The Annals of Statistics
1998, Vol. 26, No. 3, 1011–1027

ESTIMATION OF THE TRUNCATION PROBABILITY IN THE RANDOM TRUNCATION MODEL[1]

BY SHUYUAN HE AND GRACE L. YANG

Peking University and University of Maryland

Under random truncation, a pair of independent random variables X and Y is observable only if X is larger than Y. The resulting model is the conditional probability distribution $H(x, y) = P[X \le x, Y \le y | X \ge Y]$. For the truncation probability $\alpha = P[X \ge Y]$, a proper estimate is not the sample proportion but $\alpha_n = \int G_n(s)\, dF_n(s)$ where F_n and G_n are product limit estimates of the distribution functions F and G of X and Y, respectivly. We obtain a much simpler representation $\hat{\alpha}_n$ for α_n. With this, the strong consistency, an iid representation (and hence asymptotic normality), and a LIL for the estimate are established. The results are true for arbitrary F and G. The continuity restriction on F and G often imposed in the literature is not necessary. Furthermore, the representation $\hat{\alpha}_n$ of α_n facilitates the establishment of the strong law for the product limit estimates F_n and G_n.

1. Introduction. Let X and Y be two independent random variables having distribution functions $F(x)$ and $G(x)$, respectively. Consider an infinite sequence of independent random vectors $\{(X_m, Y_m);\ m = 1, 2, \ldots\}$ where X_m and Y_m are independently distributed as X and Y. For each m the pair (X_m, Y_m) is observable only when $X_m \ge Y_m$. Thus the observable random variables are a subsequence of $\{(X_m, Y_m);\ m = 1, 2, \ldots\}$. It is convenient to denote the observable subsequence by $\{(U_j, V_j);\ j = 1, 2, \ldots\}$ with $U_j \ge V_j$. The random vectors (U_j, V_j) are iid; however, the components of each vector are dependent. Here and after, (U, V) refers to any pair of (U_j, V_j). The random truncation model is defined by the joint distribution $H(x, y)$ of (U, V) as

$$H(x, y) = P[U \le x,\ V \le y] = P[X \le x,\ Y \le y | X \ge Y] \tag{1.1}$$

with marginal distributions,

$$F^*(x) = P[U \le x] = H(x, \infty) = \frac{1}{\alpha}\int_{-\infty}^{x} G(s)\, dF(s), \tag{1.2}$$

$$G^*(x) = P[V \le x] = H(\infty, x) = \frac{1}{\alpha}\int_{-\infty}^{x} \bar{F}(s-)\, dG(s), \tag{1.3}$$

where

$$\alpha = P[X \ge Y] = \int G(s)\, dF(s). \tag{1.4}$$

Received March 1996; revised January 1998.

[1]Supported in part by ONR Grant N00014-92-J-1097 and National Natural Science Foundation of China Grant 19571002.

AMS 1991 *subject classification*. 62G05.

Key words and phrases. Random truncation, nonparametric estimation, product-limit, truncation probability, strong consistency, LIL, iid representation.

The integral sign $\int_a^b$ stands for $\int_{(a,b]}$. The integral sign without limits refers to integration from $-\infty$ to $+\infty$.

The general problem is to draw statistical inference about the unknown F and G based on a sample of n iid random vectors $\{(U_j, V_j);\ j = 1, 2, \dots, n\}$ from $\{(X_m, Y_m);\ m = 1, 2, \dots, m_n\}$, where the given $n \leq m_n$ and m_n is unknown.

In the companion paper, He and Yang (1998) in the same issue of this journal, we prove the strong law of large numbers for the product limit estimate F_n given in (2.4). Under the same assumptions used in the companion paper, we address in this paper the estimation of the truncation probability,

$$\alpha = P[X \geq Y] = \int G(s)\, dF(s). \tag{1.5}$$

The problem is, of course, trivial if we have an iid sample from the original untruncated (X, Y)-data, $(X_1, Y_1), (X_2, Y_2), \dots, (X_{m_n}, Y_{m_n})$, with a known sample size m_n. Then the sample proportion of those (X_k, Y_k) with $X_k \geq Y_k$ is an optimal nonparametric estimate of α. However, under random truncation, any pair (X, Y) for which $X < Y$ is missing, and it is not known how many pairs are missing in the sample because m_n is unknown. Thus it is not at all clear that reasonable estimates for α can be found.

Equation (1.4) suggests estimating α by

$$\alpha_n = \int G_n(s)\, dF_n(s), \tag{1.6}$$

provided good estimates F_n and G_n for F and G can be obtained.

Random truncation restricts the observation range of X and Y. Only $F_0(x) = P[X \leq x | X \geq a_G]$ and $G_0(x) = P[Y \leq x | Y \leq b_F]$ can be estimated; see Woodroofe (1985), where

$$a_F = \inf\{x\colon F(x) > 0\} \quad \text{and} \quad b_F = \sup\{x\colon F(x) < 1\}$$

are the lower and upper boundaries of the support of the distribution of X. Let a_G and b_G be similarly defined.

This leads us to the introduction of the following parameter:

$$\alpha_0 = \int G_0(s)\, dF_0(s). \tag{1.7}$$

If $a_G \leq a_F$ and $b_G \leq b_F$, then $F_0 = F$, $G_0 = G$ and $\alpha_0 = \alpha$. Under these conditions and the continuity of F and G, Woodroofe (1985) proved that if F_n and G_n are product-limit estimates [given by (2.4) below], α_n converges in probability to α as $n \to \infty$. Under similar conditions, the asymptotic normality of $\sqrt{n}(\alpha_n - \alpha)$ has been investigated by several authors using different methods. Chao (1987) used influence curves and Keiding and Gill (1990) used finite Markov processes and the δ-method.

Since F_n and G_n have complicated product-limit forms, it is generally not easy to study the properties of α_n. We propose, instead, to use the relationship

$$R(x) = P[V \leq x \leq U] = P[Y \leq x \leq X | X \geq Y] = \alpha^{-1} G(x) \bar{F}(x-) \tag{1.8}$$

to obtain an estimating equation for α as

$$\alpha = G(x)\bar{F}(x-)/R(x). \tag{1.9}$$

Replacing F and G by the respective product limit estimates yields another estimate of α as

$$\hat{\alpha}_n = G_n(x)\bar{F}_n(x-)/R_n(x) \tag{1.10}$$

for all x such that $R_n(x) > 0$. [Defined by (2.3).]

An important result of this paper is that if F_n and G_n are the product-limit estimates of F_0 and G_0 defined by (2.4) below, then $\hat{\alpha}_n$ and α_n are equal. In particular, $\hat{\alpha}_n$ is independent of x, provided $R_n(x) > 0$. The proof of equivalence is presented in Section 2. It is worth noting that the equivalence is not derived from integration-by-parts. The advantage of $\hat{\alpha}_n$ over α_n is its simpler form, which makes the analysis easier and enables us to obtain further properties of the estimate. Using $\hat{\alpha}_n$, we prove in Section 3 the almost sure convergence of the estimate to α_0 and obtain a manageable iid representation for $\hat{\alpha}_n$ and a LIL. The iid representation yields immediately the asymptotic normality of the estimate.

