東海大學統計研究所

碩士論文

Nonparametric estimators of survival function

with left-truncate current status data

指導教授:沈葆聖博士

研究生:陳茂銓

中華民國一〇二年七月

東海大學碩士班研究生

論文口試委員審定書

統計學系碩士班陳茂銓君所提之論文

Nonparametric estimators of survival function with left-truncate current status data

---

經本委員會審議，認為符合碩士資格標準。

論文口試委員召集人 戴政 （簽章）

委員 [illegible]

[illegible]

____

中華民國 102 年 7 月 2 日

Nonparametric estimators of survival function
with left-truncate current status data

Director: Pao-sheng Shen
Graduate student: Mao-cyuan Chen
Department of Statistics
TunghaiUniversity,Taichung,Taiwan 40704

謝　誌

在工作了兩年後,毅然決定考研究所,幸運地讓我考上東海統研所，在這裡認識了許多人，因為轉科考，所以在基礎上相對地薄弱許多，不過在老師耐心的指導下，以及同學的支持相挺，讓我突破了兩年研究所的困境。

首先感謝指導教授 沈葆聖老師，因為有老師的鞭策，才能很快的將這篇論文完成;一開始,沒有一個大方向,老師很有耐心的安撫我們，並且一步步的將我們帶入存活的領域，亦師亦友的領導方式，讓我們不再那麼徬徨；這兩年獲益良多，從學業、價值觀甚至是未來的方向，老師都給予很多的建議，也不忘勉勵我們，時時學習，不忘本。

感謝每一個東海統計系的教授，在每一門科目，都會不厭其煩的教導,給予真誠的關懷,鼓勵我繼續努力,老師總是對我們不了解的地方，給予最細心的解釋，極有耐心的教會我們,除此之外，更主動針對一些他認為我們可能比較不懂的地方進行講解。

感謝教授 戴政老師以及教授 林正祥老師，在百忙之中能夠擔任口試委員,並且給予我論文上的指導與建議,讓我的論文能夠更加的完整;感謝學長，撥空指導我，犧牲私人時間，協助我把口試內容逐漸完善起來，並且給予寶貴的建議。

感謝碩士班的同學們，雖然大家準備的方向不同，但是同在研究室的我們，依然給予彼此鼓勵，奮鬥到三更半夜；感謝辦公室的助教們，給予很大的協助，並且安撫著我們緊張的情緒。

最後，感謝我的家人，有他們的支持，才能讓我無後顧之憂，努力奮鬥到現在；謝謝這一路上幫助我的大家，真的很感謝你們。

學生 陳茂銓 謹識於
東海大學統計研究所
中華民國 102 年 7 月

Content

Abstract

Current status data arise from the basic version of the interval censoring where the observation consists only of an examination time and knowledge of whether the failure time has occurred before the examination time. In some cases, the failure time also suffers left-truncation, which results in left-truncated current status (LTCS) data. In this article, we first point out that for LTCS data the nonparametric estimator using Turnbull's EM algorithm is not the NPMLE since left-truncation times can also be left-censoring times. However, based on innermost sets, we can still obtain a nonparametric estimate using Turnbull's algorithm. Simulation study indicates that both estimators perform adequately and are consistent.

Key Words: current status data; left truncation; interval censoring; self-consistent.

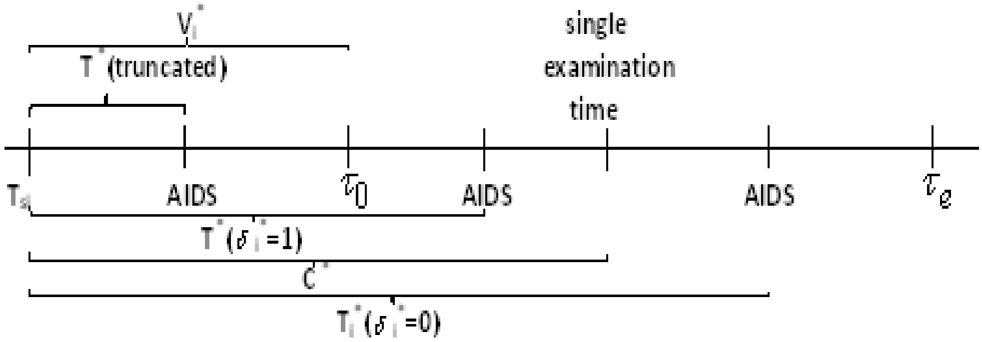

Figure 1. Schematic depiction of LTCS data

1. Introduction

Left truncated and interval-censored data often arise in epidemiology and individual follow-up studies and possibly in other fields. Current status data result from the most basic version of the interval-censoring model, known as the current status model or case 1 interval censoring. For current status data, the individual is checked only at a single point in time, denoted by C, and the status of the individual ascertained: 1 if the infection/failure time T has occurred by C and 0 otherwise. In some cases, T also suffers left-truncation. Consider the following example.

