

Strong Representations of the Survival Function Estimator for Truncated and Censored Data with Applications

IRÈNE GIJBELS*

*Institut de Statistique/CORE, Université Catholique de Louvain,
Louvain-la-Neuve, Belgium*

AND

JANE-LING WANG[†]

Division of Statistics, University of California at Davis

A strong i.i.d. representation is obtained for the product-limit estimator of the survival function based on left truncated and right censored data. This extends the result of Chao and Lo (1988, *Ann. Statist.* **16**, 661–668) for truncated data. An improved rate of the approximation is also obtained on compact sets. Applications include density and hazard rate estimation. The advantage of the improved rate of the approximation is illustrated via kernel density estimation. © 1993 Academic Press, Inc.

1. INTRODUCTION AND MAIN RESULTS

In many survival studies a subject may not be included in the study if the time origin of its lifetime, called onset time, precedes the starting time of the study. Such subjects are called left truncated. Once entered into the study the individuals are subject to the usual right censoring. In this article we study survival data which are subject to both left truncation and right censoring (LTRC). More specifically, let (X, T, C) denote random variables where X is the variable of interest, the lifetime, with continuous distribution

Received September 12, 1991; revised February 17, 1993.

AMS 1980 subject classifications: 62G05; 62E20.

Key words and phrases: asymptotic normality, Bernstein's inequality, cumulative hazard function, kernel density estimator, product-limit estimator, strong i.i.d. representation.

* Supported by the National Fund for Scientific Research Belgium, as a Senior Research Assistant.

[†] Research supported in part by the Air Force Grant AFOSR-89-0386.

function (d.f.) F ; T is the random left truncation time with arbitrary d.f. G and C is the random right censoring time with arbitrary d.f. L . It is assumed that X is independent of (T, C) but T and C may be dependent. In the random LTRC model one observes (Y, T, δ) if $Y \geq T$ where $Y = X \wedge C = \min(X, C)$ and $\delta = 1(X \leq C)$ is the indicator of censoring status. When $Y < T$ nothing is observed. Let $\alpha \equiv P(T \leq Y)$, and W denotes the d.f. of Y , i.e., $1 - W = (1 - F)(1 - L)$. Obviously one needs to assume $\alpha > 0$. Let (Y_i, T_i, δ_i) , $i = 1, \dots, n$ be an independent and identically distributed (i.i.d.) sample of (Y, T, δ) which one observes (i.e., $Y_i \geq T_i$). Let

$$C(z) = P(T \leq z \leq Y | T \leq Y) = \alpha^{-1} P(T \leq z \leq C) [1 - F(z-)], \quad (1.1)$$

and consider its empirical estimate

$$C_n(z) = n^{-1} \sum_{i=1}^n 1(T_i \leq z \leq Y_i). \quad (1.2)$$

The product-limit estimator (PLE) \hat{F}_n of F is defined in Tsai *et al.* (1987) as

$$1 - \hat{F}_n(x) = \prod_{Y_i \leq x} (1 - [nC_n(Y_i)]^{-1})^{\delta_i}. \quad (1.3)$$

Note that \hat{F}_n reduces to the Kaplan–Meier (1958) PLE when there is no left truncation ($T \equiv 0$), and to the Lynden–Bell (1971) PLE when there is no right censoring. The technical study of \hat{F}_n for LTRC data is closely related to that of \hat{F}_n for left truncated data, but now one has to account for the right censoring effect (cf. Tsai *et al.*, 1987; Gu and Lai, 1990; Lai and Ying, 1991). Properties of the Lynden–Bell PLE for left truncated data were investigated by Woodroffe (1985), Wang *et al.* (1986) and Chao and Lo (1988). Woodroffe (1985) pointed out that F is identifiable only if some conditions on the support of F and G are satisfied. In our current setting these identifiability conditions involve the support of W and G instead of that of F and G . For any d.f. K denote the left and right endpoints of its support by $a_K = \inf\{t: K(t) > 0\}$ and $b_K = \inf\{t: K(t) = 1\}$, respectively. Then F is identifiable if

$$a_G \leq a_W \quad \text{and} \quad b_G \leq b_W. \quad (1.4)$$

In this article we assume that (1.4) holds.

In the right censorship model it is difficult to handle the upper tail of F . The left truncation creates further complications so that both the upper and lower tails of F are affected under the LTRC model. For right censored data, Lo and Singh (1986) decompose the Kaplan–Meier PLE as a mean of i.i.d. random variables plus a negligible remainder term of the order

$O((n^{-1} \ln n)^{3/4})$ a.s. uniformly over compact intervals. Such an order suffices for proving most of the asymptotic properties of F , but falls short when studying estimators of the density or hazard rate function of F . The order of the remainder term is later improved to be $O(n^{-1} \ln n)$ a.s. by Burke *et al.* (1988), Major and Rejtő (1988), Lo *et al.* (1989), and Gu (1991). For left truncated data, Chao and Lo (1989) gave a similar i.i.d. strong representation for the Lynden–Bell PLE on a compact interval excluding b_F : the order of the remainder term is $O((n^{-1} \ln n)^{3/4})$ a.s. when $a_G < a_F$ and only $o(n^{-1/2})$ a.s. when $a_G = a_F$. It is the goal of this paper to establish an i.i.d. representation for the PLE \hat{F}_n in (1.3) for LTRC data and to obtain the order of the remainder term. In particular, when $a_G < a_W$ we show in Theorem 1 that the order of the remainder term is $O(n^{-1} \ln n)$ a.s., an improvement over Theorem 1 in Chao and Lo (1988). Moreover, we give a tail probability bound for the remainder term. Another strong approximation on the whole support, for a modified version of the estimator (1.3) and for independent truncation time and censoring time, is established in Gu and Lai (1990, Thm. 2). Their method, based on martingale and stochastic integral approaches, is different from ours. Although the results are not directly comparable, our rate for the remainder term seems to be better for the case $a_G < a_W$, whereas their results also cover the case $a_G = a_W$.

