*The Annals of Statistics*
1998, Vol. 26, No. 3, 992-1010

# THE STRONG LAW UNDER RANDOM TRUNCATION[1]

BY SHUYUAN HE AND GRACE L. YANG

*Peking University and University of Maryland*

The random truncation model is defined by the conditional probability distribution $H(x, y) = P[X \le x,\ Y \le y | X \ge Y]$ where $X$ and $Y$ are independent random variables. A problem of interest is the estimation of the distribution function $F$ of $X$ with data from the distribution $H$. Under random truncation, $F$ need not be fully identifiable from $H$ and only a part of it, say $F_0$, is. We show that the nonparametric MLE $F_n$ of $F_0$ obeys the strong law of large numbers in the sense that for any nonnegative, measurable function $\phi(x)$, the integrals $\int \phi(x)\, dF_n(x) \to \int \phi(x)\, dF_0(x)$ almost surely as $n$ tends to infinity. Similar results were first obtained by Stute and Wang for the right censoring model. The results are useful in establishing the strong consistency of various estimates. Some of our results are derived from the weak consistency of $F_n$ obtained by Woodroofe.

**1. Introduction.** Consider an infinite sequence of independent random vectors $(X_m, Y_m),\ m = 1, 2, \ldots,$ where the $X_m$ have a common distribution function $F$ and the $Y_m$ have a common distribution function $G$. The components $X_m$ and $Y_m$ are also independent for each $m$. Suppose both $X_m$ and $Y_m$ are observable only when $X_m \ge Y_m$. The observable pairs thus form a subsequence $\{j\}$ of the original sequence $\{m\}$. It is denoted by $\{(U_j, V_j),\ j = 1, 2, \ldots\}$. Here the subsequence is labeled consecutively for simplicity. The limitation in observation induces dependence and the constraint $U_j \ge V_j$ in each pair $j$. However, the vectors $(U_j, V_j)$ remain iid. In describing the distributional properties of any pair we shall use $(X, Y)$ to refer to any pair $(X_m, Y_m)$, and $(U, V)$ to $(U_j, V_j)$.

The random truncation model is defined by the joint distribution $H(x, y)$ of $(U, V)$. It is the conditional distribution of $(X, Y)$ given $[X \ge Y]$,

$$H(x, y) = P[U \le x,\ V \le y] = P[X \le x,\ Y \le y | X \ge Y]. \tag{1}$$

A problem of interest is to estimate the distribution function $F$ of $X$ based on a randomly truncated sample of $n$ iid observations $(U_j, V_j),\ j = 1, \ldots, n$. Truncated data occur in astronomy, economics [e.g., Woodroofe (1985), Feigelson and Babu (1992)], epidemiology, biometry [e.g., Wang, Jewell and Tsai (1986), Tsai, Jewell and Wang (1987), He and Yang (1994)] and possibly in other fields such as spike train data in neurophysiology.

Received March 1996; revised January 1998.

[1]Supported in part by NSF Grant DMS-94-041114 and the National Natural Science Foundation of China Grant 19571002.

*AMS* 1991 *subject classification*. 62G20

*Key words and phrases*. Random truncation, nonparametric estimation, strong law of large numbers, reverse supermartingale, product limit.

The truncation event $[X \geq Y]$, among other things, affects the range of observation of the $X$. Only $F_0$ defined by

$$F_0(x) = P[X \leq x | X \geq a_G] \tag{2}$$

is estimable from the truncated sample $(U_j, V_j)$, $j = 1, \ldots, n$, where

$$a_G = \inf\{y\colon G(y) > 0\}$$

is the lower boundary of $Y$. We shall denote the upper boundary of $Y$ by

$$b_G = \sup\{y\colon G(y) < 1\}. \tag{3}$$

Similar symbols, $a_F, b_F$, will be used for the boundaries of $X$.

Obviously, if $a_G \leq a_F$, $F_0 = F$. Analogously, define $G_0(y) = P[Y \leq y | Y \leq b_F]$. Thus if $b_F \geq b_G$, $G_0 = G$. Let $I[A]$ denote the indicator function of the event $A$. Let

$$\begin{aligned} &F_n^*(s) = n^{-1} \sum_{i=1}^n I[U_i \leq s], \qquad G_n^*(s) = n^{-1} \sum_{i=1}^n I[V_i \leq s], \\ &R_n(s) = G_n^*(s) - F_n^*(s-) = n^{-1} \sum_{i=1}^n I[V_i \leq s \leq U_i], \qquad -\infty < s < \infty \end{aligned} \tag{4}$$

be the empirical processes of the data.

Here and in what follows, for any real function $g$, the left limit $\lim_{y \uparrow s} g(y)$ is denoted by $g(s-)$ and the difference $g(s) - g(s-)$ by the curly brackets $g\{s\}$.

The nonparametric maximum likelihood estimates of $F_0$ and $G_0$ are given, respectively, by

$$F_n(x) = 1 - \prod_{s \leq x} \left[1 - \frac{F_n^*\{s\}}{R_n(s)}\right] \quad \text{and} \quad G_n(x) = \prod_{s > x} \left[1 - \frac{G_n^*\{s\}}{R_n(s)}\right], \tag{5}$$

where $x \in (-\infty, \infty)$ and an empty product is set equal to 1.

One of the results obtained by Woodroofe (1985) is that for any continuous $F$ and $G$,

$$\sup_x |F_n(x) - F_0(x)| \to 0 \quad \text{in probability as } n \to \infty.$$

If $F$ and $G$ are not continuous, the limit has to be modified. For arbitrary $F$ and $G$, there are two kinds of limit, $F_0$ and $F_a$, where $F_a$ is defined by

$$F_a(x) = P[X \leq x | X > a_G].$$

Under Condition B1: $a_F = a_G$, $G\{a_G\} = 0$ and $F\{a_F\} > 0$, we show in Theorem 5.6 and Corollary 5.8 that

$$\sup_x |F_n(x) - F_a(x)| \to 0 \quad \text{a.s.}$$

Otherwise,

$$\sup_x |F_n(x) - F_0(x)| \to 0 \quad \text{a.s.}$$

In Theorem 5.7, Corollaries 5.8 and 5.9, we show that for arbitrary $F$ and $G$ the strong law of large number (SLLN) holds for the MLE estimate $F_n$, that is, for any measurable nonnegative function $\varphi$,

$$\int \varphi(x)\,dF_n(x) \to \int \varphi(x)\,dF_a(x) \quad \text{as } n \to \infty, \tag{6}$$

almost surely under Condition B1. If B1 does not hold, then the convergence in (6) takes place with $F_a$ replaced by $F_0$.

The convergence in (6) implies immediately the a.s. convergence of the sample moments $\int x^k\,dF_n(x) \to E_0X^k = \int x^k\,dF_0(x)$, if $\varphi(x) = x^k$ and $E_0X^k$ are finite. Similarly, if $\varphi(x) = e^{itx}$ or $\varphi(x) = I[X \le x]$, we have the a.s. convergence of the empirical characteristic function and $F_n$. Stating the SLLN in terms of a nonnegative $\varphi$ is for convenience, since any measurable $\varphi$ can be decomposed into a positive and a negative parts.

In Sections 3 and 4, we prove the SLLN for continuous $F$ and $G$ (in the continuous case $F_0 = F_a$). The proof of the general case of $F$ and $G$ is given in Section 5.

This investigation is motivated by the work of Stute and Wang (1993) who obtained the similar strong law for the Kaplan–Meier estimator under right censoring. Although both $F_n$ and the Kaplan–Meier estimator can be written in product forms, the $R_n(s)$ that appears in $F_n$ is not a monotone function of $s$, whereas a comparable term in the Kaplan–Meier estimate is. Without the monotonicity, we cannot directly apply the martingale property of the Kaplan–Meier integrals obtained by Stute and Wang (1993).