Example: AIDS Cohort Studies

In cohort studies, we are interested in the incubation time of the disease. An individual is selected only when he (or she) has already entered status 0 (e.g. HIV-positive or diagnosis of diabetes) sometime prior to calendar time τ_0 and yet has not entered status 1 (e.g. developed AIDS or retinopathy). Hence, earlier onset of AIDS/retinopathy would then be a truncating force for the variable of interest. Suppose that for each individual i the infection time (denoted by T_{si}) can be quite accurately determined (e.g. due to blood transfusion). The recruitment starts at τ_0 and the follow-up is terminated at τ_e. For each individual i, let T_i^* denote the time from T_{si} to the calendar time of entering status 1. Let $V_i^* = \tau_0 - T_{si}$ if $T_{si} < \tau_0$ and $V_i^* = 0$ if $T_{si} \geq \tau_0$. Hence, T_i^* is observable only when $T_i^* \geq V_i^*$. Let C_i^* denote the time from T_{si} to the single examination time and $\delta_i^* = 1$ if $T_i^* \leq C_i^*$ and equal to zero otherwise. Thus, for left-truncated current status (LTCS) data, we can observe (C_i^*, δ_i^*) if $T_i^* \geq V_i^*$ and observe nothing if $T_i^* < V_i^*$. Figure 1 highlights all the different times for LTCS data as described in Example.

When there is no truncation, statistical inference methods for the nonparametric maximum likelihood estimator (NPMLE) have been extensively. For example, the algorithms for obtaining the NPMLE of the distribution function of T_i^* were proposed by Ayer et al. (1955), Peto (1973), Turnbull (1976) and Groeneboom and Wellner (1992) under the assumption that T_i^* and C_i^* are independent. Furthermore, Groenboom and Wellner (1992) showed that the NPMLE converges pointwise at rate $n^{1/3}$ to a complex limiting distribution related to Brownian motion. They also studied the efficacy of smooth functionals of the estimator.

When truncation is present, Pan and Chappell (1999) showed that the NPMLE is inconsistent from LTCS data. Using log-likelihood increment as the convergence criterion, their simulation study indicated that the NPMLE can still be seriously biased when sample size is 1000. In Section 2, we first point out that the nonparametric estimator using Turnbull's EM algorithm is not the NPMLE since left-truncation times can also be left-censoring times. However, based on a generalized definition of innermost set, we can still obtain a nonparametric estimate using Turnbull's algorithm. Furthermore, based on an integral equation, we propose a self-consistent estimator (SCE) of survival function of T_i^*. In Section 3, a simulation study is conducted to investigate the performance of the two proposed estimators. Our simulation study indicated that both estimators perform adequately and are consistent.

2. The Proposed Estimators

2.1 The Nonparametric Estimator Using Turnbull's Algorithm

Let $(V_1, C_1, \delta_1), \ldots, (V_n, C_n, \delta_n)$ denote the observed LTCS data, where $P(V_i \leq C_i) = 1$. Let $F(t)$ denote the distribution function of T_i^*, and $G(x)$ and $Q(x)$ denote the distribution function of V_i^* and C_i^*, respectively. For any distribution function W denote the left and right endpoints of its support by $a_W = inf\{t : W(t) > 0\}$ and $b_W = inf\{t : W(t) = 1\}$, respectively. Throughout this article, for identifiability of T_i^*, we assume that T_i^*, V_i^* and C_i^* are all continuous, T_i^* is independent of (V_i^*, C_i^*) and