The iid representation for $\hat{\alpha}_n$ is deduced from that of F_n and G_n. Several iid representations for F_n (and G_n) are available in the literature with different remainder terms; see Chao and Lo (1988), Gu and Lai (1990) and Stute (1993). We shall use Stute's representation, which is derived under the condition that

$$\int \frac{dF}{G^2} < \infty \quad \text{and} \quad \int \frac{dG}{\bar{F}^2} < \infty. \tag{1.11}$$

It has a sufficiently higher order remainder term of $O(\ln^3 n/n)$ that suits our purpose. This condition can be weakened to

$$\int \frac{dF}{G} < \infty \quad \text{and} \quad \int \frac{dG}{\bar{F}} < \infty, \tag{1.12}$$

provided the tails of estimates F_n and G_n are properly modified. Under (1.12), the remainder term is of lower order than $O(\log^3 n/n)$ but still good enough to yield the asymptotic normality for F_n at the rate $\sqrt{n}$ and a LIL, as shown by Gu and Lai (1990). Based on these, we obtain similar results for a modified $\hat{\alpha}_n$.

Results in Section 3 are established under the continuity of F and G but are true for discrete F and G as well. The generalization to arbitrary F and G is given in Section 4.

By construction, $\hat{\alpha}_n$ and α_n inherit asymptotic properties of F_n and G_n. Conversely, good behavior of $\hat{\alpha}_n$ induces nice properties in F_n and G_n. As shown in He and Yang (1998), the almost sure convergence of $\hat{\alpha}_n$ to α_0 leads to the SLLN for F_n in the sense that

$$\int \varphi \, dF_n \to \int \varphi \, dF_0$$

for any integrable φ.

2. The equivalence of $\hat{\alpha}_n$ and α_n. To avoid triviality, we shall always assume $a_G < b_F$, which ensures that $\alpha > 0$. In what follows, for any real monotone function g, the left continuous version of $g(s)$ is denoted by $g(s-)$ or $g_{-}(s)$, and the difference $g(s) - g(s-)$ by the curly brackets $g\{s\}$. The convergence is "with respect to $n \to \infty$" unless specified otherwise.

LEMMA 2.1. *Let α_0 be given by* (1.7) *and α by* (1.5). *Then $\alpha_0 \geq \alpha$. A necessary and sufficient condition for $\alpha_0 = \alpha$ is*

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

PROOF. For $x \in [a_G, b_F]$, we have

$$G_0(x) = P(Y \leq x)/P(Y \leq b_F), \qquad \bar{F}_0(x-) = P(X \geq x)/P(X \geq a_G).$$

Hence, it follows from (1.8) and Lemma 1 of Woodroofe (1985) that

$$\alpha_0 = \alpha[G(b_F)\bar{F}(a_G-)]^{-1}. \qquad \square$$

Let $I[A]$ denote the indicator function of the event A. Let F_n^*, G_n^* and R_n be the empirical distributions defined by

$$F_n^*(s) = n^{-1}\sum_{i=1}^{n} I[U_i \leq s], \qquad G_n^*(s) = n^{-1}\sum_{i=1}^{n} I[V_i \leq s], \tag{2.2}$$

$$R_n(s) = G_n^*(s) - F_n^*(s-) = n^{-1}\sum_{i=1}^{n} I[V_i \leq s \leq U_i]. \tag{2.3}$$

The well-known product limit estimates of F_0 and G_0 are defined by

$$F_n(x) = 1 - \prod_{s \leq x}\left(1 - \frac{F_n^*\{s\}}{R_n(s)}\right) \quad \text{and} \quad G_n(x) = \prod_{s > x}\left(1 - \frac{G_n^*\{s\}}{R_n(s)}\right), \tag{2.4}$$

where an empty product is set equal to 1. For construction of these estimates, see Woodroofe (1985) or Wang, Jewell and Tsai (1986).

The estimates F_n and G_n are step functions. The jumps of F_n occur at the distinct order statistics $U_{(1)} < U_{(2)} < \cdots < U_{(r)}$ of the sample $U_1, U_2, \ldots, U_n$ with jump size at $U_{(j)}$ (using our brackets notation) given by

$$F_n\{U_{(j)}\} = \prod_{i<j}\left(1 - \frac{F_n^*\{U_{(i)}\}}{R_n(U_{(i)})}\right)\frac{F_n^*\{U_{(j)}\}}{R_n(U_{(j)})}. \tag{2.5}$$

A similar expression for G_n can be determined from (2.4) where G_n jumps at the distinct order statistics $V_{(1)} < V_{(2)} < \cdots < V_{(q)}$ of $V_1, V_2, \ldots, V_n$.

We need to study these jumps in order to prove the following equivalence theorem.

THEOREM 2.2. *Let F_n and G_n be the product limit estimates given by* (2.4). *Let $\hat{\alpha}_n$ be defined by* (1.10) *and α_n by* (1.6). *Then $\alpha_n = \hat{\alpha}_n$, for any x such that $R_n(x) > 0$.*

PROOF. The case $\alpha_n = 0$ is easy. Note that $\alpha_n = 0$ if and only if $b_{F_n} < a_{G_n}$ or $b_{F_n} = a_{G_n}$ and $(1 - F_n(b_{F_n}-))G_n(a_{G_n}) = 0$. This is equivalent to $\hat{\alpha}_n = 0$ for all x.

Now suppose $\alpha_n > 0$. We introduce two independent random variables Z and W which have distributions $F_n(x)$ and $G_n(x)$, respectively. Then

$$\alpha_n = P[Z \geq W] = \int G_n \, dF_n.$$

The integral

$$\begin{aligned} \int_{\infty}^{x} G_n(t)\, dF_n(t) &= \sum_{j=1}^{r} G_n(U_{(j)}) F_n\{U_{(j)}\} I[U_{(j)} \leq x] \\ &= \sum_{j=1}^{r} \zeta_{n,j} F_n^*\{U_{(j)}\} I[U_{(j)} \leq x], \end{aligned} \tag{2.6}$$

where $\zeta_{n,j} = G_n(U_{(j)}) F_n\{U_{(j)}\}/F_n^*\{U_{(j)}\}$.

In Lemma 2.3 below we show that for n fixed, $\zeta_{n,j}$ is a constant in j, say ζ_0. By setting $x = \infty$ in (2.6) we obtain

$$\alpha_n = \int G_n(s)\, dF_n(s) = \zeta_0 \sum_j F_n^*\{U_{(j)}\} = \zeta_0. \tag{2.7}$$

Consequently, the conditional distribution

$$P(Z \leq x | Z \geq W) = \alpha_n^{-1} \int_{-\infty}^{x} G_n(t)\, dF_n(t) = \alpha_n^{-1} \zeta_0 F_n^*(x) = F_n^*(x). \tag{2.8}$$

By symmetry,

$$G_n^*(x) = P[W \leq x | Z \geq W].$$

Therefore, $R_n(x) = G_n^*(x) - F_n^*(x-) = P(W \leq x \leq Z | Z \geq W) = \alpha_n^{-1} G_n(x) \bar{F}_n(x-)$; that is,

$$\alpha_n = \frac{G_n(x)\bar{F}_n(x-)}{R_n(x)} = \hat{\alpha}_n$$

for all x such that $R_n(x) > 0$. □

REMARK. Once we have reached (2.8), we could use integration-by-parts to complete the proof. However, it is simpler to use random variables Z and W. Then the result follows immediately by symmetry. Note also that although α_n is an MLE, it is not obvious that $\hat{\alpha}_n$ is an MLE, since it is $F_n(x)$ and not $F_n(x-)$, that is, the MLE of F. Therefore we cannot use the MLE argument to claim that $\alpha_n = \hat{\alpha}_n$.

