \xi_{i} \} \) and \( \tilde{s} \) that \( \hat{F}(x_{0}) = \sum_{j : m_{k_{j}} \leq x_{0}} \tilde{s}_{j} \), thus

\[
(e_{x_{0}} \sum_{j} e_{x_{0}}) - 1/2(\hat{F}(x_{0}) - F(x_{0})) \xrightarrow{D} N(0, 1) \quad \text{as } n \to \infty.
\]
It follows from (2.3) that for \( x \leq \tau_r \)
\[
(e_x \Sigma_x e_x^T)^{-1/2} (\hat{F}(x) - F(x)) \xrightarrow{D} N(0,1) \quad \text{as} \quad n \to \infty. \quad \Box
\]

3.3 Comments. Some comparisons between the existing results for the C2 model and our results here are provided below.

Groeneboom and Wellner (1992) conjectured that the convergence of the GMLE is at a slower rate of \((n \log n)^{1/3}\) under the C2 model with the assumption that all the random variables are absolutely continuous. Under our assumption 2, for both C2 and IC models, the estimation of \( F \) becomes a parametric estimation problem of a finite dimensional multinomial distribution. Consequently, the GMLE should have the usual parametric \( n^{1/2} \) convergence rate. Note that under the C2 model one can only estimate \( F \) at the values of \( Y \) or \( Z \), while under the IC model we can estimate \( F \) at all values of \( X \) that are within \([\tau_l, \tau_r]\) because of the existence of exact observations.

Under the C2 model and assumptions 2 and 3, Yu et al. (1998) show that the GMLE of \( F(x) \) satisfies (3.8) for \( x \) at the values of \( Z \) or \( Y \). In addition, it is worth noting that under assumptions 2 and 3, the solution to the limiting self-consistent equation (2.5) may not be unique, i.e., Theorem 1 fails to be true under the C2 model. Thus (3.8) may not be true for arbitrary SCEs under the C2 model. However, under the IC model and assumptions 2 and 3, as we have seen in this paper, all SCEs are consistent and satisfy (3.8), due to exact observations.

Appendix

**Lemma A.** The matrix \( C \) in (3.5) is nonsingular.

**Proof.** Use the notations previously defined in Section 2. By Assumption 2, WLOG, we can assume that the supports of \( X, T, Y \) and \( Z \) are bounded by some positive number \( b \). Let \( X^* \) be a new random variable, independent of \( (X, T, Y, Z) \), with distribution function \( F_*(x) = (1 - \epsilon)F(x) + \epsilon 1(x \leq b) \), where \( \epsilon \in (0,1) \).

Define \( A_i^* = A_i \) for \( i = 1, \ldots, m - 1 \), where \( A_i \) are defined as before, and define \( A_m^* = [b, b] \) and \( A_{m+1}^* = A_m \). Let \( T, Y \) and \( Z \) be as before. Then there are \( g + 1 \) distinct values of \((L^*, R^*)\), where \( L^* \), etc. are defined accordingly in an obvious way. Let the intervals induced by these pairs be

\[
I_i^* = \begin{cases} 
I_i & \text{if } i = 1, \ldots, m - 1, \\
[b, b] & \text{if } i = m, \\
I_{i-1} & \text{if } i = m + 1, \ldots, g + 1.
\end{cases}
\]

The new weight

\[
s_j^* = \begin{cases} 
s_j(1 - \epsilon) & \text{if } j = 1, \ldots, m - 1, \\
\epsilon & \text{if } j = m, \\
s_m(1 - \epsilon) & \text{if } j = m + 1.
\end{cases}
\]
Furthermore,

\[ p_i^* = \begin{cases} 
  p_i(1 - \epsilon) & \text{if } i = 1, \ldots, m - 1, \\
  \epsilon & \text{if } i = m, \\
  p_{i-1}(1 - \epsilon) & \text{if } i = m + 1, \ldots, g + 1.
\end{cases} \]

