Statistica Sinica 6(1996), 509-530

BOOTSTRAP METHODS FOR TRUNCATED AND CENSORED DATA

Shulamith T. Gross and Tze Leung Lai

City University of New York and Stanford University

Abstract: For right censored data, Efron (1981) has shown that his "simple" and "obvious" methods of bootstrapping are equivalent. We explain why this equivalence no longer holds for truncated data. Wang (1991) generalized Efron's "obvious" bootstrap method to data that are both left truncated and right censored, under the assumption that $C \geq T$ and $C - T$ is independent of T, where T and C denote the (generic) censoring and truncation variables. We discuss how the "obvious" bootstrap method can be extended when this independence assumption is removed, and also develop an asymptotic theory of the "simple" bootstrap method for left truncated and right censored data, showing that the "simple" bootstrap approximations to the sampling distributions of various nonparametric statistics from these data are accurate to the order of $O_p(n^{-1})$.

Key words and phrases: Left truncation, right censoring, estimable functionals, Edgeworth expansions, bootstrap, asymptotic U-statistics.

1. Introduction

Let (Y_1, T_1, C_1), $(Y_2, T_2, C_2), \ldots$ be i.i.d. random vectors such that (T_i, C_i) is independent of Y_i. Let F denote the common distribution function of the random variables Y_i, and H denote the common bivariate distribution of the random vectors (T_i, C_i). Here F and its functionals are quantities of interest, but they cannot be estimated directly from the Y_i which are not completely observable because of the presence of the right censoring and left truncation variables C_i and T_i. Letting $\tilde{Y}_i = \min(Y_i, C_i)$ and $\delta_i = I(Y_i \leq C_i)$, one only observes $(\tilde{Y}_i, \delta_i)$ when $\tilde{Y}_i \geq T_i$. Thus, the data consist of n observations $(\tilde{Y}_{i,o}, \delta_{i,o}, T_{i,o})$ with $\tilde{Y}_{i,o} \geq T_{i,o}$, $i = 1, \ldots, n$. Such left truncated and right censored (l.t.r.c.) data have wide applications in biostatistics (cf. Hyde (1977), Tsai, Jewell and Wang (1987), Keiding, Holst and Green (1989), Kalbfleisch and Lawless (1989), Wang (1991), and Andersen et al. (1993)). When censoring is absent, multiplying the random variables by -1 converts a left truncated model into a right truncated model, and right truncated data have extensive applications in astronomy and econometrics (cf. Lynden-Bell (1971), Nicoll and Segal (1980), Bhattacharya, Chernoff and Yang (1983), Efron and Petrosian (1994), Tobin (1958), Goldberger (1981) and Amemiya (1985)). Throughout the sequel we let

$$\tau = \inf\{s : P(T_1 \leq s \leq C_1) > 0\},\ \tau^* = \inf\{s > \tau : P(T_1 \leq s \leq C_1) = 0\}. \quad (1.1)$$

Bootstrap methods for right censored data without left truncation (i.e., $T_i \equiv -\infty$) were introduced by Efron (1981) and Reid (1981) and subsequently studied by Akritas (1986), Lo and Singh (1986), Horváth and Yandell (1987), Babu (1991), Lai and Wang (1993), among others. Let Q denote the common distribution function of the C_i. Given the observed data $(\tilde{Y}_1, \delta_1), \ldots, (\tilde{Y}_n, \delta_n)$, Efron's "obvious" bootstrap starts by estimating the unknown F and Q from these data by the Kaplan-Meier estimates $\hat{F}$ and $\hat{Q}$, and generates independent random variables $Y_1^*, \ldots, Y_n^*$, $C_1^*, \ldots, C_n^*$ from the distributions $\hat{F}$ (for the Y_i^*) and $\hat{Q}$ (for the C_i^*) to form the bootstrap sample $\tilde{Y}_i^* = \min(Y_i^*, C_i^*)$, $\delta_i^* = I(Y_i^* \le C_i)$, $i = 1, \ldots, n$. His "simple" bootstrap simply draws independent random vectors $(\tilde{Y}_i^*, \delta_i^*)$, $i = 1, \ldots, n$, from the empirical distribution that puts weight $1/n$ at each of the observations $(\tilde{Y}_1, \delta_1), \ldots, (\tilde{Y}_n, \delta_n)$ to form the bootstrap sample. He showed that the "obvious" and "simple" bootstrap methods are actually equivalent because of certain properties of the Kaplan-Meier estimates $\hat{F}$ and $\hat{Q}$, as will be discussed further in Section 3.

Wang (1991) extended Efron's obvious bootstrap method from censored data to l.t.r.c. data under the following assumptions on the censoring and truncation variables:

$$C_i \ge T_i \text{ and } C_i - T_i \text{ is independent of } T_i, \tag{1.2}$$

$$F \text{ is continuous, } \tau < \inf\{s : F(s) > 0\} \text{ and } G^{-1}(1) < F^{-1}(1) < \tau^*, \tag{1.3}$$

where G denotes the common distribution function of the truncation variables T_i, τ and τ^* are defined in (1.1) and $F^{-1}(1) = \sup\{t : F(t) < 1\}$. Under (1.3), $S(t) := P\{Y_1 \ge t\}$ can be consistently estimated by

$$\hat{S}(t) = \prod_{s<t} \{1 - \Delta N_n(s)/R_n(s)\}, \tag{1.4}$$

where

$$R_n(s) = \sum_{i=1}^{n} I(T_{i,o} \le s \le \tilde{Y}_{i,o}), \;\; N_n(s) = \sum_{i=1}^{n} I(\tilde{Y}_{i,o} \le s, \delta_{i,o} = 1), \tag{1.5}$$

$\Delta N_n(s) = N_n(s) - N_n(s-)$ and we use the convention $0/0 = 0$. Moreover, $G(t)$ can be consistently estimated by

$$\hat{G}(t) = \Big\{ \sum_{i \le n : T_{i,o} \le t} (\hat{S}(T_{i,o}))^{-1} \Big\} / \Big\{ \sum_{i=1}^{n} (\hat{S}(T_{i,o}))^{-1} \Big\}. \tag{1.6}$$

Let $\hat{F}(t) = 1 - \hat{S}(t+)$. Because of (1.2), we need only estimate, besides F and G, the common distribution function Q of $C_i - T_i$, which is estimated by the product-limit estimator

$$\begin{aligned}\hat{Q}(t) &= 1, \quad \text{if} \quad t \geq \max_{i\leq n}(\tilde{Y}_{i,o} - T_{i,o}),\\ &= 1 - \prod_{0\leq s\leq t}\{1 - \Delta N_n'(s)/R_n'(s)\}, \quad \text{if} \quad 0 \leq t < \max_{i\leq n}(\tilde{Y}_{i,o} - T_{i,o}), \end{aligned} \tag{1.7}$$

where $N_n'(s) = \sum_{i=1}^n I(\delta_{i,o} = 0, \tilde{Y}_{i,o} - T_{i,o} \leq s)$, $R_n'(s) = \sum_{i=1}^n I(\tilde{Y}_{i,o} - T_{i,o} \geq s)$. Wang's (1991) bootstrap method generates independent random variables Y_j^*, T_j^*, D_j^* such that Y_j^* has distribution $\hat{F}$, T_j^* has distribution $\hat{G}$ and D_j^* has distribution $\hat{Q}$. Let $C_j^* = T_j^* + D_j^*$, $\delta_j^* = I(Y_j^* \leq C_j^*)$, $\tilde{Y}_j^* = \min(Y_j^*, C_j^*)$. Retain $(\tilde{Y}_j^*, \delta_j^*, T_j^*)$ in the bootstrap sample if and only if $\tilde{Y}_j^* \geq T_j^*$, so that the bootstrap sample consists of n random vectors $(\tilde{Y}_{1,o}^*, \delta_{1,o}^*, T_{1,o}^*), \ldots, (\tilde{Y}_{n,o}^*, \delta_{n,o}^*, T_{n,o}^*)$, with $\tilde{Y}_{i,o}^* \geq T_{i,o}^*$, generalizing Efron's "obvious" bootstrap method to l.t.r.c. data.

On page 140 of Wang (1991), it is mentioned that for l.t.r.c. data this obvious bootstrap method is no longer equivalent to the simple bootstrap method that samples n times with replacement from $\{(\tilde{Y}_{1,o}, \delta_{1,o}, T_{1,o}), \ldots, (\tilde{Y}_{n,o}, \delta_{n,o}, T_{n,o})\}$, and there is also the following comment concerning these two different bootstrap methods: "Some preliminary theoretical results as well as a large simulation study $\cdots$ have shown the validity of the obvious method for left-truncated and right-censored data. The appropriateness of the simple method, however, still remains unclear". In Section 2 we give an asymptotic justification of the simple bootstrap method in general l.t.r.c. models that assume neither (1.2) nor (1.3). This asymptotic theory also generalizes the recent work of Lai and Wang (1993) on censored data to l.t.r.c. data, showing that the simple bootstrap method provides an empirical Edgeworth expansion, with an $O_p(n^{-1})$ error, of the sampling distribution of a nonparametric estimate of an estimable functional of (F, H). Some numerical examples are presented in Section 2 to illustrate the applications of this theory.

Note that while the obvious bootstrap starts by estimating the distribution F of the Y_i and the bivariate distribution H of the (T_i, C_i), the simple bootstrap does not involve estimation of (F, H). In Section 3 we study the problem of estimating F and H when the independence assumption (1.2) fails and develop an alternative consistent estimator $\hat{H}$ of H without assuming (1.2). Although using $\hat{H}$ in the obvious bootstrap method enables us to remove the stringent independence assumption (1.2), $\hat{H}$ requires certain smoothness assumptions on H which may still be too restrictive in practice. The discussion in Section 3 shows that for l.t.r.c. data the simple bootstrap method, whose theoretical justification has been provided in Section 2, has important practical advantages over the obvious bootstrap method.