We now introduce some notations before we state the strong representation results. Assume without loss of generality that Y and T are non-negative random variables. Define

$$W_1(y) = P(Y \leq y, \delta = 1 \mid T \leq Y),$$

$$W_{1n}(y) = n^{-1} \sum_{i=1}^n 1(Y_i \leq y, \delta_i = 1),$$

and let $A(z) = \int_0^z dF(u)/[1 - F(u-)]$ denote the cumulative hazard function of F . It can be shown that

$$dW_1(y) = \alpha^{-1} P(T \leq y \leq C) dF(y),$$

and

$$A(z) = \int_0^z dW_1(u)/C(u), \quad (1.5)$$

where C is defined in (1.1). Hence a natural estimator of $A(z)$ is

$$\hat{A}_n(z) = \int_0^z dW_{1n}(u)/C_n(u) = \sum_{i=1}^n 1(Y_i \leq z, \delta_i = 1)/[nC_n(Y_i)], \quad (1.6)$$

which is comparable to the Nelson–Aalen estimator of the cumulative hazard function for right censored data. Moreover, \hat{A}_n is the cumulative hazard function of the PLE \hat{F}_n defined in (1.3). Define

$$\zeta(y, t, \delta, z) \equiv 1(y \leq z, \delta = 1)/C(y) - \int_0^z [1(t \leq u \leq y)/C^2(u)] dW_1(u), \quad (1.7)$$

$$\zeta_a(y, t, \delta, z) \equiv \zeta(y, t, \delta, z) - \zeta(y, t, \delta, a). \quad (1.8)$$

Then $E(\zeta(Y_i, T_i, \delta_i, z)) = E(\zeta_a(Y_i, T_i, \delta_i, z)) = 0$, and

$$\begin{aligned} \Gamma(z_1, z_2) &\equiv \text{Cov}(\zeta_a(Y, T, \delta, z_1), \zeta_a(Y, T, \delta, z_2)) \\ &= \int_a^{z_1 \wedge z_2} [C(u)]^{-2} dW_1(u). \end{aligned} \quad (1.9)$$

The following theorem is an extension of Theorem 1 of Chao and Lo (1988) to the LTRC model. It is also an improvement of that theorem since the order of the remainder term in our theorem is $O(n^{-1} \ln n)$ a.s. as compared to $O((n^{-1} \ln n)^{3/4})$ a.s. in theirs.

THEOREM 1. For $a_G < a \leq z \leq b < b_W$, it follows that $C(z) \geq \varepsilon$ for some $\varepsilon > 0$, and

$$\begin{aligned} \text{(a)} \quad & [\hat{A}_n(z) - A(z)] - [\hat{A}_n(a) - A(a)] \\ &= \int_a^z [C(u)]^{-1} dW_{1n}(u) \\ &\quad - \int_a^z [C_n(u)/C^2(u)] dW_1(u) + R'_{na}(z), \\ &= n^{-1} \sum_{i=1}^n \zeta_a(Y_i, T_i, \delta_i, z) + R'_{na}(z), \end{aligned} \quad (1.10)$$

where

$$\sup_{a \leq z \leq b} |R'_{na}(z)| = O(n^{-1} \ln n) \quad \text{a.s.} \quad (1.11)$$

(b) If $a_G < a_W$ then (a) reduces to

$$\hat{A}_n(z) - A(z) = n^{-1} \sum_{i=1}^n \zeta(Y_i, T_i, \delta_i, z) + R'_n(z), \quad (1.12)$$

where

$$\sup_{0 \leq z \leq b} |R'_n(z)| = O(n^{-1} \ln n) \quad \text{a.s.} \quad (1.13)$$

(c) If $a_G < a_W$ then

$$\hat{F}_n(z) - F(z) = [1 - F(z)] n^{-1} \sum_{i=1}^n \zeta(Y_i, T_i, \delta_i, z) + R_n(z), \quad (1.14)$$

where

$$P\left(\sup_{0 \leq z \leq b} n |R_n(z)| > x + 4\epsilon^{-2}\right) \leq K[e^{-\lambda x} + (x/50)^{-2n} + e^{-\lambda x^3}], \quad (1.15)$$

with some $\lambda > 0$, and this implies that

$$\sup_{0 \leq z \leq b} |R_n(z)| = O(n^{-1} \ln n) \quad a.s., \quad (1.16)$$

and

$$E\left(\sup_{0 \leq z \leq b} |R_n(z)|^\alpha\right) = O(n^{-\alpha}), \quad \text{for any } \alpha > 0. \quad (1.17)$$

The tail probability bound in (1.15) is similar to (1.9) in Major and Rejtő (1988). Theorem 1 can be used to derive asymptotic properties of \hat{F}_n and \hat{A}_n .

COROLLARY 1. Assume that $a_G < a_W$ and $b < b_W$.

(a) For $0 < z < b_W$, $\hat{A}_n(z) \rightarrow A(z)$ a.s.

(b) $\sup_{0 \leq z \leq b} |\hat{A}_n(z) - A(z)| = O((n^{-1} \ln \ln n)^{1/2})$ a.s.

(c) The process $n^{1/2}[\hat{A}_n(z) - A(z)]$ converges weakly on $D[0, b]$ to a mean zero Gaussian process with covariance structure $\Gamma(z_1, z_2)$, where $\Gamma(z_1, z_2)$ is given in (1.9).

(d) The conclusions in (a), (b), and (c) hold with \hat{A}_n , A , and $\Gamma(z_1, z_2)$ replaced by \hat{F}_n , F , and $[1 - F(z_1)][1 - F(z_2)] \Gamma(z_1, z_2)$, respectively.

A similar weak convergence result for $n^{1/2}(\hat{F}_n - F)$ was obtained in Tsai *et al.* (1987) under the assumptions that $P(T \leq C) = 1$ and that (T, C) has continuous distribution function.

In Section 2 we apply Theorem 1 to obtain strong i.i.d. representations for kernel estimators of the density function of F . The asymptotic properties of these estimators are shown via such strong representations.

The proof of Theorem 1 is given in Section 3. The improvement on the order of the remainder term in Theorem 1 relies on a key lemma, Lemma 1 below, which extends a result in Serfling (1980, Lemma B, p. 223) and is of independent interest itself. The proof of Lemma 1 is given in Section 3. Note that Lemma 1 can also be used to extend Lemma 6 of Major and Rejtő (1988). For a discussion on this fact, see the remark after the proof of Theorem 1 in Section 3.

LEMMA 1. Let (X_{i1}, \dots, X_{im}) , $i = 1, \dots, n$ be a random sample from a joint distribution function H whose j th marginal d.f. is denoted by F_j , for $1 \leq j \leq m$. Let $F_{nj}(t) = n^{-1} \sum_{i=1}^n (X_{ij} \leq t)$ be the corresponding j th marginal empirical d.f. where $1 \leq j \leq m$ and $h(X_{i1}, \dots, X_{im})$ be a real valued function with $E_H[(h(X_{i1}, \dots, X_{im}))^{2(p-1)}] < \infty$, for all $1 \leq i_1, \dots, i_m \leq n$, p an integer. Then

$$E_H \left\{ \left(\int \cdots \int h(x_1, \dots, x_m) \prod_{j=1}^m d[F_{nj}(x_j) - F_j(x_j)] \right)^p \right\} = O(n^{-\lfloor pm/2 \rfloor}), \quad (1.18)$$

where $\lfloor a \rfloor$ denotes the integer part of a .