LEMMA 2.3. *Let F_n and G_n be the product-limit estimates given by* (2.4). *Let $\zeta_{n,j} = G_n(U_{(j)}) F_n\{U_{(j)}\}/F_n^*\{U_{(j)}\}$ as in* (2.6). *Then for any fixed n, $\zeta_{n,j} = \zeta_{n,1}$, for $j = 2, \ldots, n$.*

PROOF. Substituting (2.5) for $F_n\{U_{(j)}\}$ in $\zeta_{n,j}$ gives

$$\zeta_{n,j} = \prod_{k=1}^{q}\left(1 - \frac{G_n^*\{V_{(k)}\}I[V_{(k)} > U_{(j)}]}{R_n(V_{(k)})}\right)\prod_{i<j}\left(1 - \frac{F_n^*\{U_{(i)}\}}{R_n(U_{(i)})}\right)\frac{1}{R_n(U_{(j)})}.$$

We show that the differences $\zeta_{n,j} - \zeta_{n,j-1} = 0$ for all j. Write the difference as a product

$$\zeta_{n,j} - \zeta_{n,j-1} = \{A_{n,j}\}\{B_{n,j}\},$$

where

$$A_{n,j} = \prod_{k=1}^{q}\left(1 - \frac{G_n^*\{V_{(k)}\}I[V_{(k)} > U_{(j)}]}{R_n(V_{(k)})}\right)\prod_{i<j-1}\left(1 - \frac{F_n^*\{U_{(i)}\}}{R_n(U_{(i)})}\right),$$

$$B_{n,j} = \left(1 - \frac{F_n^*\{U_{(j-1)}\}}{R_n(U_{(j-1)})}\right)\frac{1}{R_n(U_{(j)})} - \frac{1}{R_n(U_{(j-1)})}\prod_{l}\left(1 - \frac{G_n^*\{V_{(l)}\}I[U_{(j-1)} < V_{(l)} \le U_{(j)}]}{R_n(V_{(l)})}\right).$$

We are going to show that $B_{n,j} = 0$, which proves the lemma.

Put $h = \sum_l I[U_{(j-1)} < V_{(l)} \le U_{(j)}]$, which is the total number of V's lying in the interval $(U_{(j-1)}, U_{(j)}]$.

If $h = 0$, then $R_n(U_{(j)}) = R_n(U_{(j-1)}) - F_n^*\{U_{(j-1)}\}$. It follows that

$$B_{n,j} = \frac{R_n(U_{(j-1)}) - F_n^*\{U_{(j-1)}\}}{R_n(U_{(j-1)})}\frac{1}{R_n(U_{(j)})} - \frac{1}{R_n(U_{(j-1)})} = 0.$$

If $h > 0$, let us denote by $V'_{(1)}, V'_{(2)}, \dots, V'_{(h)}$ the distinct ordered values of V_j in $(U_{(j-1)}, U_{(j)}]$, that is,

$$U_{(j-1)} < V'_{(1)} < V'_{(2)} < \cdots < V'_{(h)} \le U_{(j)}.$$

Then,

$$\begin{aligned}\prod_{l}\left(1 - \frac{G_n^*\{V_{(l)}\}I[U_{(j-1)} < V_{(l)} \le U_{(j)}]}{R_n(V_{(l)})}\right) &= \prod_{l=1}^{h}\left(1 - \frac{G_n^*\{V'_{(l)}\}}{R_n(V'_{(l)})}\right)\\ &= \frac{R_n(V'_{(1)}-)}{R_n(V'_{(h)})}\\ &= \frac{G_n^*(U_{(j-1)}) - F_n^*(U_{(j-1)})}{G_n^*(V'_{(h)}) - F_n^*(U_{(j-1)})}.\end{aligned}$$

This implies that

$$B_{n,j} = \frac{R_n(U_{(j-1)}) - F_n^*\{U_{(j-1)}\}}{R_n(U_{(j-1)})}\frac{1}{R_n(U_{(j)})} - \frac{R_n(U_{(j-1)}) - F_n^*\{U_{(j-1)}\}}{R_n(V'_{(h)})}\frac{1}{R_n(U_{(j-1)})} = 0.$$

The last equality follows from $R_n(U_{(j)}) = R_n(V'_{(h)})$. This is because if $V'_{(h)} < U_{(j)}$, then $R_n(U_{(j)}) = G_n^*(U_{(j)}) - F_n^*(U_{(j)}-) = G_n^*(V'_{(h)}) - F_n^*(V'_{(h)}-) = R_n(V'_{(h)})$. $\square$

COROLLARY 2.4.

$$\hat{\alpha}_n = \frac{G_n(U_{(j)})\bar{F}_n\{U_{(j)}-\}}{R_n(U_{(j)})} = \frac{G_n(V_{(j)})\bar{F}_n\{V_{(j)}-\}}{R_n(V_{(j)})}, \qquad j = 1, 2, \dots, n.$$

The next corollary follows either by Theorem 4.1 of He and Yang (1998), or by applying the uniform strong convergence of F_n [see Chen, Chao and Lo (1994)].

COROLLARY 2.5. *As* $n \to \infty$,

$$\hat{\alpha}_n \to \alpha_0 \quad a.s.$$

REMARK. If we use $S_n(x) = \exp(-\int_{-\infty}^x dF_n^*/R_n)$ to estimate $1 - F_0(x)$, and $\tilde{Q}_n(x) = \exp(-\int_x^\infty dG_n^*/R_n)$ to estimate $G_0(x)$ then, by Corollary 3.2 of He and Yang (1998), for any x such that $R_n(x) > 0$,

$$c_n = \frac{\tilde{Q}_n(x)S_n(x-)}{R_n(x)}$$

is a strong consistent estimate of α_0.

3. The CLT and the LIL for $\boldsymbol{\alpha_n}$. Since α_n and $\hat{\alpha}_n$ are equivalent, the known result of asymptotic normality of $\sqrt{n}(\alpha_n - \alpha_0)$ applies to $\sqrt{n}(\hat{\alpha}_n - \alpha_0)$. On the other hand, the simple form of $\hat{\alpha}_n$ makes it possible to obtain an iid representation from which the asymptotic normality of $\hat{\alpha}_n$ follows immediately. Moreover, an LIL can be obtained.

Let F_n and G_n be defined by (2.4). Applying Theorem 2 of Stute (1993) yields the following iid representation for F_n and G_n. This result is needed for deriving the iid representation for $\hat{\alpha}_n$ as given in Theorem 3.2. It is also of independent interest.