However, the monotone property of the cumulative hazard functions $\Lambda_n$ of $F_n$ and $\Lambda_0$ of $F_0$ can be utilized. Following the approach of Stute and Wang (1993), particularly their Lemma 2.1, we show in Theorem 3.1 of Section 3 that the sequence of integrals $\int \varphi(x)\,d\Lambda_n(x)$ forms a reverse supermartingale. Based on Theorem 3.1, we establish the uniform strong convergence of the cumulative hazard function $\Lambda_n$ of $F_n$ to the cumulative hazard function $\Lambda_0$ of $F_0$. The strong limit $\Lambda_0$ is identified by utilizing the existence of the weak limit $\Lambda_0$ as proved by Woodroofe (1985). It then follows that $\tilde{F}_n(x) = 1 - \exp(-\Lambda_n(x))$ converges uniformly and strongly to $F_0(x) = 1 - \exp(-\Lambda(x))$, which in turn entails that the difference between $\tilde{F}_n(x)$ and the MLE, $F_n$, tends to zero as $n$ tends to infinity.

In Section 4, we show in Theorem 4.1 that $\sup_x |F_n(x) - F_0(x)| \to 0$ almost surely. The proof is nontrivial. It requires careful analysis of the empirical process $R_n(x)$ near the lower boundary $a_G$ of the observable $X$. Theorem 4.1 is needed for establishing the strong law stated in (6) for continuous $F$ and $G$. The proof is given in Theorem 4.3. The proof relies crucially on a special representation of the estimate $\alpha_n = \int G_n(x)\,dF_n(x)$ of the truncation probability $\alpha = P[X \ge Y]$; see He and Yang (1998).

The generalization to arbitrary $F$ and $G$ is presented in Section 5. In addition to being theoretically interesting, the results are useful for estimating $F$ from grouped data which we could treat then as from a discrete distribution $F$.

We adopted the method of proof used in Major and Rejtö (1988) by showing that the general case can be reduced to the continuous case via appropriate transformations of the random variable $X_i$ to $\hat{X}_i$ and $Y_i$ to $\hat{Y}_i$. The relationship between $X_i$, $Y_i$ and their transformations are given in Lemma 5.1. Lemma 5.1 facilitates the proof of almost sure convergence of (6) for arbitrary $F$ and $G$ as given in Theorem 5.7. The proofs of Lemmas 5.1, 5.4 and 5.5 are relegated to the Appendix.

The main result of this article is the strong law given in Theorem 5.7. The $L_1$ convergence proved by Stute and Wang (1993) for the right-censored data remains an open problem for the random truncation data.

Finally, we note that the uniform strong consistency of $F_n$ was proved by Chen, Chao and Lo (1995), but only for continuous $F$ and $G$. Our proof of the continuous case given in Theorem 4.1 is different and, we believe, simpler. The general case presented in Section 5 was not considered by Chen, Chao and Lo (1995).

**2. Assumptions and preliminaries.** The following assumptions are used in Sections 3 and 4. Assumption A1 is eliminated in Section 5.

A1. $F$ and $G$ are continuous.
A2. The supports of $F$ and $G$ are not disjoint, that is, $a_G < b_F$.

Assumption A2 is to avoid mathematical triviality. It guarantees a positive probability $\alpha = P[X \geq Y]$ of observing $X$ and $Y$, which, in turn, implies by the SLLN that infinitely many of the events $[X_i \geq Y_i]$, $i = 1, 2, \ldots$ will occur. Therefore, a sample of $n$ iid random vectors $(U_i, V_i)$ is (with probability 1) always available from the original sequence and $n$ corresponds to some $m_n$ in the original sequence $\{(X_m, Y_m),\ m = 1, 2, \ldots\}$. For further discussion, see He and Yang (1998).

The marginal distributions of $U$ and $V$ are given by

$$\begin{aligned} F^*(x) &= P[U \leq x] = P[X \leq x | X \geq Y] = \alpha^{-1} \int_{-\infty}^{x} G(s)\, dF(s), \\ G^*(x) &= P[V \leq x] = P[Y \leq x | X \geq Y] = \alpha^{-1} \int_{-\infty}^{x} \bar{F}(s-)\, dG(s), \end{aligned} \tag{7}$$

where $\int_a^b$ stands for the Lebesque integral $\int_{(a,b]}$ and $\bar{F}(s) = 1 - F(s)$. It can be checked that $a_{F_0} = a_{F^*} \geq a_G$ and $b_{F_0} = b_{F^*} = b_F$.

Put

$$R(x) = G^*(x) - F^*(x-) = P[V \leq x \leq U]. \tag{8}$$

It is seen that

$$R(x) = \alpha^{-1} \bar{F}(x-) G(x) > 0 \tag{9}$$

if and only if $x \in (a_G, b_F)$. Equation (9) holds for arbitrary $F$ and $G$; see Woodroofe (1985) for further discussion.

This relationship provides an estimating equation for the parameter $\alpha$; see He and Yang (1998) in this issue.

We shall need the cumulative hazard functions of $F_n$ and $F_0$. They are defined, respectively, by

$$\Lambda_n(x) = \int_{-\infty}^{x} \frac{dF_n(s)}{1 - F_n(s-)}, \tag{10}$$

and

$$\Lambda(x) = \int_{-\infty}^{x} \frac{dF_0(s)}{1 - F_0(s-)} = \int_{a_{F_0}}^{x} \frac{dF(s)}{1 - F(s-)}, \qquad a_{F_0} \le x < \infty. \tag{11}$$

Applying (5), (7) and (8), it is easy to check that for $x \in (a_{F^*}, b_{F^*})$, $\Lambda_n(x)$ and $\Lambda(x)$ can be written as

$$\Lambda_n(x) = \int_{-\infty}^{x} \frac{dF_n^*(s)}{R_n(s)}, \qquad \Lambda(x) = \int_{-\infty}^{x} \frac{dF^*(s)}{R(s)}. \tag{12}$$

Note that $F_n\{s\} = \bar{F}_n(s-)F_n^*\{s\}/R_n(s)$. The empirical hazard function $\Lambda_n(x)$ is a finite sum and a step function in $x$. The cumulative hazard function $\Lambda(x)$ is finite for $x \in (a_{F^*}, b_{F^*})$ and $\Lambda(x) \to \infty$ as $x \uparrow b_{F^*}$.

We use the convention $0/0 = 0$ throughout.

**3. Reverse supermartingales.** Assumptions A1 and A2 are imposed throughout Sections 3 and 4. Let $U_{1:n} \le U_{2:n} \le \cdots \le U_{n:n}$ be the ordered values of $U_j$'s among the first $n$ observations, and $V_{j:n}$, the concomitant of $U_{j:n}$ for $j = 1, \dots, n$. Let

$$\mathscr{F}_n = \sigma\{U_{j:n},\ V_{j:n},\ 1 \le j \le n, (U_k, V_k), k \ge n+1\}, \qquad n = 1, 2, \dots \tag{13}$$

be a sequence of $\sigma$-fields generated by the first $n$ ordered $U_{j:n}$, their concomitants $V_{j:n}$ and the rest of the unordered $(U_k, V_k)$ of the infinite sequence. This is a decreasing sequence, $\mathscr{F}_n \supset \mathscr{F}_{n+1}$ $\forall\, n \ge 1$, and, by definition, $\Lambda_n(t) \in \mathscr{F}_n$ for every $n$. Therefore, for any nonnegative measurable function $\varphi(x)$ the integral

$$S_n \equiv \int \varphi(x)\, d\Lambda_n(x) \in \mathscr{F}_n \qquad \forall\, n \ge 1.$$

Unless specified otherwise, the integral sign means integrating from $-\infty$ to $\infty$.