$$a_G \leq a_F \quad \text{and} \quad b_G \leq b_F \leq b_Q. \tag{2.1}$$

Based on (V_i, C_i, δ_i) $i = 1, \ldots, n$, we can define the observed interval $[L_i, R_i]$, where $[L_i, R_i] = [V_i, C_i]$ if $\delta_i = 1$ and $[L_i, R_i] = [C_i, \infty)$ if $\delta_i = 0$. Note that $[L_i, R_i] \subset [V_i, \infty]$, i.e. $V_i \leq L_i$. For arbitrarily truncated and censored data, Turnbull (1976) introduced a self-consistent algorithm to compute the NPMLE of F. Without loss of generality, suppose the observed interval $[L_i, R_i]$'s are ordered according to L_i such that $L_1 < L_2 < \cdots < L_n$. Following Turnbull (1976), Frydman (1994) and Alioum and Commenges (1996), we consider nonparametric estimation of $S_F(t) = 1 - F(t)$ using the n independent pairs $\{A_1, B_1\}, \ldots, \{A_n, B_n\}$, where $A_i = [L_i, R_i]$ and $B_i = [V_i, \infty]$. The conditional likelihood is:

$$L_c(S_F) = \prod_{i=1}^{n} \frac{P_{S_F}(A_i)}{P_{S_F}(B_i)} = \prod_{i=1}^{n} \frac{S_F(L_i-) - S_F(R_i)}{S_F(V_i-)}, \tag{2.2}$$

where $P_S(I)$ denotes the probability that is assigned to the interval I by S. We define an NPMLE as $\hat{S}_M = \text{argmax}_{S \in \mathcal{S}}\{L_c(S)\}$, where $\mathcal{S}$ denotes the class of survival functions such that $P_S(\cup_{i=1}^n B_i) = 1$ and $L_c(S)$ is defined, i.e. $P_S(B_i) > 0$ for all $i = 1, \ldots, n$. Define $\mathcal{L} = \{L_i : i = 1, \ldots, n\}$ and $\mathcal{R} = \{R_i : i = 1, \ldots, n\} \cup \{V_i : i = 1, \ldots, n\} \cup \{\infty\}$. For left-truncated and strictly interval censored data, the individual is checked more than one point, we have $V_i \neq L_i$. In this case, as pointed out by Alioum and Commenges (1999), the conditional likelihood (2.2) will be maximized when the value of $S_F(t)$ are as large as possible for $t \in \mathcal{L}$ and as small as possible for $t \in \mathcal{R}$. This can be achieved by constructing innermost sets (see Hudgens) $H_1, \ldots, H_J$ such that $H_j = [q_j, p_j]$ is to the left of $H_{j+1} = [q_{j+1}, p_{j+1}]$ for $j = 1, \ldots, J-1$, i.e. $[q_1, p_1]$,$[q_2, p_2], \ldots, [q_J, p_J]$, where $q_1 \leq p_1 < q_2 \leq p_2 < \cdots < q_J \leq p_J$. Notice that the interval $[q_j, p_j]$ can also be constructed (see Alioum and Commenges (1999)) by representing on the real line the elements of $\mathcal{L}$ and $\mathcal{R}$ by left hooks and right hooks, respectively, i.e. $q_j \in \mathcal{L}$ and $p_j \in \mathcal{R}$. By Lemma 1 of Hudgens (2005), any distribution function which increases outside $\cup_{j=1}^{J} H_j$ cannot be an NPMLE. By Lemma 2 of Hudgens (2005), for fixed value of $P_F(H_j)$, the likelihood is independent of the values of F within the region H_j. However, for LTCS data, if $R_i < \infty$, we have $V_i = L_i \in \mathcal{L}$ and $V_i \in \mathcal{R}_\rangle$, i.e. left-truncation variables belongs to both sets. Hence, for LTCS data, there may exists many intervals $[q_j, p_j]$ with $q_j = p_j = V_k$ for some k. In this case, conditional likelihood L_c would not be maximized by representing on the real line the elements of $\mathcal{L}$ and $\mathcal{R}$ by left hooks and right hooks, respectively.

Now, for each $H_j \in \mathcal{H}$, let $s_j = P_F(H_j)$ and let $\mathbf{s}$ be an J-dimension column vector with elements s_j. Define

$$L_s(\mathbf{s}) = \prod_{i=1}^{n} \frac{\sum_{j=1}^{J} \alpha_{ij} s_j}{\sum_{j=1}^{J} \beta_{ij} s_j}, \tag{2.3}$$

where $\alpha_{ij} = I[H_j \subset A_i]$, $\beta_{ij} = I[H_j \subset B_i]$ and $I[\cdot]$ is the usual indicator function. For left-truncated and strictly interval censored data, the NPMLE can be obtained by maximizing the reduced likelihood (2.3). However, for LTCS data, the estimator obtained by maximizing (2.3) is no longer the NPMLE. The goal is to maximize likelihood (2.3) subject to the constraints

$$\sum_{j=1}^{J} s_j = 1, \tag{2.4}$$

$$s_j \geq 0 \ (j = 1, \ldots, J), \tag{2.5}$$

and

$$\sum_{j=1}^{J} \alpha_{ij} s_j > 0, \ (i = 1, \ldots, n). \tag{2.6}$$

We shall use Ω to denote the parameter space that is given by constraints (2.4)-(2.6), i.e.

