The Annals of Statistics
1995, Vol. 23, No. 2, 440–449

ON STRONG UNIFORM CONSISTENCY OF THE LYNDEN-BELL ESTIMATOR FOR TRUNCATED DATA

BY KANI CHEN,[1] MIN-TE CHAO[2] AND SHAW-HWA LO[1]

Columbia University, Academia Sinica and Columbia University

In this paper, we prove that the Lynden-Bell estimator of a distribution function in the random truncation model is uniformly strong consistent over the whole half line, a problem left open by Woodroofe.

1. Introduction and main results. The purpose of the present article is to study the consistency property of the maximum likelihood estimator when data are randomly truncated. Let (X_i, Y_i), $i = 1, \dots, N$, be iid pairs of nonnegative random variables, with X_i and Y_i independent for each i. Let F and G denote the distribution functions of the X population and the Y population, respectively. For simplicity of presentation, we shall assume the continuity of F and G. In a truncation model, (X_i, Y_i) is observed only if $X_i \ge Y_i$. Based on n observations $\{(x_i, y_i); 1 \le i \le n\}$, one attempts to estimate F. This model has been considered by several authors [see Woodroofe (1985), Bhattacharya (1983), Bhattacharya, Chernoff and Yang (1983) and Chao and Lo (1988), among others]. The well known estimator of the survival function $\bar{F} = 1 - F$, due to Lynden-Bell (1971), can be expressed as follows:

$$\hat{\bar{F}}_n(x) = \prod \left(1 - \frac{r_i}{nC_n(x_i)}\right), \tag{1}$$

where $r_i = \#\{j \le n,\ x_j = x_i\}$, the product runs over all i such that $x_i \le x$ and $C_n(s) = (1/n)\#\{i \le n;\ y_i < s \le x_i\}$.

The motivation in deriving the estimator $\hat{\bar{F}}_n$ by Lynden-Bell (1971) is related to a model in astronomy which, as briefly explained in Woodroofe (1985), can be described as follows. The absolute and apparent luminosities of an astronomical object are defined to be its brightness at a fixed distance and as observed on Earth, and magnitude is defined to be the negative logarithm of luminosity. Since one can only observe those objects which are bright enough as observed on Earth, it is equivalent to state that in order to observe an astronomical object, the apparent magnitude must be small enough, say less than or equal to some constant a. It is well accepted in cosmology that apparent magnitude can be expressed as the sum of a function of redshift, denoted by X', and absolute magnitude, denoted by Y', and that the redshift and the absolute magnitude are assumed independent. If we let $X = -X'$

Received April 1993; revised September 1994.
[1]Research partially supported by AFOSR Grant F49620.
[2]Research partially supported by the National Science Council of the Republic of China.
AMS 1991 *subject classifications*. 62G05, 62E20.
Key words and phrases. Stopping time, n-symmetric, reverse submartingale.

and $Y = Y' - a$, the condition for the observability of a celestial object becomes $X \geq Y$. This problem with observational selection falls into the aforementioned framework of truncation model.

Another application arises from analysis of survival data of patients infected by the AIDS virus from contaminated blood transfusions [Lagakos, Barraj and De Gruttola (1988)]. An important feature of AIDS development is the induction period between infection with the AIDS virus and the onset of clinical AIDS. The data collected from persons infected from contaminated blood transfusions provide a unique source of information for the induction period. Of persons infected in this way, only those who have developed AIDS can be identified. Let Y denote the chronological time of infection and X^* denote the induction period and assume they are independent. Suppose that one can only observe a random sample of patients who are infected and develop AIDS in some chronological time interval $[0, \alpha]$. Let $X = \alpha - X^*$. Then the pair (X, Y) is observable if and only if $0 \leq Y \leq X$ $(\leq \alpha)$, that is, having left truncated observation of the reverse survival time X.