For simplicity, we shall suppress the statement "as $n \to \infty$" if the convergence is clearly understood to be with respect to $n$ tending to infinity.

THEOREM 3.1. *The sequence* $\{S_n, \mathscr{F}_n;\ n \ge 1\}$ *is a reverse supermartingale, that is,*

$$E[S_n | \mathscr{F}_{n+1}] \le S_{n+1} \qquad \forall\, n \ge 1.$$

PROOF. The cumulative hazard function $\Lambda_n(x)$ is a step function and has jumps only at each $U_{j:n}$, $j = 1, \dots, n$. Evaluating $\Lambda_n$, we obtain by (12),

$$\Lambda_n(U_{j:n}) = \sum_{i=1}^{j} \frac{F_n^*\{U_{i:n}\}}{R_n(U_{i:n})} \in \mathscr{F}_n, \tag{14}$$

and $\Lambda_n(u) = \Lambda_n(U_{j:n})$ for all $u \in [U_{j:n}, U_{j+1:n})$. Its mass at $U_{j:n}$, $\Lambda_n(U_{j:n}) - \Lambda_n(U_{j:n}-)$ is

$$\Lambda_n\{U_{j:n}\} = \frac{F_n^*\{U_{j:n}\}}{R_n(U_{j:n})} = \frac{1}{nR_n(U_{j:n})} \quad \text{a.s. } 1 \le j \le n.$$

Evaluating the mass of $\Lambda_n$ at the next order statistic $u = U_{j:n+1}$ yields

$$\Lambda_n\{U_{j:n+1}\} = \frac{F_n^*\{U_{j:n+1}\}}{R_n(U_{j:n+1})}, \qquad 1 \le j \le n+1.$$

One sees that the integral $S_n$ can be written as the sum

$$S_n = \sum_{j=1}^{n+1} \varphi(U_{j:n+1})\Lambda_n\{U_{j:n+1}\}. \tag{15}$$

Thus the theorem will be proved if we show

$$E[\Lambda_n\{U_{j:n+1}\}|\mathscr{F}_{n+1}] \le \Lambda_{n+1}\{U_{j:n+1}\}, \tag{16}$$

for each fixed $j$, $j = 1, 2, \dots, n+1$. Put $I_k = I[U_{n+1} = U_{k:n+1}]$. If $I_k$ occurs, then

$$\begin{aligned} U_{j:n+1} &= U_{j:n}, \\ V_{j:n+1} &= V_{j:n} \quad \text{if } j \le k-1 \end{aligned}$$

and

$$\begin{aligned} U_{j:n+1} &= U_{j-1:n}, \\ V_{j:n+1} &= V_{j-1:n} \quad \text{if } j \ge k+1. \end{aligned}$$

It follows by Lemma 2.1 of Stute and Wang (1993) that $E[I_k|\mathscr{F}_{n+1}] = 1/(n+1)$. Now we have

$$\begin{aligned} &E(\Lambda_n\{U_{j:n+1}\}|\mathscr{F}_{n+1}) \\ &\quad = \sum_{k\ne j} E\left(\left(\frac{1}{\sum_{i=1}^{k-1} I(V_{i:n} \le U_{j:n+1} \le U_{i:n}) + \sum_{i=k}^{n} I(V_{i:n} \le U_{j:n+1} \le U_{i:n})}\right) \right. \\ &\qquad\qquad \left. \times I_k|\mathscr{F}_{n+1}\right) \\ &\quad = \sum_{k\ne j} \frac{1}{(n+1)R_{n+1}(U_{j:n+1}) - I[V_{k:n+1} \le U_{j:n+1} \le U_{k:n+1}]} \frac{1}{n+1}. \end{aligned}$$

On the set $\{(n+1)R_{n+1}(U_{j:n+1}) = m\}$, $m = 1, 2, \dots, n+1$, we have $\sum_{k\ne j} I[V_{k:n+1} \le U_{j:n+1} \le U_{k:n+1}] = m-1$. Therefore, for $m > 1$,

$$E(\Lambda_n\{U_{j:n+1}\}|\mathscr{F}_{n+1}) = \frac{1}{n+1}\left(\frac{m-1}{m-1} + \frac{n-m+1}{m}\right) = \Lambda_{n+1}\{U_{j:n+1}\}.$$

For $m = 1$,

$$E(\Lambda_n\{U_{j:n+1}\}|\mathscr{F}_{n+1}) = \frac{1}{n+1}\frac{n}{(n+1)R_{n+1}(U_{j:n+1})} \le \Lambda_{n+1}\{U_{j:n+1}\}.$$

This proves (16) and hence the theorem. □

It follows by Proposition 5-3-11 of Neveu (1975) that $S_n$ converges almost surely to a limit, say $S$, which is measurable with respect to the $\sigma$-field $\mathscr{F}_\infty = \bigcap_{n\ge 1}\mathscr{F}_n$. If we take $\varphi(x) = I(-\infty, x]$, then $S_n = \Lambda_n(x)$. Since according to Woodroofe (1985) for every $x < b_{F^*}$,

$$\Lambda_n(x) \to \Lambda(x) \quad \text{in probability as } n \to \infty,$$

it must be true that $|\Lambda_n(x) - \Lambda(x)| \to 0$ a.s. By noting that $\Lambda_n(x)$ is monotone in $x$ for each fixed $n$ and $\Lambda(x)$ is continuous by assumption A1, we arrive at the following result of uniform a.s. convergence on semiclosed intervals, that is, $\forall\ b < b_{F^*}$,

$$\sup_{x\le b}|\Lambda_n(x) - \Lambda(x)| \to 0 \quad \text{a.s.} \tag{17}$$

COROLLARY 3.2. *Define* $\tilde{F}_n(x) = 1 - \exp(-\Lambda_n(x))$, *then as* $n \to \infty$,

$$\sup_x |\tilde{F}_n(x) - F_0(x)| \to 0 \quad a.s.$$

PROOF. Applying (17), we obtain for $x < b_{F^*}$, $\tilde{F}_n(x) \to 1 - \exp(-\Lambda(x)) = F_0(x)$ a.s.

Since $F_0$ is continuous and $\tilde{F}_n(x)$ is a distribution function, it must be true that

$$\sup_{x\le b_{F^*}} |\tilde{F}_n(x) - F_0(x)| \to 0 \quad \text{a.s.}$$ □

COROLLARY 3.3. *For any nonnegative measurable* $\varphi(x)$:

(a) $S = \lim_{n\to\infty} ES_n$ *exists* (*possibly infinite*);
(b) $S_n \to S$ *a.s.*;
(c) *if* $S < \infty$, *then* $\{S_n\}$ *is uniformly integrable and* $E|S_n - S| \to 0$.

PROOF. According to the Hewitt–Savage zero–one law, $\mathscr{F}_\infty$ is trivial. Hence the results follow directly by Theorem 3.1 and Proposition 5-3-11 of Neveu (1975). □

LEMMA 3.4. *For measurable* $\varphi(x) \ge 0$,

$$S = \int \varphi(x)\, d\Lambda(x).$$

PROOF. By (14) and the definition of $R_n$ given by (4),

$$\begin{aligned} ES_n &= E\sum_{j=1}^{n}\varphi(U_j)\frac{1}{nR_n(U_j)} = E\varphi(U_n)\frac{1}{R_n(U_n)} \\ &= E\varphi(U_n)\frac{n}{\sum_{i=1}^{n-1} I(V_i \le U_n \le U_i) + 1} \\ &= \int \varphi(x)E\zeta_{n-1}(x)\, dF^*(x) \qquad \forall\, n \ge 1, \end{aligned}$$

where

$$\zeta_{n-1}(x) = \frac{n}{(n-1)R_{n-1}(x) + 1}.$$

Note that $\zeta_n(x) \in \mathscr{F}_n$.