Note. Although Lemma 1 holds for general m , we only need to apply it, in the proof of Theorem 1, for the case $m = 2$.

2. APPLICATION TO DENSITY ESTIMATION

The strong i.i.d. representations in Theorem 1 can be applied to density and hazard function estimation. For LTRC data, hazard function estimation is considered in Uzunoğullari and Wang (1990) using the Hajek projection method. We are not aware of any result on density estimation even for purely truncated data. We thus concentrate on density estimation only in this section and demonstrate how to apply the strong i.i.d. representation in Theorem 1 to this goal. The derivation is similar to that of Lo *et al.* (1989) and Müller and Wang (1990). Details are thus omitted. The need of a sharp tail probability bound for the remainder term as in Theorem 1(c) will become clear (cf. Remark 2 below).

Let F be absolutely continuous with density function f . Assume $a_G < a_W$ as in Theorem 1(c) and f is $p \geq 1$ times continuously differentiable at z with $f(z) > 0$, for $a_G < z < b_W$. We consider the following kernel density estimator of $f(z)$,

$$\hat{f}_n(z) = b_n^{-1} \int K((z-t)/b_n) d\hat{F}_n(t), \quad (2.1)$$

where K is a kernel function in $L^2[-1, 1]$ of bounded variation with support in $[-1, 1]$ and satisfying the moment condition

$$\begin{aligned} 1, & \quad j = 0 \\ \int x^j K(x) dx = 0, & \quad 1 \leq j < p \\ < \infty \text{ but nonzero,} & \quad j = p. \end{aligned} \quad (2.2)$$

Note that the kernel K depends on p implicitly. The bandwidth sequence $\{b_n\}$ satisfies the usual conditions

$$b_n \rightarrow 0 \quad \text{and} \quad nb_n \rightarrow \infty. \quad (2.3)$$

Theorem 1 and integration by parts yield the following a.s. representation of $\hat{f}_n(z)$,

$$\hat{f}_n(z) = f(z) + \beta_n(z) + \sigma_n(z) + e_n(z), \quad (2.4)$$

where

$$\beta_n(z) = b_n^{-1} \int F(z - b_n x) dK(x) - f(z) \quad (2.5)$$

is essentially the bias,

$$\sigma_n(z) = (nb_n)^{-1} \sum_{i=1}^n \int [1 - F(z - b_n x)] \zeta(Y_i, T_i, \delta_i, z - b_n x) dK(x) \quad (2.6)$$

is the stochastic component of $\hat{f}_n(z)$, and $e_n(z)$ is the error of the approximation satisfying

$$\sup_{0 \leq z \leq b} |e_n(z)| = O((\ln n)(nb_n)^{-1}) \quad \text{a.s.,}$$

and

$$E \sup_{0 \leq z \leq b} |e_n(z)|^\alpha = O((nb_n)^{-\alpha}), \quad \text{for any } \alpha > 0. \quad (2.7)$$

Remark. Note that the order of the error term in (2.7) is sharper than that of (3.6) in Lo *et al.* (1989) who only considered the censored data case and has an extra $\ln n$ term. This improvement allows us to use the usual bandwidths as in (2.3) compared to the stronger requirement $(nb_n)(\ln n)^2 \rightarrow \infty$ in their condition (b2).

Denote

$$B_p \equiv [(-1)^p/p!] \int x^p K(x) dx \quad \text{and} \quad V = \int_{-1}^1 K^2(x) dx. \quad (2.8)$$

Note that $E(\sigma_n(z)) = 0$ and $\zeta(Y_i, T_i, \delta_i, t)$ is a random variable uniformly bounded for $0 \leq t \leq b < b_w$. Standard calculations and (2.2), (2.4)–(2.8) yield the following bias and variance expressions:

$$\text{bias}(\hat{f}_n(z)) = b_n^p f^{(p)}(z) B_p + o(b_n^p) + O((nb_n)^{-1}), \quad (2.9)$$

$$\text{Var}(\hat{f}_n(z)) = (nb_n)^{-1} f(z)[1 - F(z)][C(z)]^{-1} V + o((nb_n)^{-1}). \quad (2.10)$$

Strong consistency and asymptotic normality of $\hat{f}_n(z)$ thus follow by verifying the standard conditions for triangular arrays.

THEOREM 2. Let $\sigma^2 = f(z)[1 - F(z)][C(z)]^{-1} V$ and $d = \lim_{n \rightarrow \infty} nb_n^{2p+1}$. Then

- (a) $\hat{f}_n(z) \rightarrow f(z)$ a.s.
- (b) $(nb_n)^{1/2} \{ \hat{f}_n(z) - E(\hat{f}_n(z)) \} \xrightarrow{d} N(0, \sigma^2)$
- (c) $(nb_n)^{1/2} \{ \hat{f}_n(z) - f(z) \} \xrightarrow{d} N(d^{1/2} f^{(p)}(z) B_p, \sigma^2)$, if $d < \infty$.

Remarks. 1. Note that only the continuity of f at z suffices for Theorem 3(a).

2. If instead of (1.17) one has, e.g.,

$$E(R_n(z)^x) = O(n^{-\delta x}), \quad \text{for some } 1/2 < \delta < 1,$$

then the nb_n term in (2.9) is replaced by $n^\delta b_n$ and an extra term $O((n^\delta + 1/2 b_n^{3/2})^{-1}) + O((n^\delta b_n)^{-2})$ should be added to the right hand side of (2.10). Consequently, an extra condition $n^{2\delta-1} b_n \rightarrow \infty$ needs to be imposed in Theorem 3.

3. The choice of local optimal bandwidths and locally adaptive bandwidths can be carried out as in Müller and Wang (1990). We do not pursue this issue further. The above arguments should suffice to demonstrate the advantage of obtaining the improved order of convergence for the remainder term in the strong i.i.d. representation.

3. PROOFS OF THE MAIN RESULTS

The outline of the proof of Theorem 1 is similar to that of Theorem 1 of Major and Rejtő (1988), hereafter abbreviated as M & R (1988). We adopt their notations for the remainder terms, as closely as possible.