The large sample property of the product-limit estimator in the right censorship model has drawn much attention in the literature. Proofs of uniform consistency of the product-limit estimator on a compact interval or on the whole line can be found in many articles [e.g., Gill (1983), Lo and Singh (1986) and Wang (1987)]. In the truncation model, let us define $a_F = \inf\{x;\ F(x) > 0\}$ and $a_G = \inf\{x;\ G(x) > 0\}$. Woodroofe (1985) proved the weak consistency of $\hat{F}_n$ over $[0, \infty)$, as well as asymptotic normality under the additional condition that $\int_0^\infty dF/G < \infty$. It should be noticed that the only condition for the uniform weak consistency of $\hat{F}_n$ over $[a_F, \infty)$ to hold is that $a_F \geq a_G$. As pointed out in Woodroofe (1985), this is the weakest condition since there does not exist any consistent estimator of F if $a_F < a_G$. However the question of strong uniform consistency remains open. The issue involves the following three cases:

CASE 1. $a_F > a_G$. This is a relatively easy case. Both strong uniform consistency and asymptotic normality hold. In fact, in Theorem 1(iii) of Chao and Lo (1988), $\hat{F}_n - F$ is represented as an average of the sum of iid random variables with mean zero and finite variance, plus a negligible term. Therefore the strong uniform consistency and the asymptotic normality may be viewed as corollaries. The strong uniform consistency in this case has also been proved, for example, by Wang, Jewell and Tsai (1986) and Wellek (1990).

CASE 2. $a_F = a_G$. To the best of our knowledge, under no further condition the only result obtained so far is the weak uniform consistency proved in Woodroofe (1985) and Kieding and Gill (1990). Under the condition $\int_0^\infty G^{-1}\, dF < \infty$, Chao and Lo [(1988), Theorem 2(ii)] obtained an almost sure representation of $\hat{F}_n$ in terms of the iid process with remainder term $o(n^{-1/2})$. In a recent paper, Stute (1993) further proved that the remainder term is

$O(n^{-1}(\log n)^3)$ under a stronger condition $\int_0^\infty G^{-2}\,dF < \infty$. (Note that $\int_0^\infty G^{-2}\,dF < \infty$ implies $\int_0^\infty G^{-1}\,dF < \infty$.) The strong uniform consistency of $\hat{\bar{F}}_n$ can then be easily derived from this representation under the condition $\int_0^\infty G^{-1}\,dF < \infty$ [see Chao and Lo (1988)]. The condition $\int_0^\infty G^{-1}\,dF < \infty$ simply guarantees that the asymptotic variance of $\hat{\bar{F}}_n$ stays bounded as n goes to ∞. It is unknown whether the stronger uniform consistency still holds without this strong condition.

CASE 3. $a_F < a_G$. A consistent estimator of $\bar{F}$ does not exist. However, one can use $\hat{\bar{F}}'(x) \equiv \hat{\bar{F}}_n(x)/\hat{\bar{F}}_n(a_G)$ to estimate the conditional survival function of X given $X > a_G$; that is, $\bar{F}_0(x) \equiv \bar{F}(x)/\bar{F}(a_G)$ for $x > a_G$. However, in this case the property of strong uniform consistency can be easily reduced to Case 2.

Therefore, to solve the general problem, it remains to deal with Case 2. The purpose of this paper is to prove the strong uniform consistency for the Case 2 without any additional condition and thus solve the problem completely. Since our theorem is proved under the weakest conditions required in proving the weak consistency [see Woodroofe (1985)], the result contained in the following theorem is the best possible.

THEOREM 1. *Suppose F and G are continuous, $a_F = a_G = 0$. Then*

$$\sup_{x \geq 0} |\hat{\bar{F}}_n(x) - \bar{F}(x)| \to 0 \tag{2}$$

almost surely as $n \to \infty$.

REMARK. The above result can be easily extended to the case of left truncation and right censoring. The theorem still holds when the continuity assumption on F and G is relaxed to the assumption that $P(X = 0) = 0$.

2. Proofs.

We first introduce some notation to be used in our proofs.