We shall prove $\{\zeta_n(x), \mathscr{F}_n,\ n \ge 1\}$ is a reverse supermartingale. As before, put $I_k = I[U_{n+1} = U_{k:n+1}]$. For $m = 1, 2, \dots, n+1$, if $\{(n+1)R_{n+1}(x) = m\}$ occurs, then

$$E(\zeta_n(x)|\mathscr{F}_{n+1}) = \sum_{k=1}^{n+1} E(\zeta_n(x)I_k|\mathscr{F}_{n+1}) = \frac{n+2}{m+1} = \zeta_{n+1}(x).$$

If $\{(n+1)R_{n+1}(x) = 0\}$ occurs, then

$$E[\zeta_n(x)|\mathscr{F}_{n+1}] = n + 1 \le n + 2 = \zeta_{n+1}(x),$$

so that $E[\zeta_n(x)|\mathscr{F}_{n+1}] \le \zeta_{n+1}(x)$. Since $R_n(x) \to R(x)$ a.s., we have $E\zeta_n(x) \uparrow 1/R(x)$.

Now, by (12),

$$S = \lim_{n\to\infty} ES_n = \int \varphi(x)\frac{1}{R(x)}\, dF^*(x) = \int \varphi(x)\, d\Lambda(x). \qquad \square$$

**4. The strong law for $F_n$.** Assumptions A1 and A2 are imposed throughout the section.

THEOREM 4.1. *Let $F_n$, $G_n$ and $F_0$ be defined by* (5) *and* (2), *then as $n \to \infty$,*

$$\sup_x |F_n(x) - F_0(x)| \to 0 \quad \textit{and} \quad \sup_x |G_n(x) - G_0(x)| \to 0 \quad \textit{a.s.}$$

PROOF. We prove convergence only for $F_n$. The proof for $G_n$ is similar and is omitted. Put $\varphi_n = (\log n/n)^{1/2}$ and $\beta_n = \inf\{x\colon G(x) \ge \varphi_n\}$. By the definitions of $\Lambda$ and $\Lambda_n$ [see (10) and (11)], $F_n(x) \le \Lambda_n(x)$ and $F_0(x) \le \Lambda(x)$ for every $x \in (a_{F^*}, b_{F^*})$. Thus the almost sure convergence of $F_n(\beta_n)$ and $F_0(\beta_n)$ follows from that of $\Lambda_n(\beta_n)$ and $\Lambda(\beta_n)$ [see (17)]. But then $\sup_{x\le\beta_n}|F_n(x) - F_0(x)| \le F_n(\beta_n) + F_0(\beta_n) \to 0$. It remains to consider the case $x > \beta_n$.

Split the product form of $F_0$ and $\bar{F}_n = 1 - F_n$ defined by (5) as follows. Let

$$M_n(x) = \prod_{\beta_n < U_i \le x}\left[1 - \frac{1}{nR_n(U_i)}\right], \qquad \tilde{M}_n(x) = \exp\left\{-\int_{\beta_n}^{x} d\Lambda(s)\right\}.$$

Then

$$\bar{F}_n(x) = (1 - F_n(\beta_n))M_n(x), \qquad \bar{F}_0(x) = (1 - F_0(\beta_n))\tilde{M}_n(x).$$

Hence $|F_n(x) - F_0(x)| \le |F_n(\beta_n) - F_0(\beta_n)| + |M_n(x) - \tilde{M}_n(x)|$. It suffices to prove that

$$\sup_x |M_n(x) - \tilde{M}_n(x)| \to 0 \quad \text{a.s.} \tag{18}$$

Note that for every $a \in (a_{F^*}, b_{F^*})$ and for $R(x)$ given by (9),

$$\inf_{\beta_n < s \le a} R(s) \ge \alpha^{-1}\bar{F}(a)G(\beta_n) = \alpha^{-1}\bar{F}(a)\varphi_n > 0.$$

Applying Theorem 2.1.4B in Serfling (1980), we have

$$\begin{aligned}
&\lim_{n\to\infty} \sup_{\beta_n < x \le a} \left| \frac{R_n(x)}{R(x)} - 1 \right| \\
&\quad \le \lim_{n\to\infty} \frac{\alpha}{\bar{F}(a)\varphi_n} \sup_x |G_n^*(x) - F_n^*(x-) - G^*(x) + F^*(x-)| = 0 \quad \text{a.s.}
\end{aligned}$$

So, $\forall\ x \in (a_{F^*}, b_{F^*})$, with probability 1 for large $n$,

$$\inf_{\beta_n < s \le x} nR_n(s) \ge \tfrac{1}{2} n \inf_{\beta_n < s \le x} R(s) \ge M_0\sqrt{n \log n},$$

where $M_0$ is a positive constant. By Taylor's expansion,

$$\begin{aligned}
M_n(x) &= \exp\left( -\int_{\beta_n}^x \frac{dF_n^*}{R_n(s)} + O\left( \frac{n}{n \log n} \right) \right) \\
&= \exp(-\Lambda_n(x) + \Lambda_n(\beta_n) + O(1/\log n)) \\
&\to 1 - F_0(x) \quad \text{a.s.}
\end{aligned}$$

and

$$\tilde{M}_n(x) = \exp(-\Lambda(x) + \Lambda(\beta_n)) \to 1 - F_0(x).$$

Consequently, $\forall\ x \in (a_{F^*}, b_{F^*})$,

$$|M_n(x) - \tilde{M}_n(x)| \le |M_n(x) - \bar{F}_0(x)| + |\tilde{M}_n(x) - \bar{F}_0(x)| \to 0 \quad \text{a.s.}$$

The convergence must be uniform in $x$, since $M_n(x)$ and $\tilde{M}_n(x)$ are monotone in $x$ and bounded above by 1. □

LEMMA 4.2. *Suppose $\varphi(x)$ is a nonnegative measurable function satisfying*

$$\int \varphi(x) \frac{1}{1 - F_0(x)}\, dF_0(x) < \infty.$$

*Then* $\lim_{n\to\infty} \int \varphi(x)\, dF_n(x) = \int \varphi(x)\, dF_0(x)$ *a.s.*

PROOF. By Corollary 3.3, Lemma 3.4 and Theorem 4.1,

$$\begin{aligned}\lim_{n\to\infty}\int \varphi(x) dF_n(x) &= \lim_{n\to\infty}\int \varphi(x)[\bar F_n(x-) - \bar F_0(x)]\, d\Lambda_n(x)\\ &\quad + \lim_{n\to\infty}\int \varphi(x)\bar F_0(x)\, d\Lambda_n(x)\\ &= \int \varphi(x)\bar F_0(x)\, d\Lambda(x) = \int \varphi(x)\, dF_0(x) \quad \text{a.s.} \qquad \square\end{aligned}$$

The next theorem relaxes the condition of finiteness of the integral of Lemma 4.2.