LEMMA 3.1. *If F and G are continuous such that*

$$\int_{a_G}^{\infty} \frac{dF(s)}{G^2(s)} < \infty \quad and \quad \int_{-\infty}^{b_F} \frac{dG(s)}{\bar{F}^2(s)} < \infty \tag{3.1}$$

then for $x \in (a_G, b_F)$, we have the following:

(i) $F_n(x) = F_0(x) + \bar{F}_0(x)((1/n)\sum_{i=1}^n Z_i(x)) + O(\log^3 n/n)$, *a.s.*
(ii) $G_n(x) = G_0(x) - G_0(x)((1/n)\sum_{i=1}^n W_i(x)) + O(\log^3 n/n)$, *a.s.*

where

$$Z_i(x) = \frac{I[U_i \le x]}{R(U_i)} - \int_{-\infty}^{x} \frac{I[V_i \le s \le U_i]}{R^2(s)}\, dF^*(s), \qquad i = 1, 2, \dots, n,$$

are iid random variables with

$$EZ_i(x) = 0, \qquad \mathrm{Var}(Z_i(x)) = \int_{a_{F^*}}^{x} \frac{dF^*(s)}{R^2(s)}$$

and

$$W_i(x) = \frac{I[V_i > x]}{R(V_i)} - \int_x^{\infty} \frac{I[V_i \le s \le U_i]}{R^2(s)}\, dG^*(s), \qquad i = 1, 2, \ldots, n,$$

are iid random variables with

$$EW_i(x) = 0, \qquad \mathrm{Var}(W_i(x)) = \int_x^{b_{G^*}} \frac{dG^*(s)}{R^2(s)}.$$

PROOF. We prove only (i), because (ii) can be proved by symmetry.

It is easy to see that $\int_{a_G}^{\infty} dF(s)/G^2(s) < \infty$ if and only if $\int dF_0(s)/G_0^2(s) < \infty$, and $\int_{-\infty}^{b_F} dG(s)/\bar{F}^2(s) < \infty$ if and only if $\int dG_0(s)/\bar{F}_0^2(s) < \infty$. By Theorem 2 of Stute (1993),

$$F_n(x) - F_0(x) = \bar{F}_0(x) L_n(x) + O\left(\frac{\log^3 n}{n}\right), \quad \text{a.s.,}$$

with

$$L_n(x) = \int_{a_F}^{x} \frac{dF_n^*(s)}{R(s)} - \int_{a_F}^{x} \frac{R_n(s)}{R^2(s)}\, dF^*(s) = \frac{1}{n} \sum_{i=1}^{n} Z_i(x), \quad \text{a.s.}$$

Direct computation yields

$$EZ_i(x) = 0,$$

and

$$\begin{aligned} \mathrm{Var}(Z_i(x)) = \int_{a_{F^*}}^{x} \frac{dF^*}{R^2} &+ \int_{a_{F^*}}^{x} \int_{a_{F^*}}^{x} \frac{EI(s)I(t)}{R^2(s)}\, dF^*(s) \frac{1}{R^2(t)}\, dF^*(t) \\ &- 2 \int_{a_{F^*}}^{x} E\left(\frac{I[U \le x] I(s)}{R(U)}\right) \frac{dF^*(s)}{R^2(s)} = \int_{a_{F^*}}^{x} \frac{dF^*}{R^2}, \end{aligned}$$

where $I(s) = I[V \le s \le U]$. $\square$

REMARK. Evaluation of integrals similar to the above will be carried out in the proof of the next theorem.

THEOREM 3.2. *Under the assumptions of Lemma* 3.1, *as* $n \to \infty$, $\sqrt{n}(\hat{\alpha}_n - \alpha_0)$ *converges weakly to the normal distribution* $N(0, \sigma^2)$, *and with probability* 1, *the sequence* $\{\sqrt{n/2 \log\log n}(\hat{\alpha}_n - \alpha_0);\ n \ge 7\}$ *is relatively compact with its set of limit points* $[-\sigma, \sigma]$, *where*

$$\sigma^2 = \alpha_0^2 \left\{ \int_{a_{F^*}}^{x} \frac{dF^*(s)}{R^2(s)} + \int_x^{b_{G^*}} \frac{dG^*(s)}{R^2(s)} - \frac{1}{R(x)} + 2\alpha_0 - 1 \right\}$$

for $x \in (a_{G^*}, b_{F^*})$, *is a positive constant.*

PROOF. Using Lemma 3.1 and the LIL for iid partial sums, we obtain $\forall\ x \in (a_{G^*}, b_{F^*})$, with probability 1 for n large:

$$\begin{aligned}\hat{\alpha}_n - \alpha_0 &= \frac{G_n(x)(1-F_n(x-))}{R_n(x)} - \frac{G_0(x)(1-F_0(x))}{R(x)} \\ &= \frac{\bar{F}_0(x)R(x)G_0(x)}{R_n(x)R(x)}\bigg\{-\frac{1}{n}\sum_{i=1}^n W_i(x) - \frac{1}{n}\sum_{i=1}^n Z_i(x) \\ &\qquad - \frac{1}{nR(x)}\sum_{i=1}^n (I[V_i \le x \le U_i] - R(x))\bigg\} \\ &\quad + O\bigg(\frac{\log^3 n}{n}\bigg) \\ &= -\alpha_0 \frac{1}{n}\sum_{i=1}^n \zeta_i(x) + O\bigg(\frac{\log^3 n}{n}\bigg) \quad \text{a.s.,}\end{aligned}$$

where

$$\zeta_i(x) = W_i(x) + Z_i(x) + \frac{1}{R(x)}\big(I[V_i \le x \le U_i] - R(x)\big), \qquad i = 1, 2, \dots, \tag{3.2}$$

is a sequence of iid random variables with mean zero. The theorem follows by the classical CLT and the LIL for partial sums of an iid sequence, if we can show that

$$\mathrm{Var}(\zeta_i(x)) = \sigma^2 \qquad \forall\ x \in (a_{G^*}, b_{F^*}) \tag{3.3}$$

is a positive constant. This requires calculations of the moments and cross-product moments of Z_i, W_i and $I[V_i \le s \le U_i]$. The calculation is not hard but tedious. We shall give some key steps only. To proceed, we suppress the subscript i from these variables for simplicity. Put $T(s) = I(s)/R(s) - 1$. Then

$$\begin{aligned}E(T(x))^2 &= \frac{1}{R(x)} - 1, \\ EZ(x)W(x) &= \int_{a_{F^*}}^x \int_x^{b_{G^*}} \frac{E(I(s)I(t))}{R^2(t)}\, dG^*(t) \frac{1}{R^2(s)}\, dF^*(s) \\ &= \frac{1}{\alpha}\int_{a_{F^*}}^x \int_x^{b_{G^*}} \frac{G(s)\bar{F}(t)}{R^2(t)}\, dG^*(t) \frac{1}{R^2(s)}\, dF^*(s) \\ &= -\alpha\bigg(\frac{1}{G(b_{G^*})} - \frac{1}{G(x)}\bigg)\bigg(\frac{1}{\bar{F}(x)} - \frac{1}{\bar{F}(a_{F^*})}\bigg)\end{aligned}$$

and

$$\begin{aligned}E(Z(x)+W(x))T(x) &= -\frac{\alpha}{G(x)}\bigg(\frac{1}{\bar{F}(x)} - \frac{1}{\bar{F}(a_{F^*})}\bigg) \\ &\quad + \frac{\alpha}{\bar{F}(x)}\bigg(\frac{1}{G(b_{G^*})} - \frac{1}{G(x)}\bigg).\end{aligned}$$