Let $\rho = \int_0^\infty G\,dF = P(x_1 > y_1)$, where (x_1, y_1) is the first observation. Let

$$F_n(s) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{x_i \leq s\}}, \qquad \hat{\Lambda}_n(s) = \int_0^s \frac{dF_n}{C_n} \quad \text{and} \quad \tilde{\Lambda}_n(s) = \int_0^s \frac{dF_n}{C_n - 1/n}.$$

(Notice that $\tilde{\Lambda}_n$ could be ∞ instead.)

Let δ be an arbitrary but fixed positive constant such that $0 < F(\delta) < 1$. Define the events

$$A_n = \{\exists\ i \leq n, \text{such that } x_i \leq \delta,\ nC_n(x_i) \leq 2\}$$

and the integer-valued random variable

$$T = \begin{cases} 1, & \text{on } \bigcap_{i \ge 1} A_i^c, \\ n, & \text{on } A_n \bigcap_{i \ge n+1} A_i^c, \\ \infty, & \text{elsewhere.} \end{cases} \tag{3}$$

It is clear that T is the last time that A_n occurs ($T = 1$ if none of A_n, $n \ge 1$, occurs).

LEMMA 1. *Under the conditions given in Theorem* 1, *we have*

$$P(T < \infty) = 1 \tag{4}$$

and

$$E\tilde{\Lambda}_T(\delta) < \infty. \tag{5}$$

LEMMA 2. *Under the conditions given in Theorem* 1, *we have* $\hat{\Lambda}_n(\varepsilon_n) \to 0$ *almost surely as* $n \to \infty$ *for any sequence of nonincreasing positive numbers* $\{\varepsilon_n, n \ge 1\}$ *converging to* 0 *as* $n \to \infty$.

Next, we give the proof of Theorem 1, assuming the two lemmas stated above.

PROOF OF THEOREM 1. Let $\{\varepsilon_n, n \ge 1\}$ be as in Lemma 2. Then

$$\begin{aligned} -\log \hat{\bar{F}}_n(\varepsilon_n) &= \left(-\log \hat{\bar{F}}_n(\varepsilon_n) - \hat{\Lambda}_n(\varepsilon_n)\right) + \hat{\Lambda}_n(\varepsilon_n) \\ &= \sum_{x_i \le \varepsilon_n} \left(-\log\left(1 - \frac{1}{nC_n(x_i)}\right) - \frac{1}{nC_n(x_i)}\right) + \hat{\Lambda}_n(\varepsilon_n) \\ &\to 0 \end{aligned}$$

almost surely as $n \to \infty$ by Lemma 2 and the Taylor expansion. Therefore

$$\hat{\bar{F}}_n(\varepsilon_n) \to 1 \tag{6}$$

almost surely as $n \to \infty$.

It is known that [cf. Woodroofe (1985)] for any $b > 0$,

$$\sup_{x \ge b} \left| \frac{\hat{\bar{F}}_n(x)}{\hat{\bar{F}}_n(b)} - \frac{\bar{F}(x)}{\bar{F}(b)} \right| \to 0$$

almost surely. Therefore there exists a sequence of nonincreasing positive numbers b_n such that $b_n \to 0$ as $n \to \infty$ and

$$\sup_{x \ge b_n} \left| \frac{\hat{\bar{F}}_n(x)}{\hat{\bar{F}}_n(b_n)} - \frac{\bar{F}(x)}{\bar{F}(b_n)} \right| \to 0 \tag{7}$$

almost surely. It follows from (6), with ε_n replaced by b_n there, that $\hat{\bar{F}}_n(b_n) \to 1$ almost surely and also clearly $\bar{F}(b_n) \to 1$ as $n \to \infty$. With the help of (7), it can be shown that

$$\sup_{x \le b_n} \left|\hat{\bar{F}}_n(x) - \bar{F}(x)\right| \to 0 \quad \text{a.s. as } n \to \infty$$

and

$$\sup_{x \ge b_n} \left|\hat{\bar{F}}_n(x) - \bar{F}(x)\right| \to 0 \quad \text{a.s. as } n \to \infty,$$

which imply (2). The proof is thus complete. □

Now we proceed to prove the lemmas.