2. Theory and Applications of the Simple Bootstrap Method for L.T.R.C. Data

In this section we extend the l.t.r.c. model considered in Section 1 to include covariates and develop an asymptotic theory of the simple bootstrap method for this extended model. Let $(\mathbf{X}_1, Y_1)$, $(\mathbf{X}_2, Y_2), \ldots$ be i.i.d. random vectors which are not completely observable due to right censoring of the Y_i by C_i and left truncation by T_i, where the (T_i, C_i) are i.i.d. random vectors that are independent of $(\mathbf{X}_i, Y_i)$. The observations, which are available only when $\tilde{Y}_i := Y_i \wedge C_i \geq T_i$, consist of $(\mathbf{X}_{i,o}, \tilde{Y}_{i,o}, \delta_{i,o}, T_{i,o})$, with $\tilde{Y}_{i,o} \geq T_{i,o}$ and $\delta_{i,o} = I(Y_{i,o} \leq C_{i,o})$, $i = 1, \ldots, n$. Here and in the sequel we use $Y \wedge C$ to denote $\min(Y, C)$. A crucial fact underlying the theory of the simple bootstrap method for these data is the following.

Lemma 1. *The $(\mathbf{X}_{i,o}, \tilde{Y}_{i,o}, \delta_{i,o}, T_{i,o})$ are i.i.d. random vectors whose common distribution is given by*

$$\begin{aligned} & P\{\delta_{i,o} = \delta, (\mathbf{X}_{i,o}, \tilde{Y}_{i,o}, T_{i,o}) \in A\} \\ = \;& P\{I(Y_1 \leq C_1) = \delta, (\mathbf{X}_1, Y_1 \wedge C_1, T_1) \in A\} / P\{Y_1 \wedge C_1 \geq T_1\}, \end{aligned} \tag{2.1}$$

for $\delta = 0$ or 1 and all Borel sets A such that $y \geq t$ if $(\mathbf{x}, y, t) \in A$.

Proof. As in Lai and Ying (1991), we can regard $(\mathbf{X}_{i,o}, \tilde{Y}_{i,o}, \delta_{i,o}, T_{i,o})$ as $(\mathbf{X}_{\sigma(i)}, \tilde{Y}_{\sigma(i)}, \delta_{\sigma(i)}, T_{\sigma(i)})$, where

$$\sigma(i) = \inf\Big\{m \geq 1 : \sum_{j=1}^{m} I(Y_j \wedge C_j \geq T_j) = i\Big\}. \tag{2.2}$$

Note that $\sigma(1)$, $\sigma(2) - \sigma(1)$, $\sigma(3) - \sigma(2), \cdots$ are i.i.d. geometric random variables with $P\{\sigma(1) = n\} = p(1-p)^{n-1}$, where $p = P\{Y_1 \wedge C_1 \geq T_1\}$. Hence the $(\mathbf{X}_{i,o}, \tilde{Y}_{i,o}, \delta_{i,o}, T_{i,o})$ are i.i.d. random vectors and (2.1) follows from Theorem II.3.2 of Devroye (1986) on the rejection method of random variate generation.

Remark. In certain contexts, such as in astronomy (cf. Woodroofe (1985)) or in retrospective studies of disease (cf. Kalbfleisch and Lawless (1989)), it would be reasonable to suppose that the sample size n is also random with $n = \sum_{j=1}^{N} I(Y_j \wedge C_j \geq T_j)$, where N is nonrandom but unobservable, so that there are N i.i.d. random vectors $(\mathbf{X}_1, Y_1, C_1, T_1), \ldots, (\mathbf{X}_N, Y_N, C_N, T_N)$ from which one observes only those quadruples with $\tilde{Y}_i \geq T_i$, yielding the observations $(\mathbf{X}_{i,o}, \tilde{Y}_{i,o}, \delta_{i,o}, T_{i,o}) = (\mathbf{X}_{\sigma(i)}, \tilde{Y}_{\sigma(i)}, \delta_{\sigma(i)}, T_{\sigma(i)})$. In this case, the conditional distribution of $\{(\mathbf{X}_{i,o}, \tilde{Y}_{i,o}, \delta_{i,o}, T_{i,o}),\ 1 \leq i \leq n\}$ given n is the same as that as in Lemma 1. Thus all the results on the simple bootstrap method given below still hold in this case under the conditional probability measure $P(\cdot|n)$.

2.1. Asymptotic U-statistics of l.t.r.c. data and Edgeworth expansions

To develop an asymptotic theory of the simple bootstrap method for l.t.r.c. data, we make use of Lemma 1 together with the results of Lai and Wang (1993) on asymptotic U-statistics and their Edgeworth expansions. Let $\mathbf{Z}$, $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ be i.i.d. p-dimensional random vectors. A real-valued function $U_n = U_n(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$ of $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ is called an *asymptotic U-statistic* if it has the decomposition

$$U_n = \sum_{i=1}^{n}\Big\{\frac{\alpha(\mathbf{Z}_i)}{\sqrt{n}} + \frac{\alpha'(\mathbf{Z}_i)}{n^{3/2}}\Big\} + \sum_{1\le i<j\le n}\frac{\beta(\mathbf{Z}_i,\mathbf{Z}_j)}{n^{3/2}} + \sum_{1\le i<j<k\le n}\frac{\gamma(\mathbf{Z}_i,\mathbf{Z}_j,\mathbf{Z}_k)}{n^{5/2}} + \Delta_n, \quad (2.3)$$

wherre $\alpha, \alpha', \beta, \gamma$ are nonrandom Borel functions which are invariant under permutation of the arguments and

(A1) $P\{|\Delta_n| \ge n^{-1-\epsilon}\} = o(n^{-1})$ for some $\epsilon > 0$,
(A2) $E\alpha(\mathbf{Z}) = E\alpha'(\mathbf{Z}) = 0$,
(A3) $E\{\beta(\mathbf{Z}_1, \mathbf{Z}_2)|\mathbf{Z}_1\} = 0$, $E\{\gamma(\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3)|\mathbf{Z}_1, \mathbf{Z}_2\} = 0$,
(A4) $E\{|\alpha'(\mathbf{Z}_1)|^3 + |\gamma(\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3)|^4\} < \infty$.

For $t \ge a$, the following example represents the product-limit estimator $\hat{F}_a(t)$ of

$$F_a(t) := P\{Y_1 \le t | Y_1 \ge a\} = F(t)/(1 - F(a-)) \quad (2.4)$$

as an asymptotic U-statistic, which we then use to derive an Edgeworth expansion for $\hat{F}_a(t)$.

Example 1. Defining $N_n(s)$ and $R_n(s)$ by (1.5), the product-limit estimator of $F_a(t)$ is

$$\hat{F}_a(t) := 1 - \prod_{a\le s\le t}\{1 - \Delta N_n(s)/R_n(s)\} \quad (2.5)$$

for $t \ge a$. By (4.17) and (4.18) of Lai and Ying (1991), for $t \ge a$,

$$\frac{\hat{F}_a(t) - F_a(t)}{1 - F_a(t)} = \sum_{i=1}^{n}\int_a^t \frac{1 - \hat{F}_a(s-)}{1 - F_a(s)} \cdot \frac{I(R_n(s) > 0)}{R_n(s)}\, dM_i(s), \quad (2.6)$$

where letting $\Lambda(t) = \int_{-\infty}^t dF(s)/(1 - F(s-))$, we define

$$\begin{aligned} M_i(t) &= I(\tilde{Y}_{i,o} \le t, \delta_{i,o} = 1) - \int_{-\infty}^t I(T_{i,o} \le s \le \tilde{Y}_{i,o}) d\Lambda(s), \\ p(s) &= P\{T_{1,o} \le s \le \tilde{Y}_{1,o}\}, \quad w_i(s) = I(T_{i,o} \le s \le \tilde{Y}_{i,o}) - p(s). \end{aligned} \quad (2.7)$$

To simplify matters, we assume that F is continuous so that $1 - F_a(s) = 1 - F_a(s-)$. We also assume that $\tau < a < t < (\tau^* \wedge F^{-1}(1))$, where τ and τ^* are defined by (1.1). Therefore, $\inf_{a\le s\le t} p(s) > 0$. By Lemma 5 of Lai and Ying

(1991), $\{I(T_1 \le Y_1 \le t \wedge C_1) - \int_{-\infty}^t I(T_1 \le s \le Y_1 \wedge C_1) d\Lambda(s),\ -\infty < t < \infty\}$ is a continuous-parameter martingale and therefore it follows from Lemma 1 that the i.i.d. processes $\{M_i(t), -\infty < t < \infty\}$ are martingales.