Proof of Theorem 1. Recall the definition of $\hat{A}_n(z)$ from (1.6). Using the arguments as on p. 1122 of M & R (1988), we obtain that

$$\hat{A}_n(z) - \hat{A}_n(a) = \sum_{i=1}^n \frac{1(a < Y_i \leq z, \delta_i = 1)}{nC(Y_i)} \left[1 - \frac{nC_n(Y_i) - nC(Y_i)}{nC(Y_i)} \right] + R_{3n}(z),$$

where

$$|R_{3n}(z)| \leq 2 \sum_{i=1}^n \left[\frac{C_n(Y_i) - C(Y_i)}{C(Y_i)} \right]^2 \frac{1(a < Y_i \leq z, \delta_i = 1)}{nC(Y_i)}$$

on the set $A_n(z) = \bigcap_{i=1}^n \{ |C_n(Y_i) - C(Y_i)| < \frac{1}{2}C(Y_i) \text{ or } Y_i \notin (a, z] \}$. Hence

$$\hat{A}_n(z) - \hat{A}_n(a) = 2\bar{A}(z) - \bar{B}(z) + R_{3n}(z), \quad (3.1)$$

where

$$\bar{A}(z) = \sum_{i=1}^n 1(a < Y_i \leq z, \delta_i = 1) [nC(Y_i)]^{-1} = \int_a^z [C(u)]^{-1} dW_{1n}(u) \quad (3.2)$$

$$\begin{aligned} \bar{B}(z) &= \sum_{i=1}^n 1(a < Y_i \leq z, \delta_i = 1) nC_n(Y_i) [nC(Y_i)]^{-2} \\ &= \int_0^\infty \int_0^\infty \frac{1(a < x \leq z)}{C^2(x)} 1(y \leq x) dC_n(y) dW_{1n}(x) \\ &= \int_a^z \int_0^x [C(x)]^{-2} dC_n(y) dW_{1n}(x). \end{aligned} \quad (3.3)$$

Decompose

$$\begin{aligned} dC_n(y) dW_{1n}(x) &= dC(y) d[W_{1n}(x) - W_1(x)] + d[C_n(y) - C(y)] dW_1(x) \\ &\quad + d[C_n(y) - C(y)] d[W_{1n}(x) - W_1(x)] + dC(y) dW_1(x). \end{aligned}$$

As in (2.5), p. 1123 of M & R (1988) we obtain from (3.3) that

$$\bar{B}(z) = B_1(z) + B_2(z) + B_3(z) + B_4(z), \quad (3.4)$$

with

$$B_1(z) + B_4(z) = \bar{A}(z). \quad (3.5)$$

$$B_2(z) = \int_a^z [C_n(x) - C(x)]/C^2(x) dW_1(x), \quad (3.6)$$

$$\begin{aligned} B_3(z) &= \int_0^\infty \int_0^\infty \frac{1(a < x \leq z)}{C^2(x)} 1(y \leq x) \\ &\quad \times d[W_{1n}(x) - W_1(x)] d[C_n(y) - C(y)]. \end{aligned} \quad (3.7)$$

Put

$$R_{4n}(z) = -B_3(z). \quad (3.8)$$

It follows from (3.1), (3.4)–(3.8) that

$$\hat{A}_n(z) - \hat{A}_n(a) = \bar{A}(z) - B_2(z) + R_{3n}(z) + R_{4n}(z). \quad (3.9)$$

Further, (3.6) and (1.5) imply that

$$B_2(z) = \int_a^z [C_n(x)/C^2(x)] dW_1(x) - [\Lambda(z) - \Lambda(a)]. \quad (3.10)$$

Hence from (3.2), (3.9), and (3.10) we obtain (1.10) with

$$R'_{na}(z) = R_{3n}(z) + R_{4n}(z).$$

We deal with this remainder term by proving in the Appendix that

LEMMA 2. *There exist $K, \alpha > 0$ such that*

$$P\left(\sup_{a \leq z \leq b} n |R_{3n}(z)| > x\right) \leq 2ne^{-\alpha n} + Ke^{-\alpha x}.$$

LEMMA 3. *For some $K, \lambda > 0$,*

$$P\left(\sup_{a \leq z \leq b} n |R_{4n}(z)| > x\right) \leq K[(x/5)^{-2n} + e^{-\lambda(nx)^{1/2}} + e^{-\lambda x} + e^{-\lambda nx^2}].$$

Taking $x = c \ln n$ in Lemma 2 and 3, with c large enough, and applying the Borel–Cantelli lemma we find that

$$\sup_{a \leq z \leq b} n |R_{in}(z)| = O(\ln n) \text{ a.s.,} \quad \text{for } i = 3, 4,$$

which leads to (1.11). This completes the proof of part (a).

Part (b) follows from (a) by choosing $a_G < a < a_W$.

We now prove part (c). A two-term Taylor expansion yields

$$\hat{F}_n(z) - F(z) = [1 - F(z)][-\log(1 - \hat{F}_n(z)) - \Lambda(z)] + R_{1n}(z), \quad (3.11)$$

where

$$\begin{aligned} |R_{1n}(z)| &\leq |\log(1 - \hat{F}_n(z)) + \Lambda(z)|^2, \\ \text{if } |\log(1 - \hat{F}_n(z)) + \Lambda(z)| &\leq 1. \end{aligned} \quad (3.12)$$

Further, we write

$$\begin{aligned} R_{2n}(z) &= \log(1 - \hat{F}_n(z)) + \hat{\Lambda}_n(z) \\ &= \sum_{i=1}^n 1(Y_i \leq z, \delta_i = 1) \{ \log(1 - [nC_n(Y_i)]^{-1}) + [nC_n(Y_i)]^{-1} \}. \end{aligned}$$

Hence (1.14) follows with

$$R_n(z) = [1 - F(z)][-R_{2n}(z) + R_{3n}(z) + R_{4n}(z)] + R_{1n}(z). \quad (3.13)$$

Recall that $C(z) \geq \varepsilon > 0$, for $z \leq b$. For the remainder term $R_{2n}(z)$ we prove in the Appendix that

LEMMA 4. *For some $K, \lambda > 0$.*

$$P(\sup_{z \leq b} n |R_{2n}(z)| > 4\varepsilon^{-2}) \leq K e^{-\lambda n}.$$

In order to deal with the remainder term $R_{1n}(z)$, we show in the Appendix that:

LEMMA 5. *For $400[n\varepsilon^4]^{-1} < x < 2n$ and some $K, \lambda > 0$,*

$$P(\sup_{z \leq b} n |\log(1 - \hat{F}_n(z)) + A(z)|^2 > x) \leq K \{e^{-\lambda x} + [(nx)^{1/2}/25]^{-2n}\}.$$

From this we obtain,

$$P(\sup_{z \leq b} |\log(1 - \hat{F}_n(z)) + A(z)| > 1) \leq K \{e^{-\lambda n} + (n/25)^{-2n}\}. \quad (3.14)$$

Further, Lemma 5, (3.12) and (3.14) yield, for $400[n\varepsilon^4]^{-1} < x < 2n$, and some $K, \lambda > 0$,

$$P(\sup_{z \leq b} |nR_{1n}(z)| > x/2) \leq K \{e^{-\lambda x} + [(nx/2)^{1/2}/25]^{-2n} + (x/50)^{-2n}\}. \quad (3.15)$$

Applying Lemmas 2, 3, and 4 with $a_G < a < a_W$, it follows from (3.13) and (3.15) that for $400[n\varepsilon^4]^{-1} < x < 2n$,

$$\begin{aligned} P(\sup_{z \leq b} n |R_n(z)| > x + 4\varepsilon^{-2}) &\leq P(\sup_{z \leq b} n |R_{1n}(z)| > x/2) \\ &\quad + P(\sup_{z \leq b} n |R_{2n}(z)| > 4\varepsilon^{-2}) \\ &\quad + P(\sup_{z \leq b} n |R_{3n}(z)| > x/4) \\ &\quad + P(\sup_{z \leq b} n |R_{4n}(z)| > x/4) \\ &\leq K [e^{-\lambda x} + [(nx/2)^{1/2}/25]^{-2n} \\ &\quad + (x/50)^{-2n} + (x/20)^{-2n} + e^{-\lambda x^3}] \\ &\leq K [e^{-\lambda x} + (x/50)^{-2n} + e^{-\lambda x^3}]. \end{aligned} \quad (3.16)$$

For $x \leq 400[n\varepsilon^4]^{-1}$, we can choose K large enough in (3.16) so that $Ke^{-\lambda x} \geq 1$. Hence (3.16) holds for $0 \leq x < 2n$. The proof of (1.15) will be

completed if we show that (3.16) holds for $x \geq 2n$. First of all note that (1.14) implies

$$\begin{aligned} \hat{F}_n(z) - F(z) &= [1 - F(z)] \\ &\times \left\{ \int_{aw}^z [C(u)]^{-1} d[W_{1n}(u) - W_1(u)] - B_2(z) \right\} + R_n(z), \end{aligned}$$

and that $|\hat{F}_n(z) - F(z)| \leq 1$. Hence, $|nR_n(z)| > x$ implies

$$n \left| \int_{aw}^z [C(u)]^{-1} d[W_{1n}(u) - W_1(u)] - B_2(z) \right| > x/2 \quad \text{for } x \geq 2n.$$

Therefore

$$\begin{aligned} P(\sup_{z \leq b} n |R_n(z)| > x) &\leq P \left\{ \sup_{z \leq b} n \left| \int_{aw}^z [C(u)]^{-1} d[W_{1n}(u) - W_1(u)] \right| > x/4 \right\} \\ &\quad + P \{ \sup_{z \leq b} n |B_2(z)| > x/4 \} \\ &\leq P(\sup_{z \leq b} n |W_{1n}(z) - W_1(z)| > \varepsilon x/4) \\ &\quad + P(\sup_{z \leq b} n |C_n(z) - C(z)| > \varepsilon^2 x/4) \\ &\leq Ke^{-\lambda x^2/n} \leq Ke^{-2\lambda x}, \quad \text{for } x \geq 2n, \end{aligned}$$

by Lemma 2 of Dvoretzky *et al.* (1956). Hence (1.15) holds for all x .

Taking $x = c \ln n$ in (1.15), with c large enough, (1.16) follows from the Borel–Cantelli lemma.

Finally, (1.17) follows from (1.15) and the fact that

$$E(|X|^\alpha) = \alpha \int_0^\infty u^{\alpha-1} P\{|X| \geq u\} du.$$

This finishes the proof of Theorem 1. ■

Remark. Lemma 3 holds the key to the improvement of Theorem 1 in Chao and Lo (1988) and it plays the role of Lemma 3 in M & R (1988). The proof of Lemma 3 in M & R relies on their Lemma 6 which is a special case of Theorem 1 of Major (1988). This theorem cannot be applied in our situation, and its role is taken over by Lemma 1 in Section 1 (c.f. the proof of Lemma 3 in the Appendix).

Before proving Lemma 1, we first extend Lemma A on p. 222 of Serfling (1980).

LEMMA A. Let F_1, \dots, F_m and G_1, \dots, G_m be d.f.'s and $h(x_1, \dots, x_m)$ be a given real valued function. Then

$$\begin{aligned} & \int \cdots \int h(x_1, \dots, x_m) \prod_{i=1}^m d[G_i(x_i) - F_i(x_i)] \\ &= \int \cdots \int \tilde{h}(x_1, \dots, x_m) \prod_{i=1}^m dG_i(x_i), \end{aligned}$$

where the function \tilde{h} depends on F_1, \dots, F_m and

$$\int \tilde{h}(x_1, \dots, x_m) dF_i(x_i) = 0, \quad \text{for } 1 \leq i \leq m. \quad (3.17)$$

Proof. For $m=1$, take $\tilde{h}(x) = h(x) - \int h(x_1) dF_1(x_1)$.