PROOF OF LEMMA 1. Notice that for fixed k, $nC_n(x_k)$ is a nondecreasing function of n. Therefore,

$$\begin{aligned} \{T = \infty\} &= \{A_n \text{ i.o.}\} = \{A_{n-1}^c \cap A_n \text{ i.o.}\} \cup \left(\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n\right) \\ &= \{x_n \le \delta,\ nC_n(x_n) \le 2 \text{ i.o.}\} \\ &\quad \cup \left(\bigcup_{k=1}^{\infty} \{nC_n(x_k) \le 2,\ x_k \le \delta, \text{for all } n \ge k\}\right). \end{aligned} \tag{8}$$

Some easy calculations give

$$\begin{aligned} &P(x_n \le \delta,\ nC_n(x_n) \le 2) \\ &\quad = \int_0^{\delta} (1 - \rho^{-1}\bar{F}G)^{n-1} \rho^{-1} G\, dF + (n-1)\int_0^{\delta} (1 - \rho^{-1}\bar{F}G)^{n-2} \rho^{-2}\bar{F}G^2\, dF \end{aligned}$$

for any $k \le n$. Now observe

$$\sum_{n=1}^{\infty} \int_0^{\delta} n^i (1 - \rho^{-1}\bar{F}G)^n G^{i+1}\, dF < \infty \tag{9}$$

for any $i \ge 0$; hence,

$$\sum_{n=1}^{\infty} P(x_n \le \delta,\ nC_n(x_n) \le 2) < \infty. \tag{10}$$

It then follows from the Borel-Cantelli lemma that

$$P(\{nC_n(x_n) \le 2,\ x_n \le \delta \text{ i.o.}\}) = 0. \tag{11}$$

Because $P(nC_n(x_k) \le 2,\ x_k \le \delta) = P(nC_n(x_n) \le 2,\ x_n \le \delta)$ for $k \le n$, (10) also implies for fixed k, $P(nC_n(x_k) \le 2,\ x_k \le \delta) \to 0$ as $n \to \infty$. Hence for each $k \ge 1$, $P(\{nC_n(x_k) \le 2,\ x_k \le \delta \text{ for all } n \ge k\}) = 0$. Now (4) follows from (8) and (11).

To show (5), first define the events

$$B_{ij}^k = \{x_i \le \delta,\ y_j < x_i \le x_j,\ y_{k+1} < x_i \le x_{k+1},\ kC_k(x_i) = 2\},$$
$$B^k = \{\forall\ i \le k, \text{ if } x_i \le \delta, \text{ then } kC_k(x_i) \ge 2\}.$$

Clearly, $A_{k+1}^c \subseteq B^k$. Write

$$\begin{aligned} A_k \cap A_{k+1}^c &= \{\exists\ i \le k,\ x_i \le \delta,\ kC_k(x_i) \le 2\} \\ &\qquad \cap \{\forall\ i \le k+1, \text{ if } x_i \le \delta, \text{ then } (k+1)C_{k+1}(x_i) \ge 3\} \\ &\subseteq \Bigg(\bigcup_{\substack{1 \le i, j \le k \\ i \ne j}} B_{ij}^k\Bigg) \cap B^k. \end{aligned}$$

Because $\{T \le k\} \subseteq A_{k+1}^c \subseteq B^k$, $\tilde{\Lambda}_{T \vee k}$ is finite for any $k \ge 1$ and hence $\tilde{\Lambda}_{T \vee k}$ is well defined in view of (4). It follows that

$$\begin{aligned} E\tilde{\Lambda}_T(\delta) &= \sum_{k=3}^{\infty} E\tilde{\Lambda}_k(\delta) 1_{\{T=k\}} + E\tilde{\Lambda}_T(\delta) 1_{\{T \le 2\}} \\ &\le \sum_{k=3}^{\infty} E\tilde{\Lambda}_k(\delta) 1_{A_k \cap_{j \ge k+1} A_j^c} + 2 \\ &\le \sum_{k=3}^{\infty} E\tilde{\Lambda}_k(\delta) 1_{A_k \cap A_{k+1}^c} + 2 \\ &\le \sum_{k=3}^{\infty} E\tilde{\Lambda}_k(\delta) 1_{(\cup_{1 \le i, j \le k,\, i \ne j} B_{ij}^k) \cap B^k} + 2. \end{aligned}$$