THEOREM 4.3. *For any nonnegative measurable* $\varphi(x)$,

$$\lim_{n\to\infty}\int \varphi(x)\, dF_n(x) = \int \varphi(x)\, dF_0(x) \quad a.s.$$

PROOF. If $\int \varphi(x)\, dF_0(x) = \infty$, then using

$$\lim_{n\to\infty}\inf_{x\le a}\frac{\bar F_n(x-)}{\bar F_0(x)} = 1 \quad \text{for any } a < b_{F^*},$$

we obtain

$$\liminf_{n\to\infty}\int \varphi(x)\, dF_n(x) \ge \tfrac{1}{2}\liminf_{n\to\infty}\int_{-\infty}^{a} \varphi(x)\bar F_0(x)\, d\Lambda_n(x) = \tfrac{1}{2}\int_{-\infty}^{a}\varphi(x)\, dF_0(x).$$

Letting $a \to \infty$, we obtain

$$\lim_{n\to\infty}\int \varphi(x)\, dF_n(x) = \infty.$$

Suppose $\int \varphi(x)\, dF_0(x) < \infty$. Then for any $a \in (a_{F^*}, b_{F^*})$,

$$\int_{-\infty}^{a}\varphi(x)\frac{1}{1-F_0(x)}\, dF_0(x) < \infty.$$

By Lemma 4.2 and Theorem 4.1 we have

$$\liminf_{n\to\infty}\int \varphi(x)\, dF_n(x) \ge \lim_{n\to\infty}\int_{-\infty}^{a}\varphi(x)\, dF_n(x) = \int_{-\infty}^{a}\varphi(x)\, dF_0(x) \quad \text{a.s.}$$

So the theorem will be proved if we show

$$\limsup_{n\to\infty}\int \varphi(x)\, dF_n(x) \le \int \varphi(x)\, dF_0(x) \quad \text{a.s.} \tag{19}$$

Now, according to Theorem 2.2 and Corollary 2.4 of He and Yang (1998),

$$\hat\alpha_n = \frac{G_n(x)\bar F_n(x-)}{R_n(x)} \to \alpha_0 = \int G_0(x)\, dF_0(x) > 0 \quad \text{a.s.}$$

and $\hat{\alpha}_n$ is a constant in $x$. See also (9). This in conjunction with (10) and (12) yields

$$\begin{aligned}\int_a^\infty \varphi(x)\,dF_n(x) &= \hat{\alpha}_n \int_a^\infty \varphi(x)(G_n(x))^{-1}\,dF_n^*(x)\\ &\le (\hat{\alpha}_n/G_n(a))\int_a^\infty \varphi(x)\,dF_n^*(x)\\ &\to \frac{1}{G_0(a)}\int_a^\infty \varphi(x)G_0(x)\,dF_0(x).\end{aligned}$$

Hence,

$$\begin{aligned}\limsup_{n\to\infty}\int \varphi(x)\,dF_n(x) &= \limsup_{n\to\infty}\left[\int_{-\infty}^a \varphi(x)\,dF_n(x) + \int_a^\infty \varphi(x)\,dF_n(x)\right]\\ &\le \int_{-\infty}^a \varphi(x)\,dF_0(x) + \frac{1}{G_0(a)}\int_a^\infty \varphi(x)\,dF_0(x).\end{aligned}$$

Letting $a \uparrow \infty$, we obtain (19). □

**5. Arbitrary distribution functions *F* and *G*.** The results of Sections 3 and 4 are derived under the continuity assumption of $F$ and $G$ and A2. We shall prove in this section that these results remain true if the continuity assumption A1 is removed.

It is convenient to partition the real line $R$ into $A$ and its complement $A^c$ where $A = \{x_j;\ j = 1, 2, \ldots\}$ is the set of discontinuity points of either $F$ or $G$ or both. For our purpose, we may assume, without loss of generality, $F_0 = F$ and $G_0 = G$. This is because the random truncation model $H(x, y)$ is the same regardless of whether the underlying distributions are $(F, G)$ or $(F_0, G_0)$ as shown by Woodroofe (1985). This assumption simplifies the discussion considerably. Let $\{\varepsilon_i, \eta_j\colon i, j = 1, 2, \ldots\}$ be a set of iid uniform random variables on the interval [0,1]. Assume that they are independent of the $X$'s and $Y$'s. Following Major and Rejtö (1988), we use the function

$$h(x) = x + \sum_{k\,:\,x_k<x} k^{-2}, \qquad x \in R = (-\infty, \infty) \tag{20}$$

to transform $X_i$ to $\hat{X}_i$ and $Y_i$ to $\hat{Y}_i$ as follows:

$$\hat{X}_i = \begin{cases} h(X_i), & \text{if } X_i \in A^c,\\ h(x_j) + \dfrac{1}{2j^2}(1+\varepsilon_i), & \text{if } X_i = x_j \in A,\end{cases} \qquad i = 1, 2, \ldots,$$

$$\hat{Y}_i = \begin{cases} h(Y_i), & \text{if } Y_i \in A^c,\\ h(x_j) + \dfrac{1}{2j^2}\eta_i, & \text{if } Y_i = x_j \in A,\end{cases} \qquad i = 1, 2, \ldots.$$

By construction, $h$ is strictly increasing and continuous on $A^c$. Adding independent uniform random variables on $A$ makes the transformed random variables $\hat{X}_i$ and $\hat{Y}_i$ having continuous distribution functions $\hat{F}(x) = P(\hat{X}_i \le x)$

and $\hat{G}(x) = P(\hat{Y}_i \le x)$. Although the continuity of $\hat{F}$ and $\hat{G}$ is conceptually clear, to establish the strong law with respect to the product-limit estimate $F_n$ in Theorem 5.7 it is necessary to explicitly express the relationship between the events determined by $X_i$, $Y_i$ and those by the transformed variables $\hat{X}_i$, $\hat{Y}_i$. This is given in Lemma 5.1.

Applying Lemma 5.1, it is easy to prove that $\hat{F}(x)$ and $\hat{G}(x)$ are continuous in $x$. Moreover, we show that

$$F(x) = \hat{F}(h(x+)), G(x) = \hat{G}(h(x+)) \qquad \forall\ x \in R. \tag{21}$$

Suppose we have randomly truncated data of $(\hat{X}, \hat{Y})$. Denote by $\hat{F}_n$ and $\hat{G}_n$ the nonparametric MLE of $\hat{F}$ and $\hat{G}$ calculated from the observed pairs $(\hat{U}_i, \hat{V}_i)$, $i = 1, 2, \dots, n$; see (24), (25). The generalization of the strong law in Theorem 4.3 for possibly discontinuous $F$ and $G$ will make use of the fact that

$$F_n(x) = \hat{F}_n(h(x+)) \qquad \forall\ x \in R. \tag{22}$$

as will be shown in Lemma 5.4.

By definition, $h(x)$ is left continuous and has a jump of size $j^{-2}$ at $x_j$ for $j = 1, 2, \dots$ . Let $\Delta = \bigcup_j [h(x_j), h(x_j+)]$. The function $h$ is a 1–1 map from $A^c$ onto $\Delta^c = R - \Delta$.

Split $\Delta_j = [h(x_j), h(x_j+)] = \Delta_{j,1} \cup \Delta_{j,2}$ where the intervals $\Delta_{j,1} = [h(x_j), h(x_j) + 1/2j^2)$, $\Delta_{j,2} = [h(x_j) + 1/2j^2, h(x_j+)]$. Then $\Delta = \bigcup_j \Delta_j$ and the $\Delta_j$ are disjoint. Define the inverse

$$h^{-1}(u) = \sup\{x\colon h(x) \le u\}, \qquad u \in R. \tag{23}$$

Then for $u \in \Delta_j$, $h^{-1}(u) = x_j$.

The next lemma expresses the events $[\hat{X}_i \le u]$ and $[\hat{Y}_i \le u]$ in terms of the original variables $X_i$ and $Y_i$.

LEMMA 5.1.

(a)
$$[\hat{X}_i \le u] = \begin{cases} [X_i \le h^{-1}(u)], & \text{if } u \in \Delta^c, \\ [X_i < x_j] \cup [\varepsilon_i \le 2j^2(u - h(x_j)) - 1] \\ \qquad \cap [X_i = x_j], & \text{if } u \in \Delta_j, \end{cases}$$

(b)
$$[\hat{Y}_i \le u] = \begin{cases} [Y_i \le h^{-1}(u)], & \text{if } u \in \Delta^c \\ [Y_i < x_j] \cup \{[\eta_i \le 2j^2(u - h(x_j))] \\ \qquad \cap [Y_i = x_j]\}, & \text{if } u \in \Delta_j. \end{cases}$$

(c) (a) *and* (b) *remain valid if all the inequalities less than or equal to are replaced by the strict inequalities.*

For the proof, see the Appendix.