$$\Omega = \{\mathbf{s} \in R^J : \sum_{j=1}^{J} s_j = 1; s_j \geq 0 \text{ for } j = 1, \ldots, J; \sum_{j=1}^{J} \alpha_{ij} s_j > 0 \text{ for } i = 1, \ldots n\}.$$

To find the maximum likelihood estimate of the vector $\mathbf{s}$, we can use an EM algorithm and the resulting self-consistent estimate of $\mathbf{s}$ is exactly the Turnbull's (1976) self-consistency algorithm as follows:

$$s_j^{(b)} = \left\{1 + \frac{d_j(s^{(b-1)})}{M(s^{(b-1)})}\right\} s_j^{(b-1)} \ (1 \leq j \leq J), \tag{2.7}$$

where

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

and

$$M(s^{(b-1)}) = \sum_{i=1}^{n} \frac{1}{\sum_{j=1}^{J} \beta_{ij} s_j^{(b-1)}}.$$

Let $\hat{s}_j$ $(j = 1, \ldots, J)$ denote the estimators obtained from (2.7). Then based on the the estimators $\hat{s}_j$'s, an estimator $\hat{S}_M(t)$ of $S_F(t)$ can be uniquely defined for $t \in [p_j, q_{j+1})$ by $\hat{S}_M(p_j) = \hat{S}_M(q_{j+1}-) = 1 - (\hat{s}_1 + \cdots + \hat{s}_j)$, 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 $\hat{S}_M(t) = 1 - [\hat{s}_1 + \cdots + \hat{s}_{j-1} + s_j(t - q_j)/(p_j - q_j)]$ if $t \in (q_j, p_j]$ and $0 < q_j < p_j < \infty$. In simulation study, the convergence criterion was set as $|\hat{S}_M^{(r+1)}(t) - \hat{S}_M^{(r)}(t)| < 0.0001$. Although $\hat{S}_M$ is not the NPMLE, simulation study in Section 3 indicates that $\hat{S}_M$ performs adequately and is consistent.

2.2 The SCE

In this section, based on an integral equation, we propose an alternative estimator, the self-consistent estimator. Let $p = P(V_i^* \le T_i^*)$ denote the proportion of un-truncation. We have the following equation:

$$S_F(t) = P(T_i^* > t, V_i^* \le t) + P(T_i^* > t, V_i^* > t)$$

$$= pP(V_i^* \le t < C_i^*, \delta_i^* = 0 | T_i^* \ge V_i^*) + pP(C_i^* \le t, T_i^* > t | T_i^* \ge V_i^*)$$

$$+ pP(V_i^* \le t, \min(T_i^*, C_i^*) > t, \delta_i^* = 1 | T_i^* \ge V_i^*) + P(T_i^* > t, V_i^* > t)$$

$$= pP(V_i \le t < C_i, \delta_i = 0) + pP(C_i \le t, T_i > t)$$

$$+ pP(V_i \le t, \min(T_i, C_i) > t, \delta_i = 1) + P(T_i^* > t, V_i^* > t). \quad (2.8)$$

Motivated by (2.8), given p, we consider the following SCE:

$$\hat{S}(t) = \frac{1}{np^{-1}} \Big\{ \sum_{i=1}^n I_{[V_i \le t < C_i, \delta_i = 0]} + \sum_{i=1}^n I_{[C_i \le t, \delta_i = 0]} \frac{\hat{S}(t)}{\hat{S}(C_i)}$$

$$+ \sum_{i=1}^n I_{[V_i \le t < C_i, \delta_i = 1]} \frac{\hat{S}(t) - \hat{S}(C_i)}{\hat{S}(V_i) - \hat{S}(C_i)} + \sum_{i=1}^n I_{[V_i > t]} \frac{\hat{S}(t)}{\hat{S}(V_i)} \Big\}. \quad (2.9)$$