For all $k \ge 3$, we have

$$\begin{aligned} &E\tilde{\Lambda}_k(\delta) 1_{(\cup_{1 \le i, j \le k,\, i \ne j} B_{ij}^k) \cap B^k} \\ &\quad \le k(k-1) E\tilde{\Lambda}_k(\delta) 1_{B_{12}^k \cap B^k} \\ &\quad \le k(k-1) \sum_{i=1}^{2} E \frac{1_{\{x_i \le \delta\} \cap B_{12}^k \cap B^k}}{kC_k(x_i) - 1} + k(k-1) \sum_{i=3}^{k} E \frac{1_{\{x_i \le \delta\} \cap B_{12}^k \cap B^k}}{kC_k(x_i) - 1} \\ &\quad \le 2k(k-1) P(B_{12}^k) + k(k-1)(k-2) E \frac{1_{\{x_3 \le \delta\} \cap B_{12}^k \cap B^k}}{kC_k(x_3) - 1} \\ &\quad \le 2k^2 P(B_{12}^k) + 2k^3 P(\{x_3 < x_1 < \delta\} \cap B_{12}^k) + 2k^3 E \frac{1_{\{x_1 \le y_3 \le x_3 \le \delta\} \cap B_{12}^k}}{\sum_{j=3}^{k} 1_{\{y_j < x_3 \le x_j\}}} \\ &\quad = \mathrm{I}_k + \mathrm{II}_k + \mathrm{III}_k \quad \text{(say)}. \end{aligned}$$

Again by some calculations and using (9), we have

$$
\begin{aligned}
\sum_{k=3}^{\infty} \mathrm{I}_k &\le 2\sum_{k=3}^{\infty} k^2 P\big(B_{12}^k\big)\\
&= 2\sum_{k=3}^{\infty} k^2 \int_0^{\delta}\big(\rho^{-1}\overline{F}G\big)^2\big(1-\rho^{-1}\overline{F}G\big)^{k-2}\rho^{-1}G\,dF\\
&< \infty,\\
\sum_{k=3}^{\infty} \mathrm{II}_k &= 2\sum_{k=3}^{\infty} k^3 P\big(\{x_3 < x_1 \le \delta\}\cap B_{12}^k\big)\\
&= 2\sum_{k=3}^{\infty} k^3 \int_0^{\delta}\big(\rho^{-1}\overline{F}(s)G(s)\big)^2\rho^{-1}\\
&\quad\times\left(\int_0^{s} G\,dF\right)\big(1-\rho^{-1}\overline{F}(s)G(s)\big)^{k-3}\rho^{-1}G(s)\,dF(s)\\
&\le 2\sum_{k=3}^{\infty} k^3 \int_0^{\delta}\big(1-\rho^{-1}\overline{F}G\big)^{k-3}\rho^{-4}F\overline{F}^2G^4\,dF\\
&< \infty,\\
\sum_{k=3}^{\infty} \mathrm{III}_k &= \sum_{k=3}^{\infty} 2k^3 E\frac{\mathbf{1}_{\{x_1\le y_3\le x_3\le\delta\}\cap B_{12}^k}}{\sum_{j=3}^{k}\mathbf{1}_{\{y_j<x_3\le x_j\}}}\\
&= \sum_{k=3}^{\infty} 2k^3\int_0^{\delta}\big(\rho^{-1}\overline{F}(t)G(t)\big)^2\rho^{-1}G(t)\\
&\quad\times\left(E\frac{\mathbf{1}_{\{t\le y_3\le x_3\le\delta\}}\cdot\prod_{j=3}^{k}\big(\mathbf{1}_{\{x_j<t\}}+\mathbf{1}_{\{y_j\ge t\}}\big)}{\sum_{j=3}^{k}\mathbf{1}_{\{y_j<x_3\le x_j\}}}\right)dF(t)\\
&= \sum_{k=3}^{\infty} 2k^3\int_0^{\delta}\big[\rho^{-1}\overline{F}(t)G(t)\big]^2\rho^{-1}G(t)\\
&\quad\times\left(\int_t^{\delta}\sum_{j=0}^{k-3}\frac{1}{j+1}\binom{k-3}{j}\big[\overline{F}(s)(G(s)-G(t))\big]^j\right.\\
&\qquad\times\big[1-\rho^{-1}\overline{F}(t)G(t)-\overline{F}(s)(G(s)-G(t))\big]^{k-3-j}\\
&\qquad\qquad\left.\times(G(s)-G(t))\,dF(s)\right)dF(t)\\
&= \sum_{k=3}^{\infty}\frac{2k^3}{k-2}\int_0^{\delta}\Big[\big[\rho^{-1}\overline{F}(t)G(t)\big]^2\big[1-\rho^{-1}\overline{F}(t)G(t)\big]^{k-2}\\
&\qquad\times\rho^{-1}G(t)\int_t^{\delta}\overline{F}^{-1}(s)\Big(1-\Big[1-\big(\rho-\overline{F}(t)G(t)\big)^{-1}\overline{F}(s)\\
&\qquad\qquad\times(G(s)-G(t))\Big]^{k-2}\Big)\,dF(s)\Big]\,dF(t)
\end{aligned}
$$