For $m=2$, take $\tilde{h}(x_1, x_2) = h(x_1, x_2) - \int h(x_1, x_2) dF_1(x_1)$

$$- \int h(x_1, x_2) dF_2(x_2) + \iint h(x_1, x_2) dF_1(x_1) dF_2(x_2).$$

In general, take

$$\begin{aligned} \tilde{h}(x_1, \dots, x_m) &= h(x_1, \dots, x_m) - \sum_{i=1}^m \int h(x_1, \dots, x_m) dF_i(x_i) \\ &+ \sum_{1 \leq i < j \leq m} \iint h(x_1, \dots, x_m) dF_i(x_i) dF_j(x_j) - \cdots \\ &+ (-1)^m \int \cdots \int h(x_1, \dots, x_m) dF_1(x_1) \cdots dF_m(x_m). \quad \blacksquare \end{aligned}$$

Proof of Lemma 1. Choosing $G_i = F_{ni}$ in Lemma A, we obtain

$$\begin{aligned} & \int \cdots \int h(x_1, \dots, x_m) \prod_{i=1}^m d[F_{ni}(x_i) - F_i(x_i)] \\ &= \int \cdots \int \tilde{h}(x_1, \dots, x_m) \prod_{i=1}^m dF_{ni}(x_i) \\ &= n^{-m} \sum_{i_1=1}^n \cdots \sum_{i_m=1}^n \tilde{h}(X_{i_1,1} \cdots X_{i_m,m}). \end{aligned}$$

The left hand side of (1.18) equals

$$n^{-pm} \sum_{i_{11}=1}^n \cdots \sum_{i_{m1}=1}^n \cdots \sum_{i_{1p}=1}^n \cdots \sum_{i_{mp}=1}^n E_H [\tilde{h}(X_{i_{11},1}, \dots, X_{i_{m1},m}) \cdots \tilde{h}(X_{i_{1p},1}, \dots, X_{i_{mp},m})]. \quad (3.18)$$

Using the fact that $X_{i_{hl},j}$ and $X_{i_{gv},k}$ are independent if $i_{hl} \neq i_{gv}$, and (3.17), it can be shown (cf. p. 224, proof of Lemma B in Serfling (1980)) that a term in (3.18) may be possibly nonzero only if the sequence of indices $i_{11}, \dots, i_{m1}; \dots; i_{1p}, \dots, i_{mp}$ contains each member at least twice. The number of such cases is $O(n^{\lceil pm/2 \rceil})$. The lemma now follows from (3.18) and Holder's inequality. ■

APPENDIX: PROOFS OF LEMMAS 2-5

Let $W^*(t) = P(Y \leq t | T \leq Y)$ be the marginal distribution of the observed Y .

Proof of Lemma 2. The proof is similar to that of Lemma 2 in M & R (1988) and we omit the details. First of all note that $A_n(z_2) \subset A_n(z_1)$, for $z_1 \leq z_2$. Therefore, we consider

$$P(\sup_{a \leq z \leq b} n |R_{3n}(z)| > x) \leq P(\sup_{a \leq z \leq b} n |R_{3n}(z)| > x, A_n(b)) + P((A_n(b))^c).$$

For the second probability, we obtain

$$\begin{aligned} P((A_n(b))^c) &\leq \sum_{i=1}^n P \left\{ |C_n(Y_i) - C(Y_i)| \geq \frac{1}{2} C(Y_i), a < Y_i \leq b \right\} \\ &= \sum_{i=1}^n \int_a^b P \left(\left| \frac{C_n(Y_i) - C(Y_i)}{C(Y_i)} \right| \geq \frac{1}{2} \mid Y_i = t \right) dW^*(t) \\ &\leq n \int_a^b 2e^{-\lambda n C(t)} dW^*(t) \\ &\leq 2ne^{-\alpha n}, \quad \text{for some } \alpha > 0 \text{ since } C(t) \geq \varepsilon, \end{aligned} \quad (A.1)$$

where the second to the last inequality follows from Bernsteins' inequality (cf. p. 95 in Serfling (1980)).

The first probability is bounded by

$$\begin{aligned}
 & P\left(\sup_{a \leq z \leq b} 2n \sum_{i=1}^n \left[\frac{C_n(Y_i) - C(Y_i)}{C(Y_i)} \right]^2 \frac{1(a < Y_i \leq z, \delta_i = 1)}{nC(Y_i)} > x\right) \\
 & \leq P\left(2 \sum_{i=1}^n \frac{1(a < Y_i \leq b)}{\varepsilon} \left[\frac{C_n(Y_i) - C(Y_i)}{\varepsilon} \right]^2 > x\right) \\
 & \leq P(n \sup_{-\infty < t < \infty} |C_n(t) - C(t)|^2 > x\varepsilon^3/2) \\
 & \leq K_1 e^{-x\varepsilon^3/4},
 \end{aligned} \tag{A.2}$$

by Lemma 2 of Dvoretzky *et al.* (1956).

The result now follows from (A.1) and (A.2). ■

Proof of Lemma 3. We first transform the remainder term $R_{4n}(z)$. Define $W_2(y) = P(Y \leq y, \delta = 0 | T \leq Y)$. Hence $W^*(y) = W_1(y) + W_2(y)$. Consider the following transformation

$$V_i = \begin{cases} W_1(Y_i), & \text{if } \delta_i = 1 \\ 1 - W_2(Y_i), & \text{if } \delta_i = 0. \end{cases}$$

Then V_1, \dots, V_n are i.i.d. uniform $[0, 1]$ random variables. Let $U_n(x) = n^{-1} \sum_{i=1}^n 1(V_i \leq x)$, be the corresponding empirical distribution function. Then $W_{1n}(x) = U_n(W_1(x))$ for $W_1(x) \leq W_1(\infty)$. Define $W_1^{-1}(t) = \inf\{x: W_1(x) \geq t\}$.

Then $W_1^{-1}W_1(x) = x$ since W_1 is continuous (due to the fact that F is continuous). Then

$$\begin{aligned}
 -R_{4n}(z) &= B_3(z) = \int_0^\infty \int_0^{W_1(\infty)} \frac{1(a < W_1^{-1}(t) \leq z)}{C^2(W_1^{-1}(t))} \\
 &\quad \times 1(y \leq W_1^{-1}(t)) d[U_n(t) - U(t)] d[C_n(y) - C(y)] \\
 &= \int_0^\infty \int_0^\infty \frac{1(W_1(a) < t \leq W_1(z))}{C^2(W_1^{-1}(t))} \\
 &\quad \times 1(W_1(y) \leq t) d[U_n(t) - U(t)] d[C_n(y) - C(y)] \\
 &= \int_0^\infty \int_0^\infty h(t, y) d[U_n(t) - U(t)] d[C_n(y) - C(y)],
 \end{aligned}$$

where $h(t, y) \leq \varepsilon^{-2}$. Since C_n is the difference of two empirical distribution functions, an application of Lemma 1 with $m = 2$ leads to

$$E([B_3(z)]^{-2p}) = O(n^{-2p}),$$

where the bound is uniform in z . The Markov inequality then implies that for some $K > 0$ (independent of z),