Notice that the last term of the equation (2.9) is to recover the missing information due to left-truncation. Given the observation $V_i > t$, a pseudo observation is recovered by adding the weight $\hat{S}(t)/\hat{S}(V_i)$. Let $\tilde{G}(t) = P(V_i \le t)$ denote the sub-distribution function of V_i. Since $\tilde{G}(t) = p^{-1} \int_0^t S_F(v) dG(v)$. It follows that np^{-1} can be estimated by $\sum_{i=1}^n 1/S_F(V_i)$ (see Shen (2005)). Hence, an SCE $\hat{S}_n$ of S_F is defined to be the solution of the following equation:

$$\hat{S}_n(t) = \Big[\sum_{i=1}^n \frac{1}{\hat{S}_n(V_i)} \Big]^{-1} \Big\{ \sum_{i=1}^n I_{[V_i \le t < C_i, \delta_i = 0]} + \sum_{i=1}^n I_{[C_i \le t, \delta_i = 0]} \frac{\hat{S}_n(t)}{\hat{S}_n(C_i)}$$

$$+ \sum_{i=1}^n I_{[V_i \le t < C_i, \delta_i = 1]} \frac{\hat{S}_n(t) - \hat{S}_n(C_i)}{\hat{S}_n(V_i) - \hat{S}_n(C_i)} + \sum_{i=1}^n I_{[V_i > t]} \frac{\hat{S}_n(t)}{\hat{S}_n(V_i)} \Big\} \text{ and } \hat{S}_n \in \Theta, \quad (2.10)$$

where $\Theta = \{f : f$ is a nonincreasing function from $[0, \infty]$ *to* $[0, 1], f(0) = 1$ and $f(\infty) = 0\}$. Let $\tilde{G}_n(v)$ denote the empirical version of $\tilde{G}(v)$. Similarly, Let $\tilde{Q}_{0n}(c)$, $\tilde{H}_{0n}(v, c)$ and $\tilde{H}_{1n}(v, c)$ denote the empirical versions of the joint sub-distributions of $\tilde{Q}_0(c) = P(C_i \le c, \delta_i = 0)$,

$\tilde{H}_0(v,c) = P(V_i \le v, C_i \le c, \delta_i = 0)$ and $\tilde{H}_1(v,c) = P(V_i \le v, C_i \le c, \delta_i = 1)$, respectively. It follows that (2.10) can be written as

$$\hat{S}_n(t) = \left[\int \frac{1}{\hat{S}_n(v)}\tilde{G}_n(dv)\right]^{-1}\left\{\int_{v\le t<c}\tilde{H}_{0n}(dv,dc) + \int_{c\le t}\frac{\hat{S}_n(t)}{\hat{S}_n(c)}\tilde{Q}_{0n}(dc)\right.$$

$$\left. + \int_{c>t}\frac{\hat{S}_n(t)-\hat{S}_n(c)}{\hat{S}_n(v)-\hat{S}_n(c)}\tilde{H}_{1n}(dv,dc) + \int_{v>t}\frac{\hat{S}_n(t)}{\hat{S}_n(v)}\tilde{G}_n(dv)\right\}. \tag{2.11}$$

When there is no truncation, (2.11) is reduced to:

$$\hat{S}_n(t) = \left\{\int_{t<c}\tilde{Q}_{0n}(dc) + \int_{c\le t}\frac{\hat{S}_n(t)}{\hat{S}_n(c)}\tilde{Q}_{0n}(dc) + \int_{v\le t<c}\frac{\hat{S}_n(t)-\hat{S}_n(c)}{1-\hat{S}_n(c)}\tilde{Q}_{1n}(dc),\right.$$

where $\tilde{Q}_{1n}$ is the empirical function of $\tilde{Q}_1(c) = P(C_i \le c, \delta_i = 0)$.