Writing $(1-\hat{F}_a(s-))/(1-F_a(s-)) = 1-\{\hat{F}_a(s-)-F_a(s-)\}/\{1-F_a(s-)\}$ and $R_n(s) = \sum_{j=1}^n \{p(s) + w_j(s)\}$, and expanding $(p(s)+x)^{-1}$ by Taylor's theorem, we obtain from (2.6)

$$\begin{aligned}
&\sqrt{n}\frac{\hat{F}_a(t) - F_a(t)}{1 - F_a(t)} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_a^t \Big\{\frac{1}{p(s)} - \frac{n^{-1}\sum_1^n w_j(s)}{p^2(s)} + \frac{(n^{-1}\sum_1^n w_j(s))^2}{p^3(s)}\Big\} dM_i(s) \\
&\quad -\frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{j=1}^n \int_a^t \int_a^{s-} \Big\{\frac{1}{p(s)} - \frac{n^{-1}\sum_1^n w_k(s)}{p^2(s)}\Big\} \frac{1}{np(u)} dM_j(u) dM_i(s) \qquad (2.8) \\
&\quad +\frac{1}{n^{5/2}} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \int_a^t \int_a^{s-} \int_a^{u-} \frac{1}{p(s)p(u)p(v)} dM_k(v) dM_j(u) dM_i(s) + \tilde{\Delta}_n,
\end{aligned}$$

where the remainder term $\tilde{\Delta}_n$ can be shown to satisfy $P\{|\tilde{\Delta}_n| \ge n^{-1-\epsilon}\} = o(n^{-1})$ for $0 < \epsilon < 1/2$ by using exponential bounds for the empirical process $\sum_{j=1}^n w_j(s)$ and exponential inequalities for continuous-parameter martingales (cf. Shorack and Wellner (1986), page 899). Letting $\mathbf{Z}_i = (\tilde{Y}_{i,o}, \delta_{i,o}, T_{i,o})$ and making use of the identity $\sum_{j=1}^n w_j^2(s) = np(s)(1-p(s)) + (1-2p(s))\sum_{j=1}^n w_j(s)$, it follows from (2.8) that $U_n = \sqrt{n}(\hat{F}_a(t) - F_a(t))/(1-F_a(t))$ has the decomposition (2.3) with $P\{|\Delta_n| \ge n^{-1-\epsilon}\} = o(n^{-1})$ for $0 < \epsilon < 1/2$ and with

$$\begin{aligned}
\alpha(\mathbf{Z}_i) &= \int_a^t (p(s))^{-1} dM_i(s), \\
\alpha'(\mathbf{Z}_i) &= -\int_a^t \frac{w_i(s)}{p^2(s)} dM_i(s) + \int_a^t \frac{1-p(s)}{p^2(s)} dM_i(s) \\
&\quad - \int_a^t \int_a^{s-} \frac{1}{p(s)p(u)} dM_i(u) dM_i(s), \\
\beta(\mathbf{Z}_i, \mathbf{Z}_j) &= -\int_a^t \frac{w_i(s) dM_j(s) + w_j(s) dM_i(s)}{p^2(s)} \\
&\quad - \int_a^t \int_a^{s-} \frac{dM_i(u) dM_j(s) + dM_j(u) dM_i(s)}{p(s)p(u)}, \\
\gamma(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k) &= \sum_\pi \Big\{ \int_a^t \frac{w_{\pi(1)}(s) w_{\pi(2)}(s)}{p^3(s)} dM_{\pi(3)}(s) \\
&\quad + \int_a^t \int_a^{s-} \frac{w_{\pi(1)}(s)}{p^2(u)p(s)} dM_{\pi(2)}(u) dM_{\pi(3)}(s)
\end{aligned}$$

$$+\int_a^t \int_a^{s-} \int_a^{u-} \frac{1}{p(v)p(u)p(s)} dM_{\pi(1)}(v) dM_{\pi(2)}(u) dM_{\pi(3)}(s)\Big\},$$

where $\sum_\pi$ denotes summation over all six permutations of $\{i, j, k\}$.

The representation of $U_n = \sqrt{n}(\hat{F}_a(t) - F_a(t))/(1 - F_a(t))$ as an asymptotic U-statistic in Example 1 shows that U_n has a limiting normal distribution with mean 0 and variance

$$\sigma^2 := \mathrm{Var}\,(\alpha(\mathbf{Z}_1)) = P\{\tilde{Y}_1 \geq T_1\} \int_a^t \frac{1}{P\{T_1 \leq s \leq C_1\}} \frac{dF(s)}{(1 - F(s))^2}. \tag{2.9}$$

We next use it in conjunction with Theorem 1 of Lai and Wang (1993) to derive an Edgeworth expansion of the distribution of U_n under Cramér's condition

$$\limsup_{|t|\to\infty} |Ee^{it\alpha(\mathbf{Z}_1)}| < 1. \tag{2.10}$$

Note in this connection that the second equality in (2.9) follows from (4.12) of Lai and Ying (1991) and Lemma 1, recalling that F is assumed to be continuous. Moreover, from the definition of $\alpha(\mathbf{Z}_i)$, it follows that Cramér's condition (2.10) holds if F has a non-vanishing absolutely continuous component with respect to Lebesgue measure.

Letting ϕ and Φ denote the density and distribution functions of the standard normal distribution, Theorem 1 of Lai and Wang (1993) gives an Edgeworth expansion of the form

$$P\{U_n/\sigma \leq z\} = \Phi(z) - n^{-1/2}\phi(z)P_1(z) - n^{-1}\phi(z)P_2(z) + o(n^{-1}), \tag{2.11}$$

uniformly in $-\infty < z < \infty$, for asymptotic U-statistics (2.3) whose random walk component satisfies (2.10) and whose second-degree U-statistic component satisfies the following condition.

Condition (C). $E|\beta(\mathbf{Z}_1, \mathbf{Z}_2)|^r < \infty$ for some $r > 2$ and there exist K Borel functions $f_\nu : \mathbf{R}^p \to \mathbf{R}$ such that $K(r-2) > 32r - 40$, $Ef_\nu^2(\mathbf{Z}_1) < \infty$ and the covariance matrix of $(W_1, \ldots, W_K)$ is positive definite, where $W_\nu = (Lf_\nu)(\mathbf{Z}_1)$ and $(Lf)(\mathbf{z}) = E\{\beta(\mathbf{z}, \mathbf{Z}_2)f(\mathbf{Z}_2)\}$.

With the same assumptions as in Example 1 and with β given there, it can be shown by an argument similar to that of Example 1 of Lai and Wang (1993) that Condition (C) is satisfied. Hence if (2.10) holds with $\alpha(\mathbf{Z}_i)$ given in Example 1, then $U_n = \sqrt{n}(\hat{F}_a(t) - F_a(t))/(1 - F_a(t))$ has the Edgeworth expansion (2.11) in which σ is given by (2.9) and $P_1(z)$, $P_2(z)$ are polynomials in z whose coefficients involve product-moments of $\alpha(\mathbf{Z}_1)$, $\alpha'(\mathbf{Z}_1)$, $\beta(\mathbf{Z}_1, \mathbf{Z}_2)$ and $\gamma(\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3)$, cf. Theorem 1 of Lai and Wang (1993).

Example 2. Under the assumptions of Example 1, a consistent estimate of (2.9) is

$$\hat{\sigma}^2 := n \sum_{a\le s\le t} \Delta N_n(s)/R_n^2(s) = \int_a^t \frac{dN_n(s)/n}{(R_n(s)/n)^2},$$

where $N_n(s)$ and $R_n(s)$ are defined in (1.5), cf. Gross and Lai (1994). Take any sequence of positive constants ϵ_n such that $\lim_{n\to\infty}\epsilon_n = 0$. The following representation of the Studentized statistic as an asymptotic U-statistic will be proved in the Appendix:

$$\begin{aligned} V_n :&= \sqrt{n}(\hat{F}_a(t) - F_a(t))/\max\{\hat{\sigma}(1-\hat{F}_a(t)), \epsilon_n\} \\ &= \sum_{i=1}^n \Big\{\frac{\alpha_V(\mathbf{Z}_i)}{\sqrt{n}} + \frac{\alpha'_V(\mathbf{Z}_i)}{n^{3/2}}\Big\} + \sum_{1\le i<j\le n} \frac{\beta_V(\mathbf{Z}_i,\mathbf{Z}_j)}{n^{3/2}} \\ &\quad + \sum_{1\le i<j<k\le n} \frac{\gamma_V(\mathbf{Z}_i,\mathbf{Z}_j,\mathbf{Z}_k)}{n^{5/2}} + D_n, \end{aligned} \tag{2.12}$$

where, using the same notation as in Example 1, $\alpha_V(\mathbf{Z}_i) = \alpha(\mathbf{Z}_i)/\sigma$, α'_V, β_V and γ_V are given in the Appendix and $P\{|D_n| \ge n^{-1-\epsilon}\} = o(n^{-1})$ for every $0 < \epsilon < 1/2$. Hence under (2.10) on α and Condition (C) on β_V, we can again apply Theorem 1 of Lai and Wang (1993) to obtain an Edgeworth expansion of the form

$$\begin{aligned} &P\{\sqrt{n}(\hat{F}_a(t) - F_a(t))/\max[\hat{\sigma}(1-\hat{F}_a(t)), \epsilon_n] \le z\} \\ &= \Phi(z) - n^{-1/2}\phi(z)P_1(z) - n^{-1}\phi(z)P_2(z) + o(n^{-1}). \end{aligned} \tag{2.13}$$

The next two examples apply the asymptotic U-statistic representation to a wide class of statistics based on l.t.r.c. data. Let Ψ be the common distribution function of $(\mathbf{X}_{i,o}, \tilde{Y}_{i,o}, \delta_{i,o}, T_{i,o})$ and let $\mu(\Psi)$ be a functional of Ψ. A functional $J(K,H)$ of the distribution K of $(\mathbf{X}_1, Y_1)$ and the distribution H of (T_1, C_1) is said to be *estimable* if $J(K,H)$ can be expressed as a functional $\mu(\Psi)$ of Ψ. In particular, for $\tau < a < t < (\tau^* \wedge F^{-1}(1))$,

$$\begin{aligned} \Lambda_a(t) :&= \int_a^t \frac{dF(s)}{1-F(s-)} = \int_a^t \frac{P\{T_1 \le s \le C_1\}dF(s)}{P\{T_1\le s\le C_1\}(1-F(s-))} \\ &= \int_a^t \frac{dP\{T_1 \le Y_1 \le s\wedge C_1\}}{P\{T_1\le s\le Y_1\wedge C_1\}} = \int_a^t \frac{dP\{\tilde{Y}_{1,o}\le s, \delta_{1,o}=1\}}{P\{T_{1,o}\le s\le \tilde{Y}_{1,o}\}} \quad \text{(by Lemma 1)} \end{aligned}$$

is a functional of Ψ and is therefore an estimable functional of F. Hence

$$F_a(t) = 1 - e^{-\Lambda_a^c(t)} \prod_{a\le s\le t} (1 - \Delta\Lambda_a(s)) \tag{2.14}$$

is also an estimable functional. The notation Λ_a^c in (2.14) denotes the continuous part of Λ_a. Clearly $T_{i,o} \geq \tau$ and $\tilde{Y}_{i,o} \leq \tau^*$ with probability 1, and therefore only $F_\tau(y)$ with $y \leq \tau^*$ can be nonparametrically estimated from the observed data (cf. Keiding and Gill (1990) and Lai and Ying (1991)).