$$P(n |B_3(z)| > x) \leq K(x/2)^{-2p}. \quad (\text{A.3})$$

Next we aim at extending (A.3) in the sense of replacing $B_3(z)$ by $\sup_{a \leq z \leq b} |B_3(z)|$. This can be achieved by using standard discretizing arguments. By continuity of W_1 it is possible to partition $[a, b]$ as $a = t_1 < t_2 < \dots < t_n = b$ such that

$$W_1(t_{i+1}) - W_1(t_i) \leq n^{-1} \quad \text{for } i = 1, \dots, n-1.$$

From (3.7)

$$B_3(z) = \int_a^z \frac{C_n(y) - C(y)}{C^2(y)} d[W_{1n}(y) - W_1(y)].$$

Hence

$$\begin{aligned} P\left\{ \sup_{a \leq z \leq b} n |B_3(z)| > x \right\} &\leq P\left\{ \max_{1 \leq i \leq n-1} n |B_3(t_i)| > (x/2) \right\} \\ &\quad + P\left\{ \max_{1 \leq i \leq n-1} \sup_{t_i \leq t \leq t_{i+1}} n \left| \int_{t_i}^t \frac{C_n(y) - C(y)}{C^2(y)} \right. \right. \\ &\quad \left. \left. \times d[W_{1n}(y) - W_1(y)] \right| > (x/2) \right\} \\ &\leq (n-1) K(x/4)^{-2p} + P(nV_n > x/2), \end{aligned} \quad (\text{A.4})$$

for some $K > 0$, where

$$\begin{aligned} V_n &\leq \max_{1 \leq i \leq n-1} \sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t \frac{C_n(y) - C(y)}{C^2(y)} dW_{1n}(y) \right| \\ &\quad + \max_{1 \leq i \leq n-1} \sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t \frac{C_n(y) - C(y)}{C^2(y)} dW_1(y) \right| \\ &\leq \sup_{a \leq z \leq b} \left| \frac{C_n(z) - C(z)}{C^2(z)} \right| \left\{ \max_{1 \leq i \leq n-1} [W_{1n}(t_{i+1}) - W_{1n}(t_i)] \right. \\ &\quad \left. + \max_{1 \leq i \leq n-1} [W_1(t_{i+1}) - W_1(t_i)] \right\} \\ &\leq \sup_{a \leq z \leq b} \left| \frac{C_n(z) - C(z)}{C^2(z)} \right| \left\{ \max_{1 \leq i \leq n-1} | [W_{1n}(t_{i+1}) - W_1(t_{i+1})] \right. \\ &\quad \left. - [W_{1n}(t_i) - W_1(t_i)] | + 2 \max_{1 \leq i \leq n-1} [W_1(t_{i+1}) - W_1(t_i)] \right\}. \end{aligned} \quad (\text{A.5})$$

Now applying Bernstein's inequality (cf. p. 96 in Serfling (1980)), we have

$$\begin{aligned} P(n^{1/2} | [W_{1n}(t_{i+1}) - W_1(t_{i+1})] - [W_{1n}(t_i) - W_1(t_i)]| > x^{1/2}/4) \\ \leq 2e^{-\lambda(nx)^{1/2}}, \quad \text{for some } \lambda > 0. \end{aligned} \quad (\text{A.6})$$

Using Lemma 2 in Dvoretzky *et al.* (1956), we get

$$\begin{aligned} P\left(\sup_{a \leq z \leq b} n^{1/2} \left| \frac{C_n(z) - C(z)}{C^2(z)} \right| > x^{1/2}\right) \\ \leq P\left(\sup_{a \leq z \leq b} n^{1/2} |C_n(z) - C(z)| > \varepsilon^2 x^{1/2}\right) \\ \leq Ke^{-\lambda x}, \quad \text{for some } K, \lambda > 0. \end{aligned} \quad (\text{A.7})$$

Similarly,

$$P\left(\sup_{a \leq z \leq b} \left| \frac{C_n(z) - C(z)}{C^2(z)} \right| > x/8\right) \leq Ke^{-\lambda nx^2}, \quad \text{for some } K, \lambda > 0. \quad (\text{A.8})$$

From (A.5) – (A.8) we obtain

$$P(nV_n > x/2) \leq 2e^{-\lambda(nx)^{1/2}} + K_1 e^{-\lambda_2 x} + K_2 e^{-\lambda_3 nx^2}. \quad (\text{A.9})$$

Choosing $p = n$ in (A.4), it follows that $(n-1)K(x/4)^{-2p} \leq K(x/5)^{-2n}$. The lemma now follows from (A.4) and (A.9). ■

Proof of Lemma 4. First note that $|\log(1-x) + x| \leq x^2$, if $0 \leq x \leq \frac{1}{2}$. Hence on

$$\begin{aligned} E_n &= \bigcap_{i=1}^n \{nC_n(Y_i) \geq 2\}, |R_{2n}(z)| \\ &\leq \sum_{i=1}^n 1(Y_i \leq z, y_i = 1)[nC_n(Y_i)]^{-2}. \end{aligned} \quad (\text{A.10})$$

Further, $nC_n(Y_i) \geq 1$ for all i , and it follows that

$$\begin{aligned} P(E_n^c) &\leq \sum_{i=1}^n P(nC_n(Y_i) = 1) \\ &= \sum_{i=1}^n \int_{a_W}^b P(nC_n(Y_i) = 1 | Y_i = t) dW^*(t) \\ &= n \int_{a_W}^b [1 - C(t)]^{n-1} dW^*(t) \\ &\leq n(1 - \varepsilon)^{n-1} \leq Ke^{-\lambda n}, \end{aligned} \quad (\text{A.11})$$

for some $K, \lambda > 0$.

Next consider,

$$\begin{aligned}
& P\left(\sup_{z \leq b} n \left| \sum_{i=1}^n 1(Y_i \leq z, \delta_i = 1) [nC_n(Y_i)]^{-2} \right| > 4\varepsilon^{-2}\right) \\
& \leq P\left(n \sum_{i=1}^n 1(Y_i \leq b, \delta_i = 1) [nC_n(Y_i)]^{-2} > 4\varepsilon^{-2}\right) \\
& \leq P\left(n \sum_{i=1}^n 1(Y_i \leq b) [nC(Y_i) - n\varepsilon/2]^{-2} > 4\varepsilon^{-2}\right) \\
& \quad + P\left(\sup_{a \leq u \leq b} |C_n(u) - C(u)| > \frac{\varepsilon}{2}\right) \\
& \leq P\left(n \sum_{i=1}^n 1(Y_i \leq b) > n^2\right) + K_1 e^{-n\varepsilon^2/8} \\
& \leq P\left(\sum_{i=1}^n [1(Y_i \leq b) - W^*(b)] > n[1 - W^*(b)]\right) + K_1 e^{-n\varepsilon^2/8} \\
& \leq K e^{-\lambda n}, \tag{A.12}
\end{aligned}$$

for some $K, \lambda > 0$, by Bernstein's inequality. The lemma now follows from (A.10)–(A.12). ■