The new \( C^* \) matrix is now \( m \times m \) dimensional and it can be verified that

\[ C^* = \begin{pmatrix} C / (1 - \epsilon) & 0 \\ 0 & p_{g+1}^* \end{pmatrix} \]

where \( p_i^* \) is defined in an obvious way and \( C^* \) is the matrix as before. Taking partial derivatives \( \frac{\partial}{\partial p_k} \) on both sides of equations (3.5) after replacing \( C \) by \( C^* \), etc. yields (letting \( 1 - \epsilon = q \))

\[ 0 = \sum_{i=1}^{g+1} \frac{\partial p_i^*}{\partial p_h} \frac{\delta_{ij}}{\delta_{ij}} - \sum_{i=1}^{g+1} p_i^* \frac{\delta_{ij}^* \sum_{l=1}^{m+1} (\delta_{il}^* \frac{\partial s_l^*}{\partial p_h})}{(\sum_{k=1}^{m+1} \delta_{ik}^* s_k^*)^2} \]

(where \( \delta_{ij}^* = 0 \) if \( j < m \), \( \delta_{im}^* = 0 \) if \( i \neq m \) and \( \frac{\partial s_l^*}{\partial p_i^*} = 0 \))

\[ = \sum_{k=1}^{m} \frac{\delta_{kj}}{\delta_{kj}} - \sum_{k=1}^{m} \frac{\delta_{kj}}{\delta_{kj}} - \sum_{i=1}^{g} q p_i \frac{\delta_{ij} \sum_{l=1}^{m-1} (\delta_{il} - \delta_{im}^* (m-1) \frac{\partial s_l}{\partial p_h})}{(\sum_{k=1}^{m} \delta_{ik}^* s_k^*)^2}, \]

\[ j = 1, \ldots, m - 1, \ h = 1. \]

Next we rearrange the order of \( A_j^* \)'s as follows: Let \( A_j^* = A_j \) for \( j = 1, \ldots, m \) and \( A_{m+1}^* = [b, b] \); let \( I_i^* = I_i \) for \( i = 1, \ldots, g \) and \( I_{g+1}^* = [b, b] \) and define \( p_e^* \) correspondingly. The new \( m \times m \) matrix \( C_e^* \) induced by this arrangement has the \((j,l)\) entry

\[ \lambda_{jl}^e = \sum_{i=1}^{g} p_i^e \frac{\delta_{ij}^e \delta_{il}^e}{(\sum_{k=1}^{m} \delta_{ik}^e (s_k^e))^2}, \]

since \( \delta_{ij}^e \delta_{im}^e = 0 \) for \( i = 1, 2, \ldots, g + 1 \) and \( j = 1, \ldots, m \). Write \( C_e^* = U_e D_e (U_e)^t \) in an obvious way, then it can be shown that \( U_e = (3, U_2) \), thus \( C_e^* \) is non-singular. This implies that the derivative \( \frac{\partial s_e^*}{\partial p_e^*} \) exists and is unique. Note that
\[(s_1^*, ..., s_m^*, s_{m+1}^*) = (s_1^e, ..., s_m^e, s_{m+1}^e) \text{ and} \]
\[p_i^* = \begin{cases} 
  p_i^e & \text{if } i = 1, ..., m - 1, \\
  p_i^g + 1 & \text{if } i = m, \\
  p_i^g - 1 & \text{if } i = m + 1, ..., g + 1,
\end{cases} \]

and that the system of equations (2.5) related to \(s^*\) (note that there are \(m + 1\) rather than \(m\) equations) is equivalent to that related to \(s^e\) in the following sense:

Choosing any \(m - 1\) equations of it to derive \(\frac{\partial s^*}{\partial p^*}\) would be the same.

\(C^*\) corresponds to choosing \(i = 1, ..., m\) and \(C^e\) corresponds to choosing \(i = 1, ..., m - 1\) and \(m + 1\). Hence, even though \(C^* \neq C^e\), \(\frac{\partial s^*}{\partial p^*} = (C^e)^{-1}w_i^e\) implies that \(\frac{\partial s^*}{\partial p^*} = (C^e)^{-1}w_i^e\), where \(w_i^e\) are defined in an obvious way. As a consequence of the nonsingularity of \(C^e\), \(C^*\) is nonsingular, too. In view of (A.1), \(C\) is nonsingular.

Acknowledgement. The authors thank the referees for their invaluable suggestions and opinions.

References


Qiqing Yu
Department of Mathematical Sciences
State University of New York
Binghamton, NY 13902
USA
e-mail : qyu@math.binghamton.edu

Linxiong Li
Department of Mathematics
University of New Orleans
LA 70148
USA
e-mail : lli@math.uno.edu

George Y.C. Wong
Strang Cancer Prevention Center
NY 10021
USA