Using the moments of Z_i and W_i given in Lemma 3.1, we obtain

$$\mathrm{Var}(\zeta(x)) = \int_{a_{F^*}}^{x} \frac{dF^*}{R^2} + \int_{x}^{b_{G^*}} \frac{dG^*}{R^2} - \frac{1}{R(x)} + 2\alpha_0 - 1,$$

where we used the fact that $\alpha_0 = \alpha[G(b_{G^*})\bar{F}(a_{F^*})]^{-1} = \alpha[G(b_F)\bar{F}(a_G)]^{-1}$ [see the definition of G^* and F^* in (1.2), (1.3)]. Obviously, $\zeta(x)$ is not a constant and therefore $\sigma^2 > 0$. Then $\forall\ x, y \in (a_{G^*}, b_{F^*})$, the difference

$$\mathrm{Var}(\zeta(x)) - \mathrm{Var}(\zeta(y)) = \int_{x}^{y} \frac{dR}{R^2} + \frac{1}{R(y)} - \frac{1}{R(x)} = 0.$$

Hence, $\sigma^2 = E(\zeta(x))^2$ for $x \in (a_{G^*}, b_{F^*})$ is a positive constant. $\square$

REMARK. The random variable $\zeta(x)$ does not depend on x. This can be seen from the following representation:

$$\zeta(x) = \frac{1}{R(V)} - \int_{a_{G^*}}^{b_{G^*}} \frac{I(s)}{R^2(s)}\, dG^*(s) - 1 \quad \text{a.s.}$$

or

$$\zeta(x) = \frac{1}{R(U)} - \int_{a_{F^*}}^{b_{F^*}} \frac{I(s)}{R^2(s)}\, dF^*(s) - 1 \quad \text{a.s.}$$

A necessary and sufficient condition for $\sigma^2 < \infty$ is

$$\int_{a_G}^{\infty} \frac{dF}{G} < \infty \quad \text{and} \quad \int_{-\infty}^{b_F} \frac{dG}{\bar{F}} < \infty, \tag{3.4}$$

which is weaker than (3.1). To obtain results similar to Lemma 3.1 and Theorem 3.2 under the weaker condition (3.4) we can use a modified form $\tilde{\alpha}_n$ of α_n. The modification is necessary to avoid singularities at the boundaries of X. It is constructed based on the modified estimates $\tilde{F}_n$, $\tilde{G}_n$, of F_n, G_n proposed by Gu and Lai (1990) as given below:

$$\tilde{F}_n(x) = 1 - \prod_{i:\, U_i \le x} \left(1 - \frac{I[G_n^*(U_i) \ge n^{\theta-1}]}{nR_n(U_i)}\right) \tag{3.5}$$

and

$$\tilde{G}_n(x) = \prod_{i:\, V_i > x} \left(1 - \frac{I[\bar{F}_n^*(V_i-) \ge n^{\theta-1}]}{nR_n(V_i)}\right) \tag{3.6}$$

for $\theta \in (1/3, 1/2)$.

Accordingly, the modified estimate $\tilde{\alpha}_n$ for α_0 is

$$\tilde{\alpha}_n = \frac{\tilde{G}_n(x)(1 - \tilde{F}_n(x-))}{R_n(x)}, \tag{3.7}$$

for any x such that $R_n(x) > 0$. To see how $\tilde{G}_n$ is constructed, put $\tilde{X} = -Y$ and $\tilde{Y} = -X$. Thus $X \ge Y$ is precisely $\tilde{X} \ge \tilde{Y}$. Therefore, $(\tilde{X}, \tilde{Y})$ is observable if and only if $X \ge Y$, and so $(\tilde{U}, \tilde{V}) = (-V, -U)$. Then the corresponding modified estimate for $P[\tilde{X} \leq x]$ based on the data $\{(\tilde{U}_j, \tilde{V}_j);\ j = 1, \ldots, n\}$ is, according to (3.5),

$$
\begin{aligned}
1 - \prod_{i:\tilde{U}_i \leq x} & \left(1 - \frac{I[\sum_j I[\tilde{V}_j \leq \tilde{U}_i] \geq n^\theta]}{\sum_j I[\tilde{V}_j \leq \tilde{U}_i \leq \tilde{U}_j]}\right) \\
&= 1 - \prod_{i:V_i \geq -x}\left(1 - \frac{I[\sum_j I[U_j \geq V_i] \geq n^\theta]}{\sum_j I[V_j \leq V_i \leq U_j]}\right) \\
&= 1 - \prod_{i:V_i \geq -x}\left(1 - \frac{I[\bar{F}_n^*(V_i-) \geq n^{\theta-1}]}{nR_n(V_i)}\right).
\end{aligned}
$$

However, by construction $P[\tilde{X} \leq x] = P[Y \geq -x] = 1 - G_-(-x)$. Therefore, $G(x)$ is estimated by

$$
\prod_{i:V_i > x}\left[1 - \frac{I[\bar{F}_n^*(V_i-) \geq n^{\theta-1}]}{nR_n(V_i)}\right].
$$

An iid representation for $\tilde{\alpha}_n$ under the weaker condition (3.4) is achieved at the cost of lowering the order of the remainder term. However, the order remains high enough to yield the weak convergence of $\sqrt{n}(\hat{\alpha}_n - \alpha_0)$ to normality and a LIL. The proof of this is more involved than that of Theorem 3.2.

Let $Z_i(x)$ and $W_i(x)$ be defined as in Lemma 3.1.

LEMMA 3.3. *Assume that F and G are continuous and satisfy the conditions* (3.4). *Then for $\theta \in (1/3, 1/2)$, $x \in (a_G, b_F)$, as $n \to \infty$ we have the following.*

(i) $\tilde{F}_n(x) = F_0(x) + \bar{F}_0(x)((1/n)\sum_{i=1}^n Z_i(x)) + O(\eta_n)$ *a.s.*
(ii) $\tilde{G}_n(x) = G_0(x) - G_0(x)((1/n)\sum_{i=1}^n W_i(x)) + O(\eta_n)$ *a.s.*

where $\eta_n = o(\phi(n))$ a.s., $\phi(n) = \sqrt{2\log\log n/n}$ and $\eta_n = o_p(1/\sqrt{n})$.

PROOF. We first prove (i) and then show that (ii) can be obtained by means of symmetry. Taking $q = \theta$ and $c = 1$ in Theorem 2 of Gu and Lai (1990), we have, with probability 1 as $n \to \infty$,

$$
\begin{aligned}
\tilde{F}_n(x) &- F_0(x) \\
&= \frac{1}{m_n}\sum_{j=1}^{m_n} \bar{F}_0(x) \int_{\tau_n}^x \frac{1}{G(u)\bar{F}(u)} \\
&\qquad\qquad \times d\left\{I[Y_j \leq X_j \leq u] - \int_{-\infty}^u I[X_j \geq s \geq Y_j]\frac{dF(s)}{\bar{F}(s)}\right\} \\
&\quad + O(n^{\theta-1}),
\end{aligned}
$$

where $\tau_n = \inf\{s: G_n^*(s) \geq n^{\theta-1}\}$, and m_n is defined in the Introduction. By (1.2) we have

$$\begin{aligned}\tilde{F}_n(x) - F_0(x) &= \frac{\bar{F}_0(x)}{\alpha m_n}\int_{\tau_n}^{x}\frac{1}{R(u)}d\left\{nF_n^*(u) - n\int_{-\infty}^{u}\frac{R_n(s)}{R(s)}\,dF^*(s)\right\}\\ &\quad + O(n^{\theta-1})\\ &= \frac{1}{n}\bar{F}_0(x)\sum_{i=1}^{n}Z_i(x) + \frac{1}{\alpha}\left(\frac{n}{m_n} - \alpha_0\right)\bar{F}_0(x)\frac{1}{n}\sum_{i=1}^{n}Z_i(x)\\ &\quad - \frac{n}{\alpha m_n}\bar{F}_0(x)\frac{1}{n}\sum_{i=1}^{n}Z_i(\tau_n) + O(n^{\theta-1})\\ &= \bar{F}_0(x)\frac{1}{n}\sum_{i=1}^{n}Z_i(x) + J_{1,n}(x) + J_{2,n}(x) + O(n^{\theta-1}).\end{aligned}$$