Proof of Lemma 5. From the definition of $R_{2n}(z)$ and (3.2), (3.6) and (3.9), we have

$$\begin{aligned}
& P(\sup_{z \leq b} n |\log(1 - \hat{F}_n(z)) + A(z)|^2 > x) \\
& \leq P\left(\sup_{z \leq b} n \left| \int_{a_W}^z [C(u)]^{-1} d[W_{1n}(u) - W_1(u)] \right| > \frac{1}{5}(nx)^{1/2}\right) \\
& \quad + P(\sup_{z \leq b} n |B_2(z)| > \frac{1}{5}(nx)^{1/2}) \\
& \quad + P(\sup_{z \leq b} n |R_{2n}(z)| > \frac{1}{5}(nx)^{1/2}) \\
& \quad + P(\sup_{z \leq b} n |R_{3n}(z)| > \frac{1}{5}(nx)^{1/2}) \\
& \quad + P(\sup_{z \leq b} n |R_{4n}(z)| > \frac{1}{5}(nx)^{1/2}) \\
& = \text{I} + \text{II} + \text{III} + \text{IV} + \text{V}. \tag{A.13}
\end{aligned}$$

Using the fact that $C(u) \geq \varepsilon$ and Bernstein's inequality,

$$\begin{aligned} \text{I} &\leq P(\sup_{z \leq b} n |W_{1n}(z) - W_1(z)| > \tfrac{1}{3}\varepsilon(nx)^{1/2}) \\ &\leq Ke^{-\lambda \varepsilon^2 x}, \quad \text{for some } K, \lambda > 0. \end{aligned} \quad (\text{A.14})$$

Lemma 2 of Dvoretzky *et al.* (1956) implies, for some $K, \lambda > 0$,

$$\begin{aligned} \text{II} &\leq P(\sup_{z \leq b} n |C_n(z) - C(z)| > \tfrac{1}{3}\varepsilon^2(nx)^{1/2}) \\ &\leq Ke^{-\lambda \varepsilon^4 nx} \leq Ke^{-\lambda \varepsilon^2 nx}. \end{aligned} \quad (\text{A.15})$$

Lemmas 2, 3, and 4 imply that for some $K, \lambda > 0$,

$$\begin{aligned} \text{III} &\leq Ke^{-\lambda n}, \quad \text{if } x > 400[ne^4]^{-1} \\ &\leq Ke^{-\lambda x/2}, \quad \text{if } x < 2n, \end{aligned} \quad (\text{A.16})$$

$$\text{IV} \leq Ke^{-\lambda(nx)^{1/2}} \leq Ke^{-\lambda x/2}, \quad \text{if } x < 2n, \quad \text{and} \quad (\text{A.17})$$

$$\begin{aligned} \text{V} &\leq K\{[(nx)^{1/2}/25]^{-2n} + e^{-\lambda n^{3/4}x^{1/4}} + e^{-\lambda(nx)^{1/2}} + e^{-\lambda n^2x}\} \\ &\leq K\{[(nx)^{1/2}/25]^{-2n} + e^{-\lambda(1/2)^{3/4}x} + e^{-\lambda x(1/2)} + e^{-\lambda n^2x}\} \\ &\leq K\{[(nx)^{1/2}/25]^{-2n} + e^{-\lambda x}\}. \end{aligned} \quad (\text{A.18})$$

The lemma now follows from (A.13)–(A.18). ■

ACKNOWLEDGMENT

The authors thank the referee for helpful comments.

REFERENCES

- [1] BURKE, M. D., CSÖRGÖ, S., AND HORVÁTH, L. (1988). A correction to and an improvement of strong approximations of some biometric estimates under random censorship. *Probab. Theory Related Fields* **79** 51–57.
- [2] CHAO, M. T., AND LO, S. H. (1988). Some representations of the nonparametric maximum likelihood estimators with truncated data. *Ann. Statist.* **16** 661–668.
- [3] DVORETZKY, A., KIEFER, J., AND WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.
- [4] GU, M. G. (1991). The Chung–Smirnov law for the product-limit estimator under random censorship. *Chinese Ann. Math. Ser. B* **12** 96–105.
- [5] GU, M. G., AND LAI, T. L. (1990). Functional laws of the iterated logarithm for the product-limit estimator of a distribution function under random censorship or truncation. *Ann. Probab.* **18** 160–189.

- [6] KAPLAN, E. L., AND MEIER, P. (1959). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.
- [7] LAI, T. L., AND YING, Z. (1991). Estimating a distribution function with truncated and censored data. *Ann. Statist.* **19** 417–442.
- [8] LO, S. H., MACK, Y. P., AND WANG, J. L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan–Meier estimator. *Probab. Theory Related Fields* **80** 461–473.
- [9] LO, S. H., AND SINGH, K. (1986). The product-limit estimator and the bootstrap: Some asymptotic representations. *Probab. Theory Related Fields* **71** 455–465.
- [10] LYNDEN-BELL, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monthly Notices Roy. Astronom. Soc.* **155** 95–188.
- [11] MAJOR, P. (1988). On the tail behaviour of the distribution function of multiple stochastic integrals. *Probab. Theory Related Fields* **78** 419–435.
- [12] MAJOR, P., AND REJTŐ, L. (1988). Strong embedding of the estimator of the distribution function under random censorship. *Ann. Statist.* **16** 1113–1132.
- [13] MÜLLER, H. G., AND WANG, J. L. (1990). Locally adaptive hazard smoothing. *Probab. Theory Related Fields* **85** 523–538.
- [14] SERFLING, R. (1980). *Approximation Theorem of Mathematical Statistics*. Wiley, New York.
- [15] 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.
- [16] UZUNOĞULLARI, U., AND WANG, J. L. (1991). “A Comparison of Hazard Rate Estimators for Left Truncated and Right Censored Data.” Technical Report, University of California, Davis.
- [17] WANG, M. C., JEWELL, N. P., AND TSAI, W. Y. (1986). Asymptotic properties of the product limit estimate under random truncation. *Ann. Statist.* **14** 1597–1605.
- [18] WOODROOFE, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist.* **13** 163–177.