東海大學 TUNGHAI UNIVERSITY 1955

3. Simulation Results

Case 1: $C_i^* = V_i^* + 0.5$

A simulation study is conducted to investigate the performance of the estimator $\hat{S}_M$ and SCE. The simulation set-up is the same as in Pan and Chappell (1999). The left-truncation time $V_i^* \sim U(0,\theta)$ is uniformly distributed and the censoring time $C_i^* = V_i^* + 0.5$. The values of θ were set at $\theta = 4, 8$ such that proportions of left-truncation are equal to 0.53 and 0.76, respectively. The survival time T_i^* is distributed as Gamma with shape and scale parameters 2 and 1, respectively. We consider the estimation of $S_F(t_P)$ where t_P is the 100 P^{th} percentile point. Turnbull's EM algorithm is used to compute $\hat{S}_M(t_P)$ with a starting distribution which puts an equal probability mass in each s_j $(j = 1, \ldots, J)$. The convergence criterion was set as $|\hat{S}_M^{(r+1)}(t_P) - \hat{S}_M^{(r)}(t_P)| < 0.0001$. Notice that the convergence criterion for the $\hat{S}_M$ differs from that used by Pan and Chappell (1999). They use the log-likelihood increment as the criterion. To obtain an initial estimator of $\hat{S}_n$, the exponential distribution with mean equal to 2, i.e. $\hat{S}_n^{(0)} = e^{-x/2}$, was used as an initial estimator. The convergence criterion was set as $|\hat{S}_n^{(r+1)}(t_P) - \hat{S}_n^{(r)}(t_P)| < 0.0001$. The values of P are chosen as 0.2, 0.5 and 0.8 and the sample sizes are chosen as 100, 200 and 1000. The replication is 1000 times. The simulation results were reported in Table 1. Table 1 also lists proportion of truncation $P(T_i^* < V_i^*)$ (denoted by q_T) and proportion of left censoring $P(\delta_i = 1)$ (denoted by p_L).

Case 2: $C_i^* = V_i^* + D_i^*$

The distribution of T_i^* is the same as case 1. The left-truncation time V_i^* is exponentially distributed with mean θ and the censoring time $C_i^* = V_i^* + D_i^*$, where D_i^* is exponentially distributed with mean equal to 2. The values of θ were set at $\theta = 2, 4$ such that proportions of left-truncation are equal to 0.43 and 0.65, respectively. Simulation results are reported in Table 2.

Table 1. Simulation results for bias, standard deviation and rmse for left-truncated and current status data (case 1)

					$\hat{S}_n(t_P)$			$\hat{S}_M(t_P)$		
θ	n	q_T	p_L	P	bias	std	rmse	bias	std	rmse
4	100	0.53	0.24	0.2	-0.024	0.087	0.079	-0.002	0.080	0.080
4	200	0.53	0.24	0.2	-0.012	0.056	0.057	-0.001	0.050	0.050
4	1000	0.53	0.24	0.2	-0.008	0.024	0.025	0.005	0.024	0.024
8	100	0.76	0.23	0.2	-0.029	0.071	0.077	0.014	0.071	0.072
8	200	0.76	0.23	0.2	-0.016	0.051	0.053	0.006	0.045	0.045
8	1000	0.76	0.23	0.2	-0.007	0.022	0.022	0.005	0.020	0.020
4	100	0.53	0.24	0.5	-0.035	0.103	0.109	-0.004	0.095	0.095
4	200	0.53	0.24	0.5	-0.017	0.062	0.064	-0.002	0.057	0.057
4	1000	0.53	0.24	0.5	-0.006	0.031	0.031	0.003	0.028	0.028
8	100	0.76	0.23	0.5	-0.029	0.088	0.092	0.004	0.095	0.095
8	200	0.76	0.23	0.5	-0.016	0.061	0.063	0.005	0.059	0.059
8	1000	0.76	0.23	0.5	-0.011	0.025	0.027	0.006	0.023	0.024
4	100	0.53	0.24	0.8	-0.031	0.105	0.109	-0.027	0.098	0.102
4	200	0.53	0.24	0.8	-0.022	0.076	0.079	-0.015	0.071	0.073
4	1000	0.53	0.24	0.8	-0.010	0.027	0.029	0.008	0.028	0.029
8	100	0.76	0.23	0.8	-0.028	0.107	0.111	-0.024	0.101	0.104
8	200	0.76	0.23	0.8	-0.017	0.082	0.084	-0.013	0.076	0.077
8	1000	0.76	0.23	0.8	0.008	0.032	0.033	0.007	0.030	0.031

Table 2. Simulation results for bias, standard deviation and rmse for left-truncated and current status data (case 2)