$$\leq \sum_{k=3}^{\infty} 6k^3\bar{F}^{-1}(\delta)\int_0^\delta \rho^{-3}\bar{F}^2G^3\left(1-\rho^{-1}\bar{F}G\right)^{k-2}dF$$
$$< \infty.$$

By combining the above arguments, it follows that $E\tilde{\lambda}_T(\delta) < \infty$. The proof of Lemma 1 is complete. □

To prove Lemma 2 we need to introduce some notation for various σ-algebras.

A random variable is called n-symmetric if it is a function of $\{(x_1, y_1), (x_2, y_2), \dots\}$ and is unchanged under any permutations of the first n observations [cf. Hall and Heyde (1980), page 202]. Let $\mathscr{F}_n$ denote the σ-algebra generated by the n-symmetric random variables and note that $\mathscr{F}_{n+1} \subseteq \mathscr{F}_n$. Let $\mathscr{F}_n^i$ be the σ-algebra generated by $\mathscr{F}_n$ and (x_i, y_i) [i.e., $\mathscr{F}_n^i = \sigma(\mathscr{F}_n, (x_i, y_i))$]. Then clearly, $\mathscr{F}_n \subseteq \mathscr{F}_n^i$.

Proof of Lemma 2. Let ε_n be an arbitrary but fixed sequence of nonincreasing positive numbers converging to 0 as $n \to \infty$. First notice that for $1 \leq i \leq n+1$, $1 \leq k \leq n+1$, $1 \leq j \leq n+1$ and $k \neq i \neq j$,

$$P\left(y_k < x_i \leq x_k \mid \mathscr{F}_{n+1}^i\right) = P\left(y_j < x_i \leq x_j \mid \mathscr{F}_{n+1}^i\right);$$

hence,

$$P\left(y_k < x_i \leq x_k \mid \mathscr{F}_{n+1}^i\right) = \frac{1}{n}\sum_{\substack{k=1\\k\neq i}}^{n+1} 1_{\{y_k < x_i \leq x_k\}} = \frac{1}{n}\left((n+1)C_{n+1}(x_i) - 1\right).$$

Therefore, for any $i \leq n$,

$$E\left(C_n(x_i) - \frac{1}{n}\middle|\mathscr{F}_{n+1}^i\right) = \frac{1}{n}E\left(\sum_{\substack{k=1\\k\neq i}}^{n} 1_{\{y_k < x_i \leq x_k\}}\middle|\mathscr{F}_{n+1}^i\right)$$
$$= \frac{n-1}{n^2}\left((n+1)C_{n+1}(x_i) - 1\right) \leq C_{n+1}(x_i) - \frac{1}{n+1},$$

where the last inequality is due to the fact that $C_{n+1}(x_i) \geq 1/(n+1)$ for $i \leq n$.