Example 3. Gross and Lai (1994) introduced a general class of estimable functionals of the joint distribution K of $(\mathbf{X}_1, Y_1)$ and developed asymptotically normal estimates of these functionals. Closely related to Example 1 are the trimmed means $\mu(a,b) := E\{h(Y_1)|a \leq Y_1 \leq b\} = \{\int_a^b h(y)dF_a(y)\}/F_a(b)$ for smooth functions h, where $\tau < a < b < (\tau^* \wedge F^{-1}(1))$. For continuous h, Gross and Lai (1994) showed that $\sqrt{n}(\hat{\mu}_n(a,b) - \mu(a,b))$ has a limiting normal distribution, and an argument similar to that of Examples 1 and 2 can be used to show that $\sqrt{n}(\hat{\mu}(a,b) - \mu(a,b))$ is an asymptotic U-statistic, where $\hat{\mu}(a,b) = \int_a^b h(y)d\hat{F}_a(y)/\hat{F}_a(b)$ and $\hat{F}_a$ is the product-limit estimator (2.5) of F_a. Motivated by applications to regression analysis and curve fitting, they also considered estimating the parameter β^*, defined as the minimizer of $g(\beta) := E\{\rho(Y_1 - \beta^T\mathbf{X}_1)I(a \leq Y_1 \leq b)\}$ in some region D, by the solution $\hat{\beta}$ of the estimating equation

$$\sum_{i:a\leq\tilde{Y}_{i,o}\leq b} \delta_{i,o}\mathbf{X}_{i,o}\rho'(\tilde{Y}_{i,o} - \beta^T\mathbf{X}_{i,o})\hat{S}_a(\tilde{Y}_{i,o})/R_n(\tilde{Y}_{i,o}) = 0,$$

where $\hat{S}_a(t) = \prod_{a\leq s<t}\{1 - \Delta N_n(s)/R_n(s)\}$, in analogy with (1.4), $R_n(s)$ and $N_n(s)$ are defined in (1.5), and ρ is assumed to be convex and differentiable so that $g'(\beta)$ has a unique solution β^* in D. Under certain regularity conditions, the arguments used to prove the asymptotic normality of $\hat{\beta}$ in Theorem 2 of Gross and Lai (1994) can be refined along the lines of Example 1 to show that $\sqrt{n}(\hat{\beta}_n - \beta)$ is an asymptotic U-statistic.

Example 4. While Example 3 considers estimable functionals of the joint distribution K of $(\mathbf{X}_1, Y_1)$, we now consider estimable functionals of both K and the bivariate distribution H of (T_1, C_1) under the assumption that

$$F(\tau) = 0, \quad P\{T_1 \leq C_1\} = 1 = P\{T_1 \leq F^{-1}(1)\}, \tag{2.15}$$

as assumed by Wang (1991). Let $p = P\{\tilde{Y}_1 \geq T_1\} = P\{Y_1 \geq T_1\}$ (since $T_1 \leq C_1$). By Lemma 1,

$$P\{T_{1,o} \leq t\} = p^{-1}P\{T_1 \leq Y_1 \wedge t\} = p^{-1}\int_\tau^t S(u)dG(u),$$

where $S(u) = P\{Y_1 \geq u\}$ and $G(u) = P\{T_1 \leq u\}$. Therefore $p^{-1}S(t)dG(t) = dP\{T_{1,o} \leq t\}$. Combining this with the assumption $F^{-1}(1) \geq G^{-1}(1)$ $(= \tau^*$ by

(1.1)) yields

$$p = p\int_{\tau}^{G^{-1}(1)} dG(t) = \int_{\tau}^{\tau^*\wedge F^{-1}(1)} (S(t))^{-1} dP\{T_{1,o} \le t\}. \tag{2.16}$$

Since $F(\tau) = 0$ by (2.15), $S(t)$ is estimable for $t \le \tau^* \wedge F^{-1}(1)$ (cf. (2.14)) and therefore p is estimable by (2.16). Using heuristic arguments involving a "working" data set that involves unobservable censoring variables, Wang (1991) proposed to estimate p by $\hat{p} = n/\sum_{i=1}^n (\hat{S}(T_{i,o}))^{-1}$, where $\hat{S}(\cdot)$ is defined in (1.4), and showed that $\hat{p}$ is a consistent and asymptotically normal estimate of p under certain regularity conditions. Note that (2.16) provides an alternative justification of Wang's estimator $\hat{p}$. Moreover, since $G(t) = \int_{\tau}^{t} dG(u) = p\int_{\tau}^{t}(S(u))^{-1}dP\{T_{1,o} \le u\}$ and since p is estimable, it follows that $G(t)$ is also estimable and can be consistently estimated by

$$\hat{G}(t) = \hat{p}n^{-1}\sum_{i:T_{i,o}\le t} (\hat{S}(T_{i,o}))^{-1} = \Big\{\sum_{i:T_{i,o}\le t} (\hat{S}(T_{i,o}))^{-1}\Big\}/\Big\{\sum_{i=1}^{n} (\hat{S}(T_{i,o}))^{-1}\Big\},$$

which is the same as Wang's estimate (1.6) derived by heuristic "working-data" arguments. Moreover, if (2.15) is replaced by the stronger assumption that (1.3) holds and $P\{T_1 \le C_1\} = 1$, then arguments similar to those in Example 1 can be used to show that $\sqrt{n}(\hat{G}(t) - G(t))$ and $\sqrt{n}(\hat{p} - p)$ are asymptotic U-statistics.

2.2. Asymptotic theory of the simple bootstrap method

Let Ψ_n denote the empirical distribution that puts probability $1/n$ at each $\mathbf{Z}_i = (\mathbf{X}_{i,o}, \tilde{Y}_{i,o}, \delta_{i,o}, T_{i,o})$, $i = 1,\ldots,n$. The simple bootstrap sample consists of i.i.d. random vectors $\mathbf{Z}_1^*,\ldots,\mathbf{Z}_n^*$ with common distribution Ψ_n. Let $S = S(\mathbf{Z}_1,\ldots,\mathbf{Z}_n)$ be an estimate of the functional $\mu(\Psi)$ and let $\hat{\sigma} = \hat{\sigma}(\mathbf{Z}_1,\ldots,\mathbf{Z}_n)$ be an estimate of the standard error of S. The simple bootstrap method estimates the sampling distribution $P\{(S - \mu(\Psi))/\hat{\sigma} \le z\}$ by $P\{(S^* - \mu(\Psi_n))/\hat{\sigma}^* \le z|\mathbf{Z}_1,\ldots,\mathbf{Z}_n\}$, where $S^* = S(\mathbf{Z}_1^*,\ldots,\mathbf{Z}_n^*)$ and $\hat{\sigma}^* = \hat{\sigma}(\mathbf{Z}_1^*,\ldots,\mathbf{Z}_n^*)$. In most applications, $\sqrt{n}(S - \mu(\Psi))/\hat{\sigma}$ can be expressed as an asymptotic U-statistic which has a limiting standard normal distribution. The same argument can be used to represent $\sqrt{n}(S^* - \mu(\Psi_n))/\hat{\sigma}^*$ as an asymptotic U-statistic such that

$$P\{\sqrt{n}(S^* - \mu(\Psi_n))/\hat{\sigma}^* \le z|\mathbf{Z}_1,\ldots,\mathbf{Z}_n\} \xrightarrow{P} \Phi(z), \tag{2.17}$$

from which it follows that $|P\{\sqrt{n}(S^* - \mu(\Psi_n))/\hat{\sigma}^* \le z|\mathbf{Z}_1,\ldots,\mathbf{Z}_n\} - P\{\sqrt{n}(S - \mu(\Psi))/\hat{\sigma} \le z\}| \xrightarrow{P} 0$, giving an asymptotic justification of the simple bootstrap method. Furthermore, under (2.10) and Condition (C), asymptotic U-statistics and their simple bootstrap versions have Edgeworth expansions whose difference is of the order $O_p(n^{-1})$, establishing the second-order accuracy of the simple bootstrap method for l.t.r.c. data. This is a consequence of the following result of Lai and Wang (1993).