					$\hat{S}_n(t_P)$			$\hat{S}_M(t_P)$		
θ	n	q_T	p_L	P	bias	std	rmse	bias	std	rmse
2	100	0.43	0.49	0.2	-0.010	0.084	0.085	-0.016	0.088	0.089
2	200	0.43	0.49	0.2	-0.020	0.053	0.057	-0.012	0.051	0.052
2	1000	0.43	0.49	0.2	-0.012	0.024	0.027	-0.008	0.021	0.022
4	100	0.65	0.55	0.2	-0.017	0.074	0.076	-0.015	0.080	0.081
4	200	0.65	0.55	0.2	-0.032	0.049	0.059	-0.014	0.050	0.052
4	1000	0.65	0.55	0.2	-0.018	0.020	0.027	-0.016	0.019	0.025
2	100	0.43	0.49	0.5	-0.020	0.135	0.136	-0.027	0.128	0.131
2	200	0.43	0.49	0.5	-0.016	0.093	0.094	-0.019	0.089	0.091
2	1000	0.43	0.49	0.5	-0.013	0.048	0.050	-0.011	0.046	0.047
4	100	0.65	0.55	0.5	-0.035	0.132	0.136	-0.030	0.127	0.130
4	200	0.65	0.55	0.5	-0.029	0.086	0.091	-0.023	0.083	0.086
4	1000	0.65	0.55	0.5	-0.012	0.045	0.047	-0.015	0.044	0.046
2	100	0.43	0.49	0.8	-0.037	0.135	0.140	-0.032	0.131	0.135
2	200	0.43	0.49	0.8	-0.028	0.097	0.104	-0.022	0.092	0.095
2	1000	0.43	0.49	0.8	-0.014	0.043	0.045	-0.012	0.044	0.045
4	100	0.65	0.55	0.8	-0.042	0.140	0.146	-0.036	0.137	0.142
4	200	0.65	0.55	0.8	-0.031	0.102	0.107	-0.028	0.095	0.099
4	1000	0.65	0.55	0.8	-0.017	0.058	0.060	-0.014	0.056	0.058

Tables 1 and 2 indicate that (i) For case 1, when $n = 100, 200$, the biases of the estimator $\hat{S}_M$ are smaller than that of $\hat{S}_n$. For case 2, when $n = 100$, the biases of both estimators can be large. (ii) When $n = 100, 200$, in terms of rmse, the estimator $\hat{S}_M$ performs better than the SCE $\hat{S}_n$ for most of the cases considered. (iii) When $n = 1000$, the performance of the estimators $\hat{S}_n$ and $\hat{S}_M$ are close to each other.

4. Discussions

For left truncated and current status data, we have pointed out that the nonparametric estimator using Turnbull's EM algorithm is not the NPMLE since left-truncation times can also be left-censoring times. However, based on innermost sets, we can still obtain a nonparametric estimate $\hat{S}_m$ using Turnbull's algorithm and simulation results indicates that the estimator performs adequately. Furthermore, we have presented a SCE using an integral equation and simulation study indicates that the SCE performs adequately. Further research is required to establish the consistency of the SCE.

東海大學 TUNGHAI UNIVERSITY 1955

References

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

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T. and Silverman, E. (1955). An empirical distribution function for sampling with incomplete observations. *Annals of Mathematical Statistics*, **26**, 641-7.

Frydman, H. (1994). A note on nonparametric estimation of the distribution function from interval-censored and truncated data. *Journal of the Royal Statistical Society, Series B*, **56**, 71-74.

Gentleman, R. and Geyer C. J. (1994). Maximum likelihood for interval censored data: consistency and computation. *Biometrika*, **81**, 618-623.

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

Gu, M. G. and Zhang, C. H. (1993), Asymptotic properties of self-consistent estimators based on doubly censored data. *The Annals of Statistics* **21**, 611-624.

Hudgens, M. G. (2005). On nonparametric maximum likelihood estimation with interval censoring and truncation. *Journal of the Royal Statistical Society, Series B*, **67**, part 4, 573-587.

Pan, W. and Chappell, R (1999). A note on inconsistency of NPMLE of the distribution function from left truncated and case I interval censored data. *Lifetime Data Analysis*, **5**, 281-291.

Peto, R. (1973). Experimental survival curves for interval-censored data. *Applied Statistics* **22**, 86-91.

Shen, P.-S. (2005). Estimation of the truncation probability with the left-truncated and right-censored data. *Nonparametric Statistics*, **17**, No. 8, 957-969.

Schick, A. and Yu, Q. (2000). Consistency of the GMLE with Mixed Case Interval-Censored Data. *Scandinavian Journal of Statistics*, **27**, 45-55.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped censored and truncated data. *Journal of the Royal Statistical Society, Series B*, **38**, 290-295.