Without loss of generality, we can assume $\varepsilon_n < \delta$ for all $n \geq 1$. Because $A_n \in \mathscr{F}_n$, clearly $\{T = n\} \in \mathscr{F}_n$. Therefore, $\{T \geq n+1\} \in \mathscr{F}_{n+1}$, $\{T \leq n\} \in \mathscr{F}_{n+1}$ and $\tilde{\Lambda}_{T\vee n}(\varepsilon_n) \in \mathscr{F}_n$.

Now for $n \geq 1$, we have

$$E\left(\tilde{\Lambda}_{T\vee n}(\varepsilon_n) \mid \mathscr{F}_{n+1}\right)$$
$$\geq E\left(\tilde{\Lambda}_{T\vee n}(\varepsilon_{n+1}) \mid \mathscr{F}_{n+1}\right)$$
$$= E\left(\tilde{\Lambda}_n(\varepsilon_{n+1})1_{\{T\leq n\}} \mid \mathscr{F}_{n+1}\right) + E\left(\tilde{\Lambda}_{T\vee(n+1)}(\varepsilon_{n+1})1_{\{T\geq n+1\}} \mid \mathscr{F}_{n+1}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E\left(\frac{1_{\{x_i \le \varepsilon_{n+1}\}\cap\{T\le n\}}}{C_n(x_i) - 1/n} \middle| \mathscr{F}_{n+1}\right) + \tilde{\Lambda}_{T\vee(n+1)}(\varepsilon_{n+1})1_{\{T\ge n+1\}}$$

$$= \frac{1}{n}\sum_{i=1}^{n} E\left(E\left(\frac{1_{\{x_i \le \varepsilon_{n+1}\}\cap\{T\le n\}}}{C_n(x_i) - 1/n} \middle| \mathscr{F}^i_{n+1}\right) \middle| \mathscr{F}_{n+1}\right) + \tilde{\Lambda}_{T\vee(n+1)}(\varepsilon_{n+1})1_{\{T\ge n+1\}}$$

$$\ge \frac{1}{n}\sum_{i=1}^{n} E\left(\frac{1_{\{x_i \le \varepsilon_{n+1}\}\cap\{T\le n\}}}{E(C_n(x_i) - 1/n \mid \mathscr{F}^i_{n+1})} \middle| \mathscr{F}_{n+1}\right) + \tilde{\Lambda}_{T\vee(n+1)}(\varepsilon_{n+1})1_{\{T\ge n+1\}}$$

$$\ge \frac{1}{n}\sum_{i=1}^{n} E\left(\frac{1_{\{x_i \le \varepsilon_{n+1}\}\cap\{T\le n\}}}{C_{n+1}(x_i) - 1/(n+1)} \middle| \mathscr{F}_{n+1}\right) + \tilde{\Lambda}_{T\vee(n+1)}(\varepsilon_{n+1})1_{\{T\ge n+1\}}$$

$$= \frac{1}{n+1}\sum_{i=1}^{n+1} \frac{1_{\{x_i \le \varepsilon_{n+1}\}}}{C_{n+1}(x_i) - 1/(n+1)} 1_{\{T\le n\}} + \tilde{\Lambda}_{T\vee(n+1)}(\varepsilon_{n+1})1_{\{T\ge n+1\}}$$

$$= \tilde{\Lambda}_{n+1}(\varepsilon_{n+1})1_{\{T\le n\}} + \tilde{\Lambda}_{T\vee(n+1)}(\varepsilon_{n+1})1_{\{T\ge n+1\}}$$