A direct consequence of Lemma 5.1 is the corollary.

COROLLARY 5.2. *Set* $c_j = u - h(x_j)$.

(a) $$\hat{F}(u) = P(\hat{X}_i \le u) = \begin{cases} F(h^{-1}(u)), & \text{if } u \in \Delta^c, \\ F(x_j-), & \text{if } u \in \Delta_{j,1}, \\ F(x_j-) + (2j^2 c_j - 1)F\{x_j\}, & \text{if } u \in \Delta_{j,2}, \end{cases}$$

(b) $$\hat{G}(u) = P(\hat{Y}_i \le u) = \begin{cases} G(h^{-1}(u)), & \text{if } u \in \Delta^c, \\ G(x_j-) + 2j^2 c_j G\{x_j\}, & \text{if } u \in \Delta_{j,1}, \\ G(x_j), & \text{if } u \in \Delta_{j,2}. \end{cases}$$

(c) $$\hat{F}(u-) = P(\hat{X}_i < u) = \hat{F}(u),\ \hat{G}(u-) = P(\hat{Y}_i < u) = \hat{F}(u) \qquad \forall\ u \in R.$$

To facilitate the computation, the next corollary expresses, in reverse direction, $X, Y, F, G$ in terms of the transformed variables $\hat{X}, \hat{Y}$ and their distributions $\hat{F}, \hat{G}$.

COROLLARY 5.3.

(a) *For* $u \in A^c$, $[X_i \le u] = [\hat{X}_i \le h(u)]$, $[X_i < u] = [\hat{X}_i < h(u)]$,
$$[Y_i \le u] = [\hat{Y}_i \le h(u)], \qquad [X_i = u] = [\hat{X}_i = h(u)].$$

(b) *For* $x_j \in A$,
$$[X_i < x_j] = [\hat{X}_i < h(x_j)] = \left[\hat{X}_i < h(x_j) + \frac{1}{2j^2}\right],$$
$$[Y_i \le x_j] = \left[\hat{Y}_i \le h(x_j) + \frac{1}{2j^2}\right] = [\hat{Y}_i < h(x_j+)],$$
$$[X_i = x_j] = [\hat{X}_i \in \Delta_j] = [\hat{X}_i \in \Delta_{j,2}].$$

(c) $$h^{-1}(\hat{X}_j) = X_j,\ h^{-1}(\hat{Y}_j) = Y_j,\ j = 1, 2, \dots,$$

(d) $$F(u) = \hat{F}(h(u+)), \qquad G(u) = \hat{G}(h(u+)),\ \forall\ u \in R,$$
$$F(u-) = \hat{F}(h(u)), \qquad G(u-) = \hat{G}(h(u)),\ \forall\ u \in R,$$

(e) $$[\hat{X}_i \ge \hat{Y}_i] = [X_i \ge Y_i], \qquad i = 1, 2, \dots.$$

PROOF. Except for (e), the proof can be deduced directly from Lemma 5.1. For (e) we apply (a) and (b). It follows that

$$\begin{aligned} [X_i \ge Y_i] &= [X_i \ge Y_i,\ X_i \notin A] \bigcup_j [x_j \ge Y_i,\ X_i = x_j] \\ &= [h(X_i) \ge \hat{Y}_i,\ X_i \notin A] \bigcup_j \left[h(x_j) + \frac{1}{2j^2} \ge \hat{Y}_i,\ X_i = x_j\right] \\ &= [\hat{X}_i \ge \hat{Y}_i,\ X_i \notin A] \bigcup_j [\hat{X}_i \ge \hat{Y}_i,\ X_i = x_j] = [\hat{X}_i \ge \hat{Y}_i]. \qquad \square \end{aligned}$$

The product limit estimates of $\hat{F}_n$ and $\hat{G}_n$ are defined by

$$\hat{F}_n(x) = 1 - \prod_{s \le x} \left(1 - \frac{\#\{i: \hat{U}_i = s\}}{\#\{i: \hat{V}_i \le s \le \hat{U}_i\}}\right) \tag{24}$$

and

$$\hat{G}_n(x) = \prod_{s > x} \left(1 - \frac{\#\{i: \hat{V}_i = s\}}{\#\{i: \hat{V}_i \le s \le \hat{U}_i\}}\right), \tag{25}$$

where $\#\{\cdot\}$ denotes the number of $i$ satisfying the stated condition.

LEMMA 5.4. *For any* $x \in R$, $F_n(x) = \hat{F}_n(h(x+))$, $G_n(x) = \hat{G}_n(h(x+))$.

For the proof, see the Appendix.

The behavior of the distributions at the lower boundaries requires special attention. It will be shown in Theorem 5.6 that the MLE $F_n$ given by (5) converges to either $F_a$ or $F$, depending on whether $F$ and $G$ have jumps at the lower boundaries $a_F$ and $a_G$. The following lemma provides a detailed analysis of the boundaries.

LEMMA 5.5.

(a) $h(a_F) \le a_{\hat{F}} \le h(a_F+)$, $\quad h(a_G) \le a_{\hat{G}} \le h(a_G+)$;
(b) $a_G < a_F$ *implies* $a_{\hat{G}} < a_{\hat{F}}$;
(c) *if* $a_G = a_F$, *a necessary and sufficient condition for* $a_{\hat{G}} > a_{\hat{F}}$ *is*

$$G\{a_G\} = 0 \quad \text{and} \quad F\{a_F\} > 0,$$

*and in this case* $a_{\hat{G}} = h(a_G+)$.

For the proof, see the Appendix.

As defined in the Introduction,

$$F_a(x) = P[X \le x | X > a_G]. \tag{26}$$

THEOREM 5.6. *Suppose* $a_G \le a_F$.

(a) *If* $a_{\hat{F}} \ge a_{\hat{G}}$ *we have*

$$\sup_x |F_n(x) - F(x)| \to 0 \quad a.s.$$

(b) *If* $a_{\hat{F}} < a_{\hat{G}}$, *we have*

$$\sup_x |F_n(x) - F_a(x)| \to 0 \quad a.s.$$

PROOF. To avoid triviality, we assume that $F\{a_F\} < 1$. Since $\hat{F}$ and $\hat{G}$ are continuous, by Theorem 4.1,

$$\sup_x |\hat{F}_n(x) - \hat{F}_0(x)| \to 0 \quad \text{a.s.,} \tag{27}$$

where

$$\hat{F}_0(x) = P[\hat{X}_i \le x | \hat{X}_i \ge a_{\hat{G}}]. \tag{28}$$

Note that $a_{\hat{F}} \geq a_{\hat{G}}$ implies $\hat{F}(h(x+)) = \hat{F}_0(h(x+))$ for all $x$. For (a), using (d) of Corollary 5.3 and Lemma 5.4, we have

$$\sup_x |F_n(x) - F(x)| = \sup_x |\hat{F}_n(h(x+)) - \hat{F}_0(h(x+))| = \sup_t |\hat{F}_n(t) - \hat{F}_0(t)| \to 0.$$

(b) If $a_{\hat{F}} < a_{\hat{G}}$, we show that $F_a(x) = \hat{F}_0(h(x+))$ for all $x \in R$. Consider first the case $x < a_G$. Then by Lemma 5.5, $h(x+) \leq h(a_G+) = a_{\hat{G}}$. Thus

$$\hat{F}_0(h(x+)) \leq P(\hat{X}_i \leq h(a_G+) | \hat{X}_i \geq a_{\hat{G}}) = P(\hat{X}_i \leq a_{\hat{G}} | \hat{X}_i \geq a_{\hat{G}}) = 0.$$

Consequently, $F_a(x) = \hat{F}_0(h(x+)) = 0$.