Lemma 2. *Let $\mathbf{Z}_1, \mathbf{Z}_2, \ldots,$ be i.i.d. p-dimensional random vectors with common distribution Ψ and let $U_n = U_n(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$ be an asymptotic U-statistic defined by* (2.3) *and* (A1)-(A4). *Suppose that α satisfies* (2.10) *and β satisfies Condition* (C). *Let $\Psi_n(B) = n^{-1}\sum_{i=1}^n I(\mathbf{Z}_i \in B)$ denote the empirical distribution, and let $(\mathbf{Z}_1^*, \ldots, \mathbf{Z}_n^*)$ be i.i.d. with common distribution Ψ_n. Suppose that there exist functions $\hat{\alpha}_n$, $\hat{A}_n$, $\hat{\beta}_n$, $\hat{\gamma}_n$, Δ_n, depending on Ψ_n and invariant under permutation of arguments, such that*

$$n^{-1}\sum_{i=1}^n |\hat{A}_n(\mathbf{Z}_i)|^3 + n^{-3}\sum_{1\le i<j<k\le n} |\hat{\gamma}_n(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k)|^4 = O_p(1),$$

$$\begin{aligned}
\sum_{i=1}^n \hat{\alpha}_n(\mathbf{Z}_i) &= \sum_{i=1}^n \hat{A}_n(\mathbf{Z}_i) = 0 = \sum_{i=1}^n \hat{\beta}_n(\mathbf{z}_1, \mathbf{Z}_i) \\
&= \sum_{i=1}^n \hat{\gamma}_n(\mathbf{z}_1, \mathbf{z}_2, \mathbf{Z}_i), \quad \textit{for any} \quad \mathbf{z}_1, \mathbf{z}_2 \in \mathcal{S}(\Psi), \\
n^{-1}\sum_{i=1}^n \hat{\alpha}_n^2(\mathbf{Z}_i) &= 1 = E\alpha^2(\mathbf{Z}_1),
\end{aligned}$$

$$\sup_{\mathbf{z}\in\mathcal{S}(\Psi)} \frac{|\hat{\alpha}_n(\mathbf{z}) - \alpha(\mathbf{z})|}{\mathbf{1} + |\alpha(\mathbf{z})|} + \sup_{\mathbf{z},\mathbf{z}'\in\mathcal{S}(\Psi)} |\hat{\beta}_n(\mathbf{z}, \mathbf{z}') - \beta(\mathbf{z}, \mathbf{z}')| = O_p(n^{-1/2}),$$

where $\mathcal{S}(\Psi)$ denotes the support of Ψ. Let

$$\begin{aligned}
U_n^* &= \sum_{i=1}^n \Big\{\frac{\hat{\alpha}_n(\mathbf{Z}_i^*)}{\sqrt{n}} + \frac{\hat{A}_n(\mathbf{Z}_i^*)}{n^{3/2}}\Big\} + \sum_{1\le i<j\le n} \frac{\hat{\beta}_n(\mathbf{Z}_i^*, \mathbf{Z}_j^*)}{n^{3/2}} \\
&\quad + n^{-5/2}\sum_{1\le i<j<k\le n} \hat{\gamma}_n(\mathbf{Z}_i^*, \mathbf{Z}_j^*, \mathbf{Z}_k^*) + \Delta_n(\mathbf{Z}_1^*, \ldots, \mathbf{Z}_n^*), \qquad (2.18)
\end{aligned}$$

where $nP\{|\Delta_n(\mathbf{Z}_1^*, \ldots, \mathbf{Z}_n^*)| \ge n^{-1-\epsilon}|\Psi_n\} \xrightarrow{P} 0$ for some $\epsilon > 0$. Then $P\{U_n \le z\}$ has the Edgeworth expansion (2.11) *and*

$P\{U_n^* \le u|\Psi_n\} = \Phi(u) - n^{-1/2}\phi(u)P_1(u) + O_p(n^{-1})$ *uniformly in* $-\infty < u < \infty$.

Consequently, $\sup_u |P\{U_n \le u\} - P\{U_n^* \le u|\Psi_n\}| = O_p(n^{-1})$.

As an illustration of the applications of Lemma 2, we now show that $P\{\sqrt{n}$ $(\hat{F}_a(t) - F_a(t))/\max[\hat{\sigma}(1 - \hat{F}_a(t)), \epsilon_n] \le u\}$ in Example 2 can be approximated by the simple bootstrap estimate with an error of the order $O_p(n^{-1})$. Let $\mathbf{Z}_i = (\mathbf{X}_{i,o}, \tilde{Y}_{i,o}, \delta_{i,o}, T_{i,o})$ and let Ψ denote the common distribution of the $\mathbf{Z}_i$. As pointed out in (2.14), $F_a(t)$ can be expressed as a functional $\mu(\Psi)$ of Ψ since $\Lambda_a(s)$ can be expressed as a functional of Ψ. Comparison of (2.5) with such representations of Λ_a and F_a shows that $\hat{F}_a(t) = \mu(\Psi_n)$. Moreover, the same argument as in the Appendix shows that $U_n^* = \sqrt{n}(\hat{F}_a^*(t) - \hat{F}_a(t))/\max\{\hat{\sigma}^*(1 - \hat{F}_a^*(t)), \epsilon_n\}$ has the representation (2.18) with $\hat{\alpha}_n$, $\hat{A}_n$, $\hat{\beta}_n$, $\hat{\gamma}_n$ satisfying the assumptions of Lemma 2, as can be shown by an argument similar to that on page 528 of Lai and Wang (1993).

2.3. Numerical examples of applications of the simple bootstrap to l.t.r.c. data

In the following two examples, we apply the simple bootstrap method to construct confidence intervals for survival probabilities in a real data set and to estimate the bias and standard error of a regression estimator in several simulated data sets.

Example 5. Table 1 of Hyde (1977) shows ages in months at death of 97 men in Channing House in Palo Alto, California. The truncation times are ages in months at entry into the community in 1965. Censoring times are ages in months at the end of the study on July 1, 1975, or the age at withdrawal from the community. Confidence intervals for survival probabilities have been computed from these data by Tsai, Jewell and Wang (1987) using normal approximations and by Wang (1991) using the obvious bootstrap method that assumes independence between $C - T$ and T, where T and C are the truncation and censoring variables respectively. This independence assumption may not be justified in the present case because there were 5 censored cases of unknown cause, although censoring due to termination of the study, with left truncation caused by death prior to study initiation, does not violate the independence assumption. We therefore applied, instead, the simple bootstrap method to these data. Estimation was performed conditionally given that $Y \geq 867$ months. This conditioning was also used by Tsai, Jewell and Wang (1987) and Wang (1991), and the choice of 867 months yields risk set sizes $R_n(s)$ that are not too small for $s \geq 867$. As explained in Gross and Lai (1994), it is not possible to estimate the entire survival distribution when left truncation and right censoring are present.

In Table 1 we display three types of (pointwise) 95% confidence intervals for the conditional survival probabilities $P\{Y \geq y | Y \geq 867\}$ at several ages y between 909 and 1012 months. The first two types are (simple-)bootstrap confidence intervals using the percentile-t and the percentile methods (cf. Hall (1988)), while the third type of confidence intervals is based on the normal approximation to $\sqrt{n}(\hat{F}_a(y-) - F_a(y-))/\max\{\hat{\sigma}(1 - \hat{F}_a(y-)), \epsilon_n\}$ (see Example

2). Here $a = 867$ and we take $\epsilon_n = 0.01$. Each bootstrap estimate reported in the table is based on 1000 bootstrap samples. The percentile-t bootstrap confidence intervals for $1 - F_a(y-)$ tend to be asymmetric about the estimate $1 - \hat{F}_a(y-)$ and to be narrower than the other two types of confidence intervals. For example, at $y = 998$, the 95% percentile-t bootstrap confidence interval for $P\{Y \geq 998|Y \geq 867\}$ (which is estimated to be 0.502) is $[0.332, 0.621]$, while the 95% bootstrap confidence interval using the percentile method is $[0.358, 0.644]$ which, like the normal confidence interval, is almost symmetric about 0.502.

Table 1. Estimates $\hat{p}$ of conditional survival probabilities $P\{Y \geq y|Y \geq 867\}$ from Hyde's data, and the associated confidence intervals $[p_L, p_U]$ obtained by the bootstrap percentile-t and the percentile methods, and by using normal approximations.

		Percentile-t		Percentile		Normal Approx	
Age y	$\hat{p}$	p_L	p_U	p_L	p_U	p_L	p_U
909	0.784	0.674	0.957	0.652	0.938	0.624	0.944
911	0.761	0.640	0.938	0.631	0.921	0.592	0.930
927	0.715	0.609	0.903	0.593	0.895	0.547	0.884
932	0.696	0.571	0.878	0.567	0.878	0.504	0.887
945	0.676	0.557	0.845	0.545	0.855	0.487	0.866
948	0.657	0.519	0.813	0.529	0.828	0.454	0.861
957	0.638	0.488	0.798	0.505	0.809	0.427	0.850
966	0.620	0.468	0.779	0.479	0.793	0.411	0.828
969	0.603	0.448	0.745	0.469	0.767	0.395	0.810
971	0.588	0.420	0.734	0.442	0.746	0.377	0.799
983	0.573	0.403	0.712	0.424	0.723	0.363	0.782
985	0.558	0.383	0.687	0.410	0.701	0.352	0.764
989	0.544	0.371	0.670	0.397	0.684	0.341	0.746
993	0.530	0.359	0.647	0.382	0.667	0.332	0.727
998	0.502	0.332	0.621	0.358	0.644	0.299	0.705
1009	0.488	0.319	0.610	0.352	0.631	0.289	0.687
1012	0.472	0.298	0.587	0.332	0.612	0.275	0.668

Example 6. In this example we apply the simple bootstrap method to estimate, from each of six sets of simulated data, the bias and variance of the estimator $\hat{\beta}$, defined by the estimating equation (2.24), with $\rho(u) = u^2$, of the minimizer β of $E\{(Y - \beta^T\mathbf{X})^2 I(a \leq Y \leq b)\}$. These data sets have been generated from the following simulation experiment. Let X be uniformly distributed on $[0, 4]$ and let W be uniform on $[0, 1]$ and independent of X. Let $Y = 1 + X + \gamma X^2 + 2W^k$. A pair (X, Y) generated in this way is retained if $Y \geq T$, where $T + 0.5$ has

the exponential distribution with mean λ and is independent of (X, Y). Triplets (X, Y, T) are generated until $n = 50$ such triplets are obtained. For each such triplet, generate $C = T + 1 + 15U$, where U is uniformly distributed on $[0, 1]$ and is independent of (X, Y, T), and let $\tilde{Y} = Y \wedge C$, $\delta = I(Y \leq C)$. Note that $C \geq T$ and that $\tau = -0.5$, $\tau^* = \infty$ (cf. (1.1)) while $(\tau <)1 \leq Y \leq 7 + 16\gamma(< \tau^*)$. Thus, the support of Y is inside (τ, τ^*) and we can, therefore, take $a = -\infty$ and $b = \infty$ in the definitions of β and $\hat{\beta}$ (cf. Gross and Lai (1994)). With $a = -\infty$ and $b = \infty$, the minimizer $\beta = (\beta_1, \beta_2)^T$ of $E(Y - \beta_1 - \beta_2 X)^2$ is given by

$$\beta_2 = \{E(XY) - (EX)(EY)\}/\mathrm{Var}\,(X), \quad \beta_1 = E(Y) - \beta_2 E(X), \tag{2.19}$$

which are the slope and intercept terms, respectively, of the best linear approximation to the regression function. Note that when the support of Y is not inside $[\tau, \tau^*]$, this best linear approximations is not estimable and one has to consider as in Example 3 the minimizer of $E[(Y - \beta_1 - \beta_2 X)^2 I(a \leq Y \leq b)]$ instead. The use of $a > \tau$ and $b < \tau^*$ also avoids the difficulties caused by small risk set sizes near τ and τ^*, as in the treatment of the Channing House data in Example 5.