According to Corollary 4 of Gu and Lai (1990), for any $\delta \in (a_G, x]$,

$$\limsup_{n\to\infty}\frac{1}{\phi(n)}\sup_{a_G\leq s\leq\delta}\left|\frac{1}{n}\sum_{i=1}^{n}Z_i(s)\right| \leq \left(\int_{a_G}^{\delta}\frac{dF(s)}{\alpha\bar{F}^2(s)G(s)}\right)^{1/2}\quad\text{a.s.}$$

Now

$$\int_{a_G}^{\delta}\frac{dF(s)}{\alpha\bar{F}^2(s)G(s)} \leq \frac{1}{\bar{F}^2(x)}\int_{a_G}^{\delta}\frac{dF(s)}{G(s)} \to 0\quad\text{as }\delta\to a_G,$$

$n/m_n \to \alpha_0$ a.s, and the classical LIL implies that $J_{i,n}(x) = o(\phi(n))$, a.s. for $i = 1, 2$. On the other hand, the CLT and the Chebyshev inequality imply that $J_{i,n}(x) = o_p(1/\sqrt{n})$, for $i = 1, 2$. This completes the proof of (i). We now turn to the proof of (ii). As noted earlier,

$$Q(x) \equiv P[\tilde{X}_j \leq x] = 1 - G(-x),\qquad K(x) \equiv P[\tilde{Y}_j \leq x] = 1 - F(-x).$$

Thus $a_Q = -b_G$, $b_Q = -a_G$, $a_K = -b_F$, $b_K = -a_F$, and $a_G < b_F$ if and only if $-b_Q < -a_K$.

Therefore,

$$\int_{a_K}^{\infty}\frac{dQ(x)}{K(x)} = \int_{-b_F}^{\infty}\frac{-dG(-x)}{1-F(-x)} = \int_{-\infty}^{b_F}\frac{dG(x)}{1-F(x)} < \infty.$$

Since $(a_G, b_F) = (-b_Q, -a_K)$, so if $x \in (a_G, b_F)$ then $-x \in (a_K, b_Q)$. Setting $y = -x$ and applying (i), we obtain

$$\tilde{Q}_n(y) = Q_0(y) + (1 - Q_0(y))\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{Z}_i(y)\right) + O(\eta(n))$$

where $\tilde{Q}_n(y) = 1 - \tilde{G}_n((-y)-) = 1 - \tilde{G}_n(x-)$, $Q_0(y) = P[\tilde{X}_j \leq y | \tilde{X}_j \geq a_K] = \bar{G}_0(x-)$ and

$$\tilde{Z}_i(y) = \frac{I[\tilde{U}_i \leq y]}{\tilde{R}(\tilde{U}_i)} - \int_{a_K}^{y}\frac{I[\tilde{V}_i \leq s \leq \tilde{U}_i]}{\tilde{R}^2(s)}\,dQ^*(s),$$

with $\tilde{R}(s) = P[\tilde{V}_i \le s \le \tilde{U}_i] = R(-s)$ and $Q^*(s) = P[\tilde{U}_i \le s] = 1 - G^*(-s)$. Hence

$$\tilde{Z}_i(y) = \frac{I[V_i \ge x]}{R(V_i)} - \int_{-b_F}^{-x} \frac{I[V_i \le -s \le U_i]}{R^2(-s)}\, d(1 - G^*(-s)) = W_i(x-),$$

where W_i is given in Lemma 3.1.

Therefore

$$\tilde{G}_n(x-) = G_0(x-) - G_0(x-)\frac{1}{n}\sum_{i=1}^{n} W_i(x-) + O(\eta_n) \quad \text{a.s.} \qquad \square$$

The following theorem is proved the same way as Theorem 3.2.

THEOREM 3.4. *Suppose the assumptions of Lemma* 3.3 *hold*. *As* $n \to \infty$, $\sqrt{n}(\tilde{\alpha}_n - \alpha_0)$ *converges weakly to the normal distribution* $N(0, \sigma^2)$, *and with probability* 1, *the sequence* $\{\phi^{-1}(n)(\tilde{\alpha}_n - \alpha_0);\ n \ge 7\}$ *is relatively compact with its set of limit points* $[-\sigma, \sigma]$, *where* σ^2 *is defined in Theorem* 3.2.

4. Arbitrary F and G. We shall relax the continuity condition on F and G of Theorems 3.2 and 3.4. The results are given in Theorems 4.1 and 4.2. As noted by Woodroofe (1985), the truncation model $H(x, y)$ is the same if the underlying distributions F and G are replaced by F_0 and G_0. Thus for studying H, we may, without loss of generality, assume that $F_0 = F$ and $G_0 = G$. It follows that $\alpha_0 = \alpha$. This simplifies the discussion.

The proof for arbitrary F and G uses a technique of Major and Rejtö (1988). Namely, we transform X, Y to $\hat{X}, \hat{Y}$ via a certain specially constructed real function $h(x)$. The transformed random variables have continuous distribution functions to be denoted by $\hat{F}$ and $\hat{G}$. As such, the foregoing theorems apply to the product-limit estimates $\hat{F}_n, \hat{G}_n$ of $\hat{F}, \hat{G}$, where the product-limit estimates $\hat{F}_n, \hat{G}_n$ are computed, with (2.4), based on the transformed sample, $\hat{X}_i, \hat{Y}_i$ or $\hat{U}_i, \hat{V}_i$. It is shown in He and Yang (1998) that

$$F(u) = \hat{F}(h(u+)), \qquad G(u) = \hat{G}(h(u+)),$$
$$F_n(u) = \hat{F}_n(h(u+)), \qquad G_n(u) = \hat{G}_n(h(u+)) \quad \text{for any real number } u.$$

By way of this relationship, we show that the results of previous sections hold for arbitrary F and G. To proceed, we need to express R_n and α_n in terms of the transformed variables as well. Using the same notation as in He and Yang, we put the symbol ˆ on quantities derived from the transformed data and let $A = \{x_j;\ j \ge 1\}$ be the set of jump points of F and G.