For $x \geq a_G$, we have

$$\begin{aligned} F_a(x) = \frac{F(x) - F(a_G)}{1 - F(a_G)} &= \frac{\hat{F}(h(x+)) - \hat{F}(a_{\hat{G}})}{1 - \hat{F}(a_{\hat{G}})} \\ &= P(\hat{X}_i \leq h(x+) | \hat{X}_i \geq a_{\hat{G}}) = \hat{F}_0(h(x+)). \end{aligned}$$

Then (b) follows from (27) and a similar proof used in (a) with $F$ replaced by $F_a$. □

We now prove the strong law for arbitrary $F$ and $G$.

THEOREM 5.7. *Suppose $a_F \geq a_G$ and $\varphi(x)$ is any nonnegative measurable function.*

(a) *If $a_{\hat{F}} \geq a_{\hat{G}}$, we have*

$$\lim \int \varphi \, dF_n = \int \varphi \, dF \quad a.s.$$

(b) *If $a_{\hat{F}} < a_{\hat{G}}$, then*

$$\lim \int \varphi \, dF_n = \int \varphi \, dF_a \quad a.s.$$

PROOF. Assume that $F\{a_F\} < 1$ to avoid triviality. Let $X_{(1)}, X_{(2)}, \ldots, X_{(\nu)}$ be the distinct values of $X_1, X_2, \ldots, X_{m_n}$ and $Y_{(1)}, Y_{(2)}, \ldots, Y_{(\nu)}$ be their concomitants. By Lemma 5.4 and Corollary 5.3 we have

$$\begin{aligned} \int \varphi(x) \, dF_n(x) &= \int \varphi(x) \, d\hat{F}_n(h(x+)) \\ &= \sum_{j=1}^{\nu} \varphi(X_{(j)}) \{\hat{F}_n(h(X_{(j)}+)) - \hat{F}_n(h(X_{(j)})-)\} I[X_{(j)} \geq Y_{(j)}] \\ &= \sum_{j:\hat{X}_{(j)} \in \Delta^c} \varphi(h^{-1}(\hat{X}_{(j)})) \{\hat{F}_n(\hat{X}_{(j)}) - \hat{F}_n(\hat{X}_{(j)}-)\} I[\hat{X}_{(j)} \geq \hat{Y}_{(j)}] \\ &\quad + \sum_{j=1}^{\nu} \sum_k \int_{\Delta_k} \varphi(h^{-1}(u)) \, d\hat{F}_n(u) I[X_{(j)} = x_k \geq Y_{(j)}] \end{aligned} \tag{29}$$

$$= \int_{\Delta^c} \varphi(h^{-1}(u))\, d\hat{F}_n(u) + \int_{\Delta} \varphi(h^{-1}(u))\, d\hat{F}_n(u)$$
$$= \int \varphi(h^{-1}(u))\, d\hat{F}_n(u).$$

Here we have used the fact that $h^{-1}(u) = x_j$ for $x_j \in \Delta_j$ and $[X_{(j)} = x_k \geq Y_{(j)}] = [\hat{X}_{(j)} \in \Delta_k,\ \hat{X}_{(j)} \geq \hat{Y}_{(j)}]$.

Finally, we are ready to complete the proof of the theorem.

(a) Using Theorem 4.3 and Corollary 5.3, we have

$$\lim_{n\to\infty} \int \varphi\, dF_n = \int \varphi(h^{-1}(u))\, d\hat{F}(u) = E\varphi(h^{-1}(\hat{X}_j)) = E\varphi(X_j) = \int \varphi\, dF.$$

(b) Using Theorem 4.3 and $a_{\hat{G}} = h(a_G+)$, we have from (29) that

$$\lim_{n\to\infty} \int \varphi\, dF_n = \int \varphi(h^{-1}(u))\, d\hat{F}_0(u) = \int_{a_{\hat{G}}}^{\infty} \varphi(h^{-1}(u)) \frac{d\hat{F}(u)}{1 - \hat{F}(a_{\hat{G}})}.$$

Note that

$$[u > a_{\hat{G}}] = [u > h(a_G+)] = [h^{-1}(u) > a_G].$$

Therefore,

$$\int_{a_{\hat{G}}}^{\infty} \varphi(h^{-1}(u)) \frac{d\hat{F}(u)}{1 - \hat{F}(a_{\hat{G}})} = \int \varphi(h^{-1}(u)) I[h^{-1}(u) > a_G]\, d\hat{F}(u)/(1 - \hat{F}(a_{\hat{G}}))$$
$$= E\varphi(h^{-1}(\hat{X}_i)) I[h^{-1}(\hat{X}_i) > a_G]/(1 - F(a_G))$$
$$= E\varphi(X_i) I[X_i > a_G]/(1 - F(a_G))$$
$$= \int \varphi(u)\, dF_a(u). \qquad \square$$

Recall Condition B1: $a_F = a_G$, $G\{a_G\} = 0$ and $F\{a_F\} > 0$ in the Introduction. The case where B1 does not hold will be called Condition B2. Note that in what follows we do not suppose $a_G \leq a_F$ and $b_G \leq b_F$, as is customarily assumed in the literature.

COROLLARY 5.8. *Let $\varphi$ be any nonnegative measurable function. If Condition* B2 *holds*,

$$\lim_{n\to\infty} \sup_x |F_n(x) - F_0(x)| = 0 \quad \textit{and} \quad \lim_{n\to\infty} \int \varphi\, dF_n = \int \varphi\, dF_0 \quad \textit{a.s.}$$

*If Condition* B1 *holds*,

$$\lim_{n\to\infty} \sup_x |F_n(x) - F_a(x)| = 0 \quad \textit{and} \quad \lim_{n\to\infty} \int \varphi\, dF_n = \int \varphi\, dF_a \quad \textit{a.s.}$$

Similarly, let $G_b(x) = P[Y \leq x | Y < b_{F_0}]$. We say Condition C1 holds if $b_{F_0} = b_{G_0}$, $G_0\{b_{G_0}\} > 0$, $F_0\{b_{F_0}\} = 0$. The complement of Condition C1 is called Condition C2.

COROLLARY 5.9. *Let $\varphi(x)$ be any nonnegative measurable function. Then under Condition* C2,

$$\lim_{n\to\infty}\sup_x |G_n(x)-G_0(x)| = 0 \quad and \quad \lim_{n\to\infty}\int \varphi\, dG_n = \int \varphi\, dG_0 \quad a.s.$$

*Under Condition* C1,

$$\lim_{n\to\infty}\sup_x |G_n(x)-G_b(x)| = 0 \quad and \quad \lim_{n\to\infty}\int \varphi\, dG_n = \int \varphi\, dG_b \quad a.s. \qquad \square$$