Table 2 considers six such simulated data sets $(\mathbf{X}_{i,o}, \tilde{Y}_{i,o}, \delta_{i,o}, T_{i,o})$, $i = 1, \ldots, 50$. Data sets I and II are generated using the following values of the parameters γ, k and $\lambda : \gamma = 0.2$, $k = 3$ and $\lambda = 4$, which give the truncation probability $P\{T > Y\} = 0.298$ and the censoring probability $P\{C > Y | Y \geq T\} = 0.144$. Data sets III and IV are generated using $\gamma = 0.2$, $k = 3$ and $\lambda = 6$, which give $P\{T > Y\} = 0.435$ and $P\{C > Y | Y \geq T\} = 0.133$. Data sets V and VI are generated using $\gamma = 0$, $k = 3$ and $\lambda = 4$, which give $P\{T > Y\} = 0.365$ and $P\{C > Y | Y \geq T\} = 0.081$. These truncation and censoring probabilities and the corresponding values of β_1 and β_2 in (2.19), together with the means and standard deviations (also reported in Table 2) of the sampling distributions of $\hat{\beta}_1$ and $\hat{\beta}_2$, were computed by the Monte Carlo method involving 500 simulations for each result. From each of the six data sets, we computed the estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ and their estimated standard errors using the asymptotic formulas (3.6) and (3.7) of Gross and Lai (1994). We also used the simple bootstrap method to estimate the bias $(= E(\hat{\beta}_j - \beta_j))$ and the standard deviation $\mathrm{SD}(\hat{\beta}_j)$ of $\hat{\beta}_j$ $(j = 1, 2)$. The results are given in Table 2, where each bootstrap estimate is based on 500 (simple-)bootstrap samples. Particularly noteworthy is data set III, in which the estimate $\hat{\beta}_1$ differs substantially from β_1 but the bootstrap estimate of the bias is still quite close to the population value of the bias. When the standard deviations of $\hat{\beta}_j$ estimated from the asymptotic theory are reasonably close to the population values, they are also close to the bootstrap estimates. However, when $\mathrm{SD}(\hat{\beta}_j)$ is not well estimated by the asymptotic theory (as in $\hat{\beta}_2$ for data sets II, III and V and $\hat{\beta}_1$ for data set III), the bootstrap method seems to provide substantial improvement.

Table 2. Estimates $\hat{\beta}_j$ of β_j, bias and standard deviation (SD) of $\hat{\beta}_j$, bootstrap estimates $\text{Bias}^*(\hat{\beta}_j)$ and $\text{SD}^*(\hat{\beta}_j)$ of $\text{Bias}(\hat{\beta}_j)$ and $\text{SD}(\hat{\beta}_j)$, and estimate $\widehat{\text{SD}}(\hat{\beta}_j)$ of $\text{SD}(\beta_j)$ using asymptotic normality of $\hat{\beta}_j$, for $j = 1, 2$, from six sets of simulated l.t.r.c. data with $n = 50$.

Data Set	I	II	III	IV	V	VI
β_1	0.969	0.969	0.969	0.969	1.491	1.491
$\hat{\beta}_1$	0.841	1.200	0.227	0.812	1.573	1.380
$\text{Bias}(\hat{\beta}_1)$	−0.013	−0.013	−0.006	−0.006	0.001	0.001
$\text{Bias}^*(\hat{\beta}_1)$	−0.012	−0.005	−0.017	−0.020	0.004	−0.007
$\text{SD}(\hat{\beta}_1)$	0.205	0.205	0.228	0.228	0.185	0.185
$\text{SD}^*(\hat{\beta}_1)$	0.196	0.180	0.140	0.206	0.168	0.142
$\widehat{\text{SD}}(\hat{\beta}_1)$	0.185	0.178	0.121	0.210	0.194	0.148
β_2	1.798	1.798	1.798	1.798	1	1
$\hat{\beta}_2$	1.919	1.713	2.101	1.882	1.006	1.051
$\text{Bias}(\hat{\beta}_2)$	0.104	0.104	−0.001	−0.001	0.004	0.004
$\text{Bias}^*(\hat{\beta}_2)$	0.002	0.005	−0.007	−0.003	−0.001	0.003
$\text{SD}(\hat{\beta}_2)$	0.088	0.088	0.094	0.094	0.077	0.077
$\text{SD}^*(\hat{\beta}_2)$	0.077	0.081	0.056	0.088	0.073	0.068
$\widehat{\text{SD}}(\hat{\beta}_2)$	0.079	0.102	0.042	0.100	0.108	0.083

3. Estimation of F, H and the Obvious Bootstrap

It will be assumed throughout this section that $C_i \geq T_i$, as in Wang (1991). First consider the problem of estimating the common distribution function F of the Y_i and the common bivariate distribution H of (T_i, C_i) without assuming independence between T_i and $C_i - T_i$ as in (1.2). This estimation problem is basic to the obvious bootstrap method for l.t.r.c. data, and we shall explain why the simple and obvious bootstrap methods are no longer equivalent even when the data are only subject to truncation but not to censoring. It will be shown that consistent estimation of (F, H) requires complicated smoothing and restrictive assumptions, leading us to the conclusion that the simple bootstrap method for l.t.r.c. data is preferable to the obvious bootstrap method in practice.

3.1. Nonparametric maximum likelihood estimate of (F, H)

Let $\Psi(c|t) = P\{C_1 \leq c | T_1 = t\}$. Note that $H(t, c) = \int_{-\infty}^{t} \Psi(c|s) dG(s)$, where G is the distribution function of T_1. The likelihood function L of (F, G, Ψ) is given by

$$L = \prod_{i=1}^{n} \Big\{ \Delta F(\tilde{Y}_{i,o}) \Delta G(T_{i,o}) (1 - \Psi(\tilde{Y}_{i,o} - |T_{i,o}))/p \Big\}^{\delta_{i,o}} \times \Big\{ \Delta \Psi(\tilde{Y}_{i,o} | T_{i,o}) \Delta G(T_{i,o}) (1 - F(\tilde{Y}_{i,o}))/p \Big\}^{1-\delta_{i,o}},$$

where $\Delta\Psi(c|t) = \Psi(c|t) - \Psi(c-|t)$ and

$$p = P\{\tilde{Y}_1 \geq T_1\} = P\{Y_1 \geq T_1\} = \int_{-\infty}^{\infty} (1 - F(t-))dG(t), \tag{3.1}$$

since $C_1 \geq T_1$. Let $t_1 < \cdots < t_m$ denote the distinct values of the $T_{i,o}$, and let $k_j = \sum_{i=1}^n I(T_{i,o} = t_j)$. As in Wang (1991), we decompose L into three factors, yielding

$$L = L_1 L_2 L_3, \text{ where } L_1 = \prod_{i=1}^n \{(\Delta F(\tilde{Y}_{i,o}))^{\delta_{i,o}} (1 - F(Y_{i,o}))^{1-\delta_{i,o}} / (1 - F(T_{i,o}-))\},$$

$$L_2 = \prod_{j=1}^m \{\Delta G(t_j)(1 - F(t_j-))/p\}^{k_j},$$

$$L_3 = \prod_{j=1}^m \Big\{ \prod_{i:T_{i,o}=t_j} (1 - \Psi(\tilde{Y}_{i,o} - |t_j))^{\delta_{i,o}} (\Delta\Psi(\tilde{Y}_{i,o}|t_j))^{1-\delta_{i,o}} \Big\},$$

from which it follows that the maximizer $(\hat{F}, \hat{G}, \hat{\Psi}; \hat{p})$ of L can be characterized as follows:

(i) $1 - \hat{F}(t-) = \hat{S}(t)$, where $\hat{S}$ is given by (1.4), in view of L_1.

(ii) $\hat{G}$ is a step function with jumps at $t_j (j = 1, \ldots, m)$, so $\hat{p} = \sum_{i=1}^n (1 - \hat{F}(T_{i,o}-))\Delta G(T_{i,o})$ in view of (3.1).

(iii) $\Delta\hat{G}(t_j)\hat{S}(t_j)$ is proportional to k_j in view of L_2, and therefore $\hat{G}$ is given by (1.6).

(iv) $\hat{\Psi}(\cdot|t_j)$ is the product-limit estimator of the censoring distribution $\Psi(\cdot|t_j)$ based on $\{(\tilde{Y}_{i,o}, 1 - \delta_{i,o}) : T_{i,o} = t_j\}$, in view of L_3.

Under (1.3), $\hat{F}$ and $\hat{G}$ are consistent estimates of F and G, as shown by Wang (1991). However, when the bivariate distribution H of (T_1, C_1) is continuous, no more than one censored $\tilde{Y}_{i,o}$ can be associated with each t_j and therefore $\hat{\Psi}(\cdot|t_j)$ fails to give a consistent estimate of $\Psi(\cdot|t_j)$. To circumvent this difficulty, one way is to impose certain independence assumptions as in (1.2) so that one can replace the conditional distribution $\Psi(\cdot|t_j)$ by some marginal distribution that can be estimated. Another approach is to impose suitable smoothness assumptions on H so that $\Psi(c|t)$ is well approximated by $P\{C_1 \leq c | G(t) - \Delta_n \leq G(T_1) \leq G(t) + \Delta_n\}$ which can be consistently estimated when Δ_n approaches 0 at a certain rate depending on the sample size n as $n \to \infty$.