If $a_{G^*} < b_{F^*}$, then for $x \in (a_{G^*}, b_{F^*}) - A$, with probability 1 for large n we have

$$\begin{aligned} 0 < R_n(x) &= \frac{1}{n}\sum_{i=1}^{m_n} I[Y_i \le x \le X_i] = \frac{1}{n}\sum_{i=1}^{m_n} I[\hat{Y}_i \le h(x) \le \hat{X}_i] \\ &= \frac{1}{n}\sum_{i=1}^{n} I[\hat{V}_i \le h(x) \le \hat{U}_i] \equiv \hat{R}_n(h(x)). \end{aligned} \tag{4.1}$$

Define $\hat{F}^*(x) = P(\hat{U}_i \le x)$, $\hat{G}^*(x) = P(\hat{V}_i \le x)$ and $\hat{R}(x) = \hat{G}^*(x) - \hat{F}^*(x-)$. It follows that $R(x) = \hat{R}(h(x))$ and

$$\hat{\alpha}_n = \frac{(1 - F_n(x-))G_n(x)}{R_n(x)} = \frac{(1 - \hat{F}_n(h(x)-))\hat{G}_n(h(x))}{\hat{R}_n(h(x))}. \tag{4.2}$$

Here we have used the fact that $h(x) = h(x+)$ for $x \in A^c$, the continuity set of F and G.

Parallel to Theorems 3.2 and 3.4, we arrive at the following general results in Theorems 4.1 and 4.2 for arbitrary F, G under condition (3.1) and (3.4), respectively. Note that for continuous F and G, the σ^2 formula in Theorem 4.1 coincides with that in Theorem 3.2.

THEOREM 4.1. *If*

$$\int \frac{dF(s)}{G^2(s)} < \infty, \qquad \int \frac{dG(s)}{(1 - F(s-))^2} < \infty \tag{4.3}$$

and $a_{G^*} < b_{F^*}$, *then as* $n \to \infty$, $\sqrt{n}(\hat{\alpha}_n - \alpha)$ *converges weakly to the normal distribution* $N(0, \sigma^2)$. *Moreover, with probability* 1, *the sequence* $\{\phi(n)^{-1}(\hat{\alpha}_n - \alpha);\ n \ge 7\}$ *is relatively compact with the set of its limit points* $[-\sigma, \sigma]$, *where*

$$\sigma^2 = \alpha^2 \left\{ \int_{-\infty}^{x} \frac{dF(s)}{R(s)\bar{F}(s)} + \int_{x}^{\infty} \frac{dG(s)}{R(s)G(s-)} - \frac{1}{R(x)} + 2\alpha - 1 \right\}$$

is a positive constant for $x \in (a_{G^*}, b_{F^*}) - A$.

PROOF. We first show that $\hat{F}_0 = \hat{F}$ and $\hat{G}_0 = \hat{G}$. Applying Corollary 5.3 of He and Yang (1998) gives

$$\begin{aligned} \int \frac{dF}{G^2} &= \int_{A^c} \frac{d\hat{F}(h(x))}{\hat{G}^2(h(x))} + \sum_j \frac{\hat{F}(h(x_j+)) - \hat{F}(h(x_j))}{\hat{G}^2(h(x_j+))} \\ &= \int_{\Delta^c} \frac{d\hat{F}}{\hat{G}^2} + \sum_j \left\{ \int_{\Delta_{j,1}} \frac{d\hat{F}}{\hat{G}^2} + \int_{\Delta_{j,2}} \frac{d\hat{F}}{\hat{G}^2} \right\} = \int \frac{d\hat{F}}{\hat{G}^2} < \infty, \end{aligned}$$

and similarly

$$\int \frac{dG}{(1 - F_-)^2} = \int \frac{d\hat{G}}{(1 - \hat{F})^2} < \infty,$$

where $\Delta^c, \Delta_{j,1}, \Delta_{j,2}$ are defined in Section 5 of He and Yang (1998). It follows that $a_{\hat{F}} \ge a_{\hat{G}}$ and $b_{\hat{G}} \le b_{\hat{F}}$. Hence $\hat{F}_0 = \hat{F}$, $\hat{G}_0 = \hat{G}$. By Theorem 3.2, (3.9) and the fact that $\alpha = P(X_i \ge Y_i) = P(\hat{X}_i \ge \hat{Y}_i)$, the weak convergence of $\sqrt{n}(\hat{\alpha}_n - \alpha)$ to $N(0, \sigma_1^2)$ follows. Also, with probability 1, the sequence $\{\phi(n)^{-1}(\hat{\alpha}_n - \alpha);\ n \ge 7\}$ is relatively compact with the set of its limit points $[-\sigma_1, \sigma_1]$, where

$$\sigma_1^2 = \alpha^2 \left\{ \int_{-\infty}^{h(x)} \frac{d\hat{F}^*}{\hat{R}^2} + \int_{h(x)}^{\infty} \frac{d\hat{G}^*}{\hat{R}^2} - \frac{1}{\hat{R}(h(x))} + 2\alpha - 1 \right\} \tag{4.4}$$

is a positive constant for $h(x) \in (a_{\hat{G}^*}, b_{\hat{F}^*})$.

It remains to prove that $\sigma^2 = \sigma_1^2$. For $x \in A^c$ with $R(x) > 0$, we know that $h(x) \in \Delta^c$ and

$$\hat{R}(h(x)) = P(\hat{V}_i \le h(x) \le \hat{U}_i) = \alpha^{-1} P(\hat{Y}_i \le h(x) \le \hat{X}_i) = R(x) > 0. \tag{4.5}$$

For $s \in \Delta_{j,2}$,

$$\alpha \hat{R}(s) = G(x_j)(1 - F(x_j-) - [2j^2(s - h(x_j)) - 1]F\{x_j\}),$$
$$\hat{F}(s) = F(x_j-) + [2j^2(s - h(x_j)) - 1]F\{x_j\}.$$

We show that the two integrals in σ_1^2 equal the corresponding ones in σ^2. The first integral is

$$\int_{-\infty}^{h(x)} \frac{d\hat{F}^*}{\hat{R}^2} = \int_{-\infty}^{h(x)} \frac{d\hat{F}}{\hat{R}(1-\hat{F})} = E\left(\frac{I[\hat{X} \le h(x)]}{\hat{R}(\hat{X})(1-\hat{F}(\hat{X}))}\right) \equiv B + D, \tag{4.6}$$

where

$$\begin{aligned} B &= E\left(\frac{I[\hat{X} \le h(x),\ X \in A^c]}{\hat{R}(\hat{X})(1-\hat{F}(\hat{X}))}\right) = E\left(\frac{I[h(X) \le h(x),\ X \in A^c]}{\hat{R}(h(X))(1-\hat{F}(h(X)))}\right) \\ &= E\left(\frac{I[X \le x,\ X \in A^c]}{R(X)(1-F(X))}\right) = \int_{-\infty}^{x} I_{A^c} \frac{dF(s)}{R(s)\bar{F}(s)} \end{aligned}$$

and

$$\begin{aligned} D &= E\left(\frac{I[\hat{X} \le h(x),\ X \in A]}{\hat{R}(\hat{X})(1-\hat{F}(\hat{X}))}\right) \\ &= \sum_{k:\, x_k < x} E\left(\frac{I[\hat{X} \in \Delta_k]}{\hat{R}(\hat{X})(1-\hat{F}(\hat{X}))}\right) = \sum_{k:\, x_k < x} E \int_{\Delta_k} \frac{d\hat{F}}{\hat{R}(1-\hat{F})} \\ &= \sum_{k:\, x_k < x} \alpha \int_{\Delta_{k,2}} \frac{d(2k^2(s - h(x_k)) - 1)F\{x_k\}}{G(x_k)(1 - F(x_k-) - [2k^2(s - h(x_k)) - 1]F\{x_k\})^2} \\ &= \sum_{k:\, x_k < x} \frac{\alpha}{G(x_k)} \int_0^{F\{x_k\}} \frac{ds}{(1 - F(x_k-) - s)^2} = \sum_{k:\, x_k < x} \int_{\{x_k\}} \frac{dF}{R\bar{F}}. \end{aligned}$$