$$= \tilde{\Lambda}_{T\vee(n+1)}(\varepsilon_{n+1}).$$

Combining with (5), we have shown $\{\tilde{\Lambda}_{T\vee n}(\varepsilon_n), \mathscr{F}_n;\ n \ge 1\}$ is an L^1 reversed nonnegative submartingale. By the reverse submartingale convergence theorem [cf. Hall and Heyde (1980)], there exists some random variable ξ, such that $\tilde{\Lambda}_{T\vee n}(\varepsilon_n)$ converges to ξ almost surely and in mean. Because on the set $\{T \le n\}$, we have $nC_n(x_i) \ge 2$ and hence $nC_n(x_i) - 1 \le nC_n(x_i) \le 2(nC_n(x_i) - 1)$ for all $i \le n$ such that $x_i \le \delta$. Therefore,

$$\tfrac{1}{2}\tilde{\Lambda}_{T\vee n}(\varepsilon_n) \le \hat{\Lambda}_{T\vee n}(\varepsilon_n) \le \tilde{\Lambda}_{T\vee n}(\varepsilon_n). \tag{12}$$

With some easy calculations [cf. Woodroofe (1985)], we have

$$E\hat{\Lambda}_n(\varepsilon_n) = -\log \bar{F}(\varepsilon_n) - \int_0^{\varepsilon_n}(1 - \rho^{-1}\bar{F}G)^n \bar{F}^{-1}\,dF \le -\log \bar{F}(\varepsilon_n) \to 0$$

as $n \to \infty$, whence $\hat{\Lambda}_n(\varepsilon_n)$ converges to 0 in probability. Since T is almost surely finite in view of (4), we know that $\hat{\Lambda}_{T\vee n}(\varepsilon_n)$ also converges to 0 in probability. So it follows from (12) that $\xi \le 0$, but clearly ξ is nonnegative; therefore, $\xi = 0$. Now again by (12) and the almost sure convergence of $\tilde{\Lambda}_{T\vee n}(\varepsilon_n)$, we know that $\hat{\Lambda}_{T\vee n}(\varepsilon_n)$ converges to 0 almost surely as $n \to \infty$. Therefore, $\hat{\Lambda}_n(\varepsilon_n)$ converges to 0 almost surely by finiteness of T. The proof is thus complete. □

REFERENCES

BHATTACHARYA, P. K. (1983). Justification for a *K-S* type test for the slope of truncated regression. *Ann. Statist.* **11** 697–701.

BHATTACHARYA, P. K., CHERNOFF, H. and YANG, S. S. (1983). Nonparametric estimation of the slope of a truncated regression. *Ann. Statist.* **11** 505–514.

CHAO, M. T. and LO, S. H. (1988). Some representations of the nonparametric maximum likelihood estimators with truncated data. *Ann. Statist.* **16** 661–668.

GILL, R. D. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.* **11** 49–58.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application.* Academic Press, New York.

KIEDING, N. and GILL, R. D. (1990). Random truncation models and Markov processes. *Ann. Statist.* **18** 582–602.

LAGAKOS, S. W., BARRAJ, L. M. and DE GRUTTOLA, V. (1988). Nonparametric analysis of truncated survival data with application to AIDS. *Biometrika* **75** 515–523.

LO, S-H. and SINGH, K. (1986). The product-limit estimator and the bootstrap: some asymptotic representations. *Probab. Theory Related Fields* **17** 455–465.

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** 95–118.

STUTE, W. (1993). Almost sure representations of the product-limit estimator for truncated data. *Ann. Statist.* **21** 146–156.

WANG, J. G. (1987). A note on the uniform consistency of the Kaplan–Meier estimator. *Ann. Statist.* **15** 1313–1316.

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.

WELLEK, S. (1990). A nonparametric model for product-limit estimation under right censoring and left truncation. *Comm. Statist. Models* **6** 561–592.

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

K. CHEN
S-H. LO
DEPARTMENT OF STATISTICS
BOX 10 MATHEMATICS
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027

M-T. CHAO
INSTITUTE OF STATISTICAL SCIENCE
ACADEMIA SINICA
TAIPEI, TAIWAN
REPUBLIC OF CHINA