To begin with, note that the nonparametric maximum likelihood estimator $\hat{\Psi}(\cdot|t_j)$ can be expressed as

$$\hat{\Psi}(c|t) = 1 - \prod_{s \leq c} \Big\{ 1 - \frac{\sum_{j=1}^n I(\tilde{Y}_{j,o} = s, \delta_{j,o} = 0) I(T_{j,o} = t)}{\sum_{j=1}^n I(\tilde{Y}_{j,o} \geq s \geq T_{j,o}) I(T_{j,o} = t)} \Big\}. \tag{3.2}$$

If G is continuous, the denominator in (3.2) has at most one nonzero summand, causing the inconsistency of $\hat{\Psi}(c|t)$. To increase the number of nonzero summands, we shall replace $I(T_{j,o} = t)$ in (3.2) by $a_n^{-1}K(\{\hat{G}(t) - \hat{G}(T_{j,o})\}/a_n)$, where K is a smooth probability density, $\{a_n\}$ is a sequence of positive constants converging to 0 and $\hat{G}$ is the estimate (1.6) of G. Thus, instead of estimating $\Psi(c|t)$ by (3.2), we use the estimate

$$\bar{\Psi}(c|t) = 1 - \prod_{s \le c} \Big\{1 - \frac{\sum_{j=1}^n I(\tilde{Y}_{j,o} = s, \delta_{j,o} = 0) K(\{\hat{G}(t) - \hat{G}(T_{j,o})\}/a_n)}{\sum_{j=1}^n I(\tilde{Y}_{j,o} \ge s \ge T_{j,o}) K(\{\hat{G}(t) - \hat{G}(T_{j,o})\}/a_n)}\Big\}. \quad (3.3)$$

Noting that $H(t,c) = \int_{-\infty}^t \Psi(c|s) dG(s)$, we estimate H by

$$\hat{H}(t,c) = \sum_{i=1}^n I(T_{i,o} \le t) \bar{\Psi}(c|T_{i,o}) \Delta\hat{G}(T_{i,o}). \quad (3.4)$$

This estimator is similar to that of Akritas (1994) for the problem of estimating the bivariate distribution of (X, Y) when Y is subject only to right censoring (and no left truncation) and X is a completely observable covariate. Under (1.3) and quite strong smoothness assumptions on H similar to those of Akritas on the bivariate distribution of (X, Y), it can be shown that if the a_n in (3.3) satisfy $na_n^4 \to 0$ but $na_n^3 |\log a_n|^{7/2} \to \infty$ then

$$\sup_{t \le c \le F^{-1}(1)(<\tau^*)} |\hat{H}(t,c) - H(t,c)| = O_p(n^{-1/2}).$$

The arguments are similar to those of Akritas (1994) and the details are omitted here.

3.2. The obvious bootstrap method for censored and truncated data

We begin with a brief review of why the simple and obvious bootstrap methods are equivalent for censored data. Suppose that Y_i are subject only to right censoring and let Q denote the common distribution function of the censoring variables C_i. Assume that Q and the distribution function F of the Y_i have no common discontinuities. The Kaplan-Meier estimates $\hat{F}$, $\hat{Q}$ of F, Q satisfy

$$\{1 - \hat{F}(t-)\}\{1 - \hat{Q}(t-)\} = \hat{R}(t), \quad (3.5)$$

where $\hat{R}(t) = n^{-1} \sum_{i=1}^n I(\tilde{Y}_i \ge t)$. Moreover, since F and Q have no common discontinuities, the set of discontinuities of $\hat{F}$ (= set of uncensored $\tilde{Y}_i$) is disjoint from the set of discontinuities of $\hat{G}$ (= set of censored $\tilde{Y}_i$). Therefore if $Y_1^*, \ldots, Y_n^*$, $C_1^*, \ldots, C_n^*$ are independent random variables with $Y_i^* \sim \hat{F}$ and $C_i^* \sim \hat{Q}$, then (3.5) implies that $\tilde{Y}_i^* (= Y_i^* \wedge C_i^*) \sim \hat{R}$ (= empirical distribution of $\tilde{Y}_1, \ldots, \tilde{Y}_n$).

Moreover, $I(\tilde{Y}_i^* = Y_i^*) = I(\tilde{Y}_j = Y_j)$ on the event $\{\tilde{Y}_i^* = \tilde{Y}_j\}$ since $\hat{F}$ and $\hat{Q}$ have no common discontinuities. Thus, even though the obvious bootstrap method samples independently Y_i^* from $\hat{F}$ and C_i^* from $\hat{Q}$, no new pairings between $\tilde{Y}_i^*$ and δ_i^* are added to the set of pairings $\{(\tilde{Y}_i, \delta_i), 1 \leq i \leq n\}$ in the original data, and by symmetry every member of this set of pairings is equally likely to be drawn by the obvious bootstrap. Hence the obvious and the simple bootstrap methods are equivalent.

We next consider the case of left truncated data in the absence of right censoring. Suppose that the $Y_i(\sim F)$ are subject only to left truncation by truncation variables $T_i(\sim G)$. Let $\hat{F}$, $\hat{G}$ be the product-limit estimators of F, G based on the observed data $\{(Y_{i,o}, T_{i,o}), 1 \leq i \leq n\}$ with $Y_{i,o} \geq T_{i,o}$. The obvious bootstrap method generates Y_i^*, T_i^* independently from $\hat{F}$ and $\hat{G}$, respectively. If $T_i^* \leq Y_i^*$, the pair is retained in the bootstrap sample, and (Y_i^*, T_i^*) is discarded if $T_i^* > Y_i^*$. This procedure is repeated until n such pairs $(Y_{i,o}^*, T_{i,o}^*)$ with $Y_{i,o}^* \geq T_{i,o}^*$ are generated to form the bootstrap sample. Since $\hat{G}$ assigns positive probability to every element of $\{T_{1,o}, \ldots, T_{n,o}\}$ while $\hat{F}$ assigns positive probability to every element of $\{Y_{1,o}, \ldots, Y_{n,o}\}$, it follows that any pair $(Y_{i,o}, T_{j,o})$ with $T_{j,o} \leq Y_{i,o}$ has positive probability of being included in the bootstrap sample. Hence, unlike the simple bootstrap sample which assigns probability $1/n$ to n observed pairs $(Y_{i,o}, T_{i,o})$, $1 \leq i \leq n$, the obvious bootstrap puts positive weights to the larger set of pairings $\{(Y_{i,o}, T_{j,o}) : 1 \leq i, j \leq n, Y_{i,o} \geq T_{j,o}\}$.

Since we assume no censoring (i.e., $C_i \equiv \infty$), (1.1) reduces to $\tau = \inf\{s : G(s) > 0\}$ and $\tau^* = \infty$. If (1.3) holds, then $\hat{F}$ and $\hat{G}$ are consistent estimates of F and G, respectively. However, without (1.3), $\hat{F}(t)$ converges in probability to $F(t)/(1 - F(\tau))$ while $\hat{G}(t)$ converges in probability to $G(t)/G(F^{-1}(1))$ (cf. Woodroofe (1985)), and it is not possible to estimate F and G consistently. Moreover, even when $F(\tau) = 0$ and $G(F^{-1}(1)) = 1$, $\hat{F}$ may still be a very poor estimator of F for moderate sample sizes because of the small risk set sizes $R_n(s)$ for s near τ and because of the non-monotonic oscillations in $R_n(s)$, as shown in a simulation study of Lai and Ying (1991), pages 440-441; and a similar comment also applies to $\hat{G}$. Lai and Ying (1991) proposed to modify the product-limit estimator $\hat{F}$ to circumvent this difficulty caused by small risk set sizes near τ and the simulation study suggests substantial improvement in using this modified version of the product-limit estimator. These results suggest that perhaps one should use instead of the product-limit estimators $\hat{F}$, $\hat{G}$ their modified versions in the obvious bootstrap method described above.

When the Y_i are subject to right censoring in addition to left truncation, difficulties in estimating the joint censoring-truncation distribution H increase substantially, although F can still be consistently estimated by the product-limit estimator $\hat{F}$ under (1.3). The discussion in Subsection 3.1 highlights the additional smoothness or independence assumptions that are needed for consistent estimation of H, and these assumptions seem to be too restrictive in practice. Therefore, for l.t.r.c. data, the simple bootstrap method is preferable to the obvious bootstrap method not only because it is substantially simpler to implement but also because it completely dispenses with the stringent assumptions that are needed for consistent estimation of (F, H) in the obvious bootstrap method. These advantages of the simple bootstrap method become even more pronounced if there are also covariates so that the observed data consist of $(\mathbf{X}_{i,o}, \tilde{Y}_{i,o}, \delta_{i,o}, T_{i,o})$, $i = 1, \ldots, n$, as in Section 2. Here the obvious bootstrap method would entail estimation of the joint distribution K of $(\mathbf{X}, Y)$ in addition to the bivariate distribution H of (T, C). Estimation of K is much harder than that of F (and even H). It involves further refinements and extensions of the ideas of Akritas (1994) and requires even more stringent smoothness assumptions. On the contrary, the simple bootstrap method can be directly applied to l.t.r.c. data with covariates, with no additional effort and no additional requirements in comparison with the well understood case of right censored data.

Acknowledgement

The research of Tze Leung Lai is supported by the National Science Foundation and the National Security Agency.