## APPENDIX

**Proofs.**

PROOF OF LEMMA 5.1. Let $\Delta_j^0$ and $\Delta_{j,2}^0$ be the interior of $\Delta_j$ and $\Delta_{j,2}$. Let $\bar{\Delta}_{j,1}$ be the closure of $\Delta_{j,1}$. To ease the notation, put $c_j = u - h(x_j)$, $\Omega_j = [\varepsilon_i \le 2j^2c_j - 1] \cap [X_i = x_j]$ and let $A - B$ denote $A \cap B^c$ for any sets $A$ and $B$. We have

$$\begin{aligned}
[\hat{X}_i \le u] &= [\hat{X}_i \le u,\ X_i \notin A] \bigcup_j [h(x_j) + (1+\varepsilon_i)/2j^2 \le u,\ X_i = x_j] \\
&= [\hat{X}_i \le u,\ X_i \notin A] \bigcup_{j:\, c_j \ge j^{-2}} [X_i = x_j] \bigcup_{j:\, 0<c_j<j^{-2}} [\varepsilon_i \le 2j^2c_j - 1,\ X_i = x_j] \\
&= [h(X_i) \le u,\ X_i \notin A]\Bigg(\bigcup_{j:\, h(x_j)\le u} [X_i = x_j] - \bigcup_{j:\, u-j^{-2}<h(x_j)\le u} [X_i = x_j]\Bigg) \\
&\quad + \bigcup_{j:\, u\in\Delta_j^0} \Omega_j \\
&= \Bigg\{[h(X_i) \le u] - \Bigg(\bigcup_{j:\, u\in\bar{\Delta}_{j,1}} [X_i = x_j] + \bigcup_{j:\, u\in\Delta_{j,2}^0} [X_i = x_j]\Bigg)\Bigg\} + \bigcup_{j:\, u\in\Delta_j^0} \Omega_j \\
&= \begin{cases} [X_i \le h^{-1}(u)], & \text{if } u \in \Delta^c, \\ [X_i < x_j] \cup \{[\varepsilon_i \le 2j^2c_j - 1] \cap [X_i = x_j]\}, & \text{if } u \in \Delta_j. \end{cases}
\end{aligned}$$

The proofs for (b) and (c) are similar. $\square$

PROOF OF LEMMA 5.4. Write $F_n$ [defined by (5)] in terms of $X$ and $Y$,

$$1 - F_n(x) = \prod_{s\le x}\left(1 - \frac{\#\{i\colon X_i = s,\ X_i \ge Y_i,\ 1 \le i \le m_n\}}{\#\{i\colon Y_i \le s \le X_i,\ 1 \le i \le m_n\}}\right) = \prod_{s\le x}\left[1 - \frac{n_s}{D_s}\right],$$

where $m_n$ is defined at the beginning of Section 2. Note that the above formula is self-adjusting for tied observations. We now apply Corollary 5.3. We shall, subject to $s \le x$, treat the cases $s \notin A$ and $s \in A$ separately. The numerator

$$(30) \qquad n_s = \begin{cases} \#\{i\colon \hat{X}_i = h(s),\ \hat{X}_i \ge \hat{Y}_i,\ 1 \le i \le m_n\}, & \text{if } s \notin A, \\ \#\{i\colon \hat{X}_i \in \Delta_{j,2},\ \hat{X}_i \ge \hat{Y}_i,\ 1 \le i \le m_n\}, & \text{if } s = x_j \in A. \end{cases}$$

The denominator

$$D_s = \begin{cases} \#\{i: \hat{Y}_i \le h(s) \le \hat{X}_i,\ 1 \le i \le m_n\}, & \text{if } s \notin A, \\ \#\{i: \hat{Y}_i \le h(s_j) + j^{-2}/2 \le \hat{X}_i,\ 1 \le i \le m_n\}, & \text{if } s = x_j \in A. \end{cases} \tag{31}$$

Note that

$$\begin{aligned} &1 - \frac{\#\{i: \hat{X}_i \in \Delta_{j,2},\ \hat{X}_i \ge \hat{Y}_i,\ 1 \le i \le m_n\}}{\#\{i: \hat{Y}_i \le h(x_j) + (1/2j^2) \le \hat{X}_i,\ 1 \le i \le m_n\}} \\ &\quad = \prod_{u \in \Delta_{j,2}} \left(1 - \frac{\#\{i: \hat{X}_i = u,\ \hat{X}_i \ge \hat{Y}_i,\ 1 \le i \le m_n\}}{\#\{i: \hat{Y}_i \le u \le \hat{X}_i,\ 1 \le i \le m_n\}}\right). \end{aligned} \tag{32}$$

It follows from (32) that

$$\begin{aligned} 1 - F_n(x) &= \prod_{\substack{u \le h(x) \\ u \in \Delta^c}} \left(1 - \frac{\#\{i: \hat{U}_i = u\}}{\#\{i: \hat{V}_i \le u \le \hat{U}_i\}}\right) \prod_{\substack{u \le h(x+) \\ u \in \Delta}} \left(1 - \frac{\#\{i: \hat{U}_i = u\}}{\#\{i: \hat{V}_i \le u \le \hat{U}_i\}}\right) \\ &= 1 - \hat{F}_n(h(x+)). \end{aligned}$$

The proof of $G_n(x) = \hat{G}_n(h(x+))$ is similar. □

PROOF OF LEMMA 5.5. (a) By definition of $\hat{X}_i$, we have $h(X_i) \le \hat{X}_i \le h(X_i+)$. This implies that $h(a_F) \le a_{\hat{F}} \le h(a_F+)$. Similarly, we can prove $h(a_G) \le a_{\hat{G}} \le h(a_G+)$.

(b) It follows from (a) that $a_{\hat{G}} \le h(a_G+) < h(a_F) \le a_{\hat{F}}$. Hence (b) is true.

(c) If $G\{a_G\} = 0$, $F\{a_F\} > 0$, then $a_F \in A$. This implies that $a_{\hat{F}} = h(a_F) + 1/2k^2$. Now

$$P[\hat{Y}_i \ge h(a_G+)] \ge P[Y_i > a_G,\ \hat{Y}_i \ge h(a_G+)] = P[Y_i > a_G] = 1.$$

Hence $a_{\hat{G}} \ge h(a_G+)$, and it follows that $a_{\hat{G}} > a_{\hat{F}}$. On the other hand, suppose $G\{a_G\} > 0$. Then, $\forall\ \lambda > 0$ we have

$$P[\hat{Y}_i < h(a_G) + \lambda] \ge P[Y_i = a_G,\ \hat{Y}_i < h(a_G) + \lambda] = G\{a_G\} P[\eta_i < 2k^2\lambda] > 0,$$

so that, $a_{\hat{G}} = h(a_G) \le a_{\hat{F}}$.

If $F\{a_F\} = 0$, we can suppose $G\{a_F\} = 0$. So, $a_{\hat{G}} = h(a_F) \le a_{\hat{F}}$. □

**Acknowledgments.** We thank a referee for very helpful comments and suggestions. They are incorporated in the companion paper in this issue.

## REFERENCES

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

CHEN, K., CHAO, M. T. and LO, S.-H. (1995). On strong uniform consistency of the Lynden-Ball estimator for truncated data. *Ann. Statist.* **23** 440–449.

FEIGELSON, E. D. and BABU, G. J., eds. (1992). *Statistical Challenges in Modern Astronomy*. Springer, New York.

HE, S. and YANG, G. L. (1994). Estimating a lifetime distribution under different sampling plans. In *Statistical Decision Theory and Related Topics* (S. S. Gupta and J. O. Berger, eds.) **5** 73–85. Springer, New York.

HE, S. and YANG, G. L. (1998). Estimation of the truncation probability in the random truncation model. *Ann. Statist.* **26** 1011–1027.

MAJOR, P. and REJTÖ, L. (1988). Strong embedding of the estimator of the distribution function under random censorship. *Ann. Statist.* **16** 1113–1132.

NEVEU, J. (1975). *Discrete-Parameter Martingales*. North-Holland, Amsterdam.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

STUTE, W. and WANG, J. L. (1993). The strong law under random censorship. *Ann. Statist.* **21** 1591–1607.

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., JEWELL, N. P. and TSAI, W.-Y. (1986). Asymptotic properties of the product limit estimate under random truncation. *Ann. Statist.* **14** 1597–1605.

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

DEPARTMENT OF PROBABILITY AND STATISTICS
PEKING UNIVERSITY
BEIJING
CHINA

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND 20742
E-MAIL: gly@math.umd.edu