Therefore,

$$\int_{-\infty}^{h(x)} \frac{d\hat{F}^*}{\hat{R}^2} = B + D = \int_{-\infty}^{x} \frac{dF}{R\bar{F}}. \tag{4.7}$$

Evaluation of the second integral requires the specification of the values at s in the intervals $\Delta_{j,1}$. We proceed as above by computing the integral separately over A and A^c as follows:

$$\int_{h(x)}^{\infty} \frac{d\hat{G}^*}{\hat{R}^2} = \int_{h(x)}^{\infty} \frac{d\hat{G}}{\hat{R}\hat{G}} = E\left(\frac{I[\hat{Y} > h(x)]}{\hat{R}(\hat{Y})\hat{G}(\hat{Y})}\right) \equiv B_1 + D_1, \tag{4.8}$$

where

$$B_1 = E\left(\frac{I[\hat{Y} > h(x),\ Y \in A^c]}{\hat{R}(\hat{Y})\hat{G}(\hat{Y})}\right), \qquad D_1 = E\left(\frac{I[\hat{Y} > h(x),\ Y \in A]}{\hat{R}(\hat{Y})\hat{G}(\hat{Y})}\right).$$

After some tedious computations similar to those of B and D, we arrive at

$$B_1 + D_1 = \int_x^\infty I_{A^c}\frac{dG(s)}{R(s)G(s)} + \int_x^\infty I_A \frac{dG(s)}{R(s)G(s-)} = \int_x^\infty \frac{dG(s)}{R(s)G(s-)}.$$

The equalities (4.5) through (4.9) show that $\sigma^2 = \sigma_1^2$. This completes the proof of Theorem 4.1. $\square$

For the modified estimate

$$\tilde{\alpha}_n = \frac{\tilde{G}_n(x)(1-\tilde{F}_n(x-))}{R_n(x)} \quad \text{for any } x \text{ such that } R_n(x) > 0,$$

we have the following result.

THEOREM 4.2. *For possibly discontinuous F and G, if*

$$\int \frac{dF}{G} < \infty, \qquad \int \frac{dG}{1-F_-} < \infty \tag{4.9}$$

and $a_{G^*} < b_{F^*}$, then as $n \to \infty$, $\sqrt{n}(\tilde{\alpha}_n - \alpha)$ converges weakly to the normal distribution $N(0, \sigma^2)$. Moreover, with probability 1, *the sequence $\{\phi(n)^{-1}(\tilde{\alpha}_n - \alpha);\ n \geq 7\}$ is relatively compact with the set of its limit points $[-\sigma, \sigma]$, where σ^2 is given in Theorem* 4.1.

PROOF. For s belonging to the continuity set A^c, we apply Corollary 5.3 of He and Yang (1998) to obtain $\{Y_i \leq s,\ X_i \geq Y_i\} = \{\hat{Y}_i \leq h(s),\ \hat{X}_i \geq \hat{Y}_i\}$ and $\{Y_i \leq x_j,\ X_i \geq Y_i\} = \{\hat{Y}_i \leq t,\ \hat{X}_i \geq \hat{Y}_i\}$, $\forall\, t \in \Delta_{j,2}$. It follows that $G_n^*(s) = \hat{G}_n^*(h(s))$, $\forall\, s \in A^c$ and $G_n^*(x_j) = \hat{G}_n^*(h(t))$, $\forall\, t \in \Delta_{j,2}$.

Hence by (3.5) for $x \in A^c$, we have

$$\begin{aligned}
1 - \tilde{F}_n(x) &= \prod_{s \leq x}\left(1 - \frac{{}^\#\{i; U_i = s,\ 1 \leq i \leq n\} I[G_n^*(s) \geq n^{\theta-1}]}{{}^\#\{i; V_i \leq s \leq U_i,\ 1 \leq i \leq n\}}\right) \\
&= \prod_{s \leq x,\, s \in A^c}\left(1 - \frac{{}^\#\{i; \hat{X}_i = h(s),\ \hat{X}_i \geq \hat{Y}_i,\ 1 \leq i \leq m_n\} I[\hat{G}_n^*(h(s)) \geq n^{\theta-1}]}{{}^\#\{i; \hat{Y}_i \leq h(s) \leq \hat{X}_i,\ 1 \leq i \leq m_n\}}\right) \\
&\quad \times \prod_{j:\, x_j < x}\left(1 - \frac{{}^\#\{i; \hat{X}_i \in \Delta_{j,2},\ \hat{X}_i \geq \hat{Y}_i,\ 1 \leq i \leq m_n\} I[\hat{G}_n^*(h(x_j+)) \geq n^{\theta-1}]}{{}^\#\{i; \hat{Y}_i \leq h(x_j) + 1/2j^2 \leq \hat{X}_i,\ 1 \leq i \leq m_n\}}\right) \\
&= \prod_{s \leq h(x)}\left(1 - \frac{{}^\#\{i; \hat{U}_i = s,\ 1 \leq i \leq n\} I[\hat{G}_n^*(s) \geq n^{\theta-1}]}{{}^\#\{i; \hat{V}_i \leq s \leq \hat{U}_i,\ 1 \leq i \leq n\}}\right).
\end{aligned}$$

The second equality is obtained by translating U_i, V_i into X_i, Y_i and then into $\hat{X}_i, \hat{Y}_i$. By the same token,

$$\tilde{G}_n(x) = \prod_{s>h(x)} \left(1 - \frac{^{\#}\{i;\ \hat{V}_i = s\} I[1 - \hat{F}_n^*(s-) \geq n^{\theta-1}]}{^{\#}\{i;\ \hat{V}_i \leq s \leq \hat{U}_i\}}\right).$$

Finally, Theorem 3.4 and (4.5) imply Theorem 4.2. □

REMARK. If $a_{G^*} = b_{F^*}$, then the observations $(U_i, V_i) = (a_{G^*}, a_{G^*})$, $i = 1, 2, \ldots, n$. This implies that $\alpha_n = \hat{\alpha}_n = \tilde{\alpha}_n = 1$. Since $a_{G^*} = b_{F^*}$ implies that $F\{a_{G^*}\} = 1$ and $G\{a_{G^*}\} = 1$, hence $\alpha = 1$. Therefore, $\hat{\alpha}_n = \tilde{\alpha}_n = \alpha$.

REFERENCES

CHAO, M.-T. (1987). Influence curves for randomly truncated data. *Biometrika* **74** 426–429.

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 Bell estimator for truncated data. *Ann. Statist.* **23** 440–449.

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.

HE, S. and YANG, G. L. (1998). The strong law under random truncation. *Ann. Statist.* **26** 992–1010.

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

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

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

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