Appendix: Proof of (2.12)

By Lemma 1, $E(R_n(s)/n) = p(s) = (1-F(s))P\{T_1 \le s \le C_1\}/P\{Y_1 \wedge C_1 \ge T_1\}$, since F is assumed to be continuous. From (2.7) and (2.9), it follows that

$$\hat{\sigma}^2 - \sigma^2 = n^{-1} \sum_{i=1}^{n} \int_a^t (R_n(s)/n)^{-2} dM_i(s) + \int_a^t \{(R_n(s)/n)^{-1} - (p(s))^{-1}\} d\Lambda(s). \tag{A.1}$$

Since $R_n(s)/n = p(s) + n^{-1}\sum_{j=1}^n w_j(s)$, it follows from (A.1) and Taylor's expansions that

$$\begin{aligned}
\sigma/\hat{\sigma} &= \{1 + (\hat{\sigma}^2 - \sigma^2)/\sigma^2\}^{-1/2} \\
&= 1 - \frac{1}{2\sigma^2}\Big\{\frac{1}{n}\sum_{i=1}^{n}\int_a^t [\frac{1}{p^2(s)} - \frac{2}{np^3(s)}\sum_{j=1}^{n} w_j(s)] dM_i(s) \\
&\qquad\qquad + \int_a^t [-\frac{\sum_1^n w_j(s)}{np^2(s)} + \frac{(\sum_1^n w_i(s))^2}{n^2p^3(s)}] d\Lambda(s)\Big\} \\
&\quad + \frac{3}{8\sigma^4 n^2}\Big\{\sum_{i=1}^{n}\sum_{j=1}^{n}\int_a^t\int_a^t \frac{1}{p^2(s)p^2(u)} dM_i(s) dM_j(u)
\end{aligned}$$

$$+\int_a^t\int_a^t\Big(\frac{\sum_1^n w_i(s)}{p^2(s)}\Big)\Big(\frac{\sum_1^n w_j(u)}{p^2(u)}\Big)d\Lambda(s)d\Lambda(u)$$
$$-2\Big[\sum_{i=1}^n\int_a^t dM_i(s)/p^2(s)\Big]\Big[\sum_{j=1}^n\int_a^t (w_j(u)/p^2(u))d\Lambda(u)\Big]\Big\}+\cdots \qquad \text{(A.2)}$$

Letting $U_n=\sqrt{n}(\hat{F}_a(t)-F_a(t))/(1-F_a(t))$, note that

$$(1-F_a(t))/(1-\hat{F}_a(t))=(1-U_n/\sqrt{n})^{-1}=1+U_n/\sqrt{n}+U_n^2/2n+\cdots \qquad \text{(A.3)}$$

By the exponential inequalities for continuous-parameter martingales, $P\{\hat{\sigma}(1-\hat{F}_a(t))<\epsilon_n\}=o(\rho^n)$ for some $0<\rho<1$. From the representation of U_n in Example 1 and (A.2) and (A.3), (2.12) follows with $\alpha_V(\mathbf{Z}_i)=\alpha(\mathbf{Z}_i)/\sigma$ (using the same notation as in Example 1) and

$$\alpha_V'(\mathbf{Z}_i) = \alpha'(\mathbf{Z}_i)/\sigma-\alpha(\mathbf{Z}_i)\Big\{\int_a^t (p(s))^{-2}dM_i(s)-\int_a^t (w_i(s)/p^2(s))d\Lambda(s)$$
$$+\int_a^t[(1-p(s))/p^2(s)]d\Lambda(s)\Big\}/(2\sigma^3)+\alpha^2(\mathbf{Z}_i)/\sigma,$$

$$\beta_V(\mathbf{Z}_i,\mathbf{Z}_j)$$
$$=\beta(\mathbf{Z}_i,\mathbf{Z}_j)/\sigma-\Big\{\alpha(\mathbf{Z}_i)\int_a^t(p(s))^{-2}dM_j(s)+\alpha(\mathbf{Z}_j)\int_a^t(p(s))^{-2}dM_i(s)\Big\}/2\sigma^3$$
$$+\Big\{\alpha(\mathbf{Z}_i)\int_a^t\frac{w_j(s)}{p^2(s)}d\Lambda(s)+\alpha(\mathbf{Z}_j)\int_a^t\frac{w_i(s)}{p^2(s)}d\Lambda(s)\Big\}/2\sigma^3+\frac{2\alpha(\mathbf{Z}_i)\alpha(\mathbf{Z}_j)}{\sigma},$$

$$\gamma_V(\mathbf{Z}_i,\mathbf{Z}_j,\mathbf{Z}_k)$$
$$=\gamma(\mathbf{Z}_i,\mathbf{Z}_j,\mathbf{Z}_k)/\sigma+\sum_\pi\alpha(\mathbf{Z}_{\pi(i)})\Big\{\int_a^t(w_{\pi(j)}(s)/p^3(s))dM_{\pi(k)}(s)/\sigma^3$$
$$-\int_a^t\frac{w_{\pi(j)}(s)w_{\pi(k)}(s)}{2\sigma^3p^3(s)}d\Lambda(s)+\frac{3}{8}\int_a^t\int_a^t\frac{dM_{\pi(j)}(s)dM_{\pi(k)}(u)}{\sigma^4p^2(s)p^2(u)}$$
$$+\frac{3}{8}\int_a^t\int_a^t\frac{w_{\pi(j)}(s)w_{\pi(k)}(u)}{\sigma^4p^2(s)p^2(u)}d\Lambda(s)d\Lambda(u)$$
$$-\frac{3}{4\sigma^3}\Big(\int_a^t\frac{dM_{\pi(j)}(s)}{p^2(s)}\Big)\Big(\int_a^t\frac{w_{\pi(k)}(u)}{p^2(u)}d\Lambda(u)\Big)$$
$$+\frac{\alpha(\mathbf{Z}_{\pi(j)})\alpha(\mathbf{Z}_{\pi(k)})}{2\sigma}+\frac{\beta(\mathbf{Z}_{\pi(j)},\mathbf{Z}_{\pi(k)})}{\sigma}\Big\}$$
$$-\frac{1}{2\sigma^3}\sum_\pi\Big\{\int_a^t\frac{dM_{\pi(i)}(s)}{p^2(s)}-\int_a^t\frac{w_{\pi(i)}(s)}{p^2(s)}d\Lambda(s)\Big\}$$
$$\times\Big\{\frac{1}{2}\beta(\mathbf{Z}_{\pi(j)},\mathbf{Z}_{\pi(k)})+\alpha(\mathbf{Z}_{\pi(j)})\alpha(\mathbf{Z}_{\pi(k)})\Big\},$$

where $\sum_\pi$ denotes summation over all six permutations of $\{i,j,k\}$.

References

Akritas, M. G. (1986). Bootstrapping the Kaplan-Meier estimator. *J. Amer. Statist. Assoc.* **81**, 1032-1038.

Akritas, M. G. (1994). Nearest neighbor estimation of a bivariate distribution under random censoring. *Ann. Statist.* **22**, 1299-1327.

Amemiya, T. (1985). *Advanced Econometrics.* Harvard Univ. Press.

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.

Babu, G. J. (1991). Asymptotic theory for estimators under random censorship. *Probab. Theory Related Fields* **90**, 275-290.

Bhattacharya, P. K., Chernoff, H. and Yang, S. S. (1983). Nonparametric estimation of the slope of a truncated regression. *Ann. Statist.* **11**, 505-514.

Devroye, L. (1986). *Non-uniform Random Variate Generation.* Springer-Verlag, New York.

Efron, B. (1981). Censored data and the bootstrap. *J. Amer. Statist. Assoc.* **76**, 312-319.

Efron, B. and Petrosian, V. (1994). Survival analysis of the gamma-ray burst data. *J. Amer. Statist. Assoc.* **89**, 452-462.

Goldberger, A. S. (1981). Linear regression after selection. *J. Econometrics* **15**, 357-366.

Gross, S. and Lai, T. L. (1994). Nonparametric estimation and regression analysis with left truncated and right censored data. To appear in *J. Amer. Statist. Assoc.*

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.* **16**, 927-953.

Horváth, L. and Yandell, B. S. (1987). Convergence rates for bootstrapped product-limit process. *Ann. Statist.* **15**, 1155-1173.

Hyde, J. (1977). Testing survival under right censoring and left truncation. *Biometrika* **64**, 225-230.

Kalbfleisch, J. D. and Lawless, J. F. (1989). Inference based on retrospective ascertainment: An analysis of data on transfusion-related AIDS. *J. Amer. Statist. Assoc.* **84**, 360-372.

Keiding, N. and Gill, R. D. (1990). Random truncation models and Markov processes. *Ann. Statist.* **18**, 582-602.

Keiding, N., Holst, C. and Green, A. (1989). Retrospective estimation of diabetes incidence from information in a current prevalent population and historical mortality. *Amer. J. Epidemiol.* **130**, 588-600.

Lai, T. L. and Wang, J. Q. (1993). Edgeworth expansions for symmetric statistics with applications to bootstrap methods. *Statist. Sinica* **3**, 517-542.

Lai, T. L. and Ying, Z. (1991). Estimating a distribution function with truncated and censored data. *Ann. Statist.* **19**, 417-442.

Lo, S. H. and Singh, K. (1986). The product-limit estimator and the bootstrap: Some asymptotic representations. *Probab. Theory Related Fields* **71**, 455-465.

Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monthly Notices Roy. Astron. Soc.* **155**, 95-118.

Nicoll, J. F. and Segal, I. E. (1980). Nonparametric elimination of the observational cutoff bias. *Astron. Astrophys.* **82**, L3-L6.

Reid, N. (1981). Estimating the median survival time. *Biometrika* **68**, 601-608.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* John Wiley, New York.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica* **26**, 24-36.

Tsai, W. Y., Jewell, N. P. and Wang, M. C. (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika* **74**, 883-886.

Wang, M. C. (1991). Nonparametric estimation from cross-sectional survival data. *J. Amer. Statist. Assoc.* **86**, 130-143.

Woodroofe, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist.* **13**, 163-177.

Department of Statistics and Computer Information Systems, Baruch College, The City University of New York, New York, NY 10010, U.S.A.

Department of Statistics, Stanford University, Stanford, CA 94305, U.S.A.

(Received October 1994; accepted July 1995)