The Annals of Statistics
1988, Vol. 16, No. 2, 661–668

SOME REPRESENTATIONS OF THE NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATORS WITH TRUNCATED DATA

By Min-Te Chao[1] and Shaw-Hwa Lo

Academia Sinica and Harvard University

The nonparametric maximum likelihood estimators of the distribution functions of observations in the truncation model are represented as iid mean processes, with a remainder term of order $o(n^{-1/2})$ a.s.

1. Introduction and main results. Let (X_i, Y_i), $i = 1, \dots, N$, be iid positive random variables with X_i and Y_i independent for each i. In the random truncation model, one observes (X, Y) only if $X > Y$. Based on these observed n pairs of samples $\{(x_i, y_i), i = 1, 2, \dots, n\}$, one attempts to estimate F, G and N, where F and G denote the population distributions of X and Y, respectively. This model has been considered by several authors [see, for example, Woodroofe (1985) and Wang and Jewell (1985)]. The following nonparametric MLE $\hat{S}$ $(= 1 - \hat{F})$, due to Lynden-Bell (1971) and formally studied by Woodroofe (1985), is used to estimate S $(= 1 - F)$ on the basis of the data $\{(x_i, y_i), 1 \le i \le n\}$:

$$\hat{S}(x) = \prod\nolimits^* (1 - r(x_i)/nC_n(x_i)), \quad \text{for } 0 \le x < \infty,$$

where $r(x_i) = \#\{k \le n: x_k = x_i\}$ and $\prod^*$ indicates product over i such that $x_i < x$. The function C_n plays an important role in the truncation model and is defined by $C_n(z) = G_n(z) - F_n(z^-)$, where G_n and F_n are the usual empirical distributions based on $\{y_i\}$ and $\{x_i\}$, respectively.

Woodroofe (1985) proved the uniform weak consistency of $\hat{S}$ as an estimate of S over the half-line $[0, \infty)$. Wang and Jewell (1985) proved the uniform strong consistency of $\hat{S}$ over $[a, \infty)$, where $a > a_F = \inf\{x > 0, F(x) > 0\}$. Woodroofe also showed the weak convergence of the process $n^{1/2}(\hat{F} - F)$ to be a Gaussian process on $D[0, b]$ under the condition $\int_0^\infty dF/G < \infty$. Some general, related applications and motivation of this model can be found in Lynden-Bell (1971), Bhattacharya (1983), Bhattacharya, Chernoff and Yang (1983), Nicoll and Segal (1980), Segal (1975), Woodroofe (1985) and Wang and Jewell (1985). Since the estimator $\hat{S}$ is the analogue of the product–limit estimator of Kaplan and Meier (1958) for censored data, one expects $\hat{S}$ would have similar properties for the truncated case. As Woodroofe (1985) pointed out, there are some differences. One can only estimate F if the pair (F, G) belongs to certain class $\mathscr{H}_0$ to be defined later. In addition, one can always estimate F well at the left end (near 0) in the censored case, but it is not so in the truncated case, especially when we consider the weak convergence of the process $n^{1/2}(\hat{F} - F)$.

Received December 1985; revised June 1987.

[1]Partially supported by the National Science Council of the Republic of China.

AMS 1980 *subject classifications*. 62G05, 62E20.

Key words and phrases. Functional weak convergence, functional LIL, Lynden-Bell estimates, nonparametric MLE, truncated data.

The basic aim of this article is to further study the estimator $\hat{F}$. We first express the closely related hazard process $-\log \hat{S}(z)$ as iid means of random variables with remainders, which are of order $O(n^{-3/4}l_n^{3/4})$ a.s. ($l_n = \log n$ for brevity), uniformly for $z \in [a, b]$ (Theorem 1), where $a > a_F$. We then extend the same representation to $[0, b]$ under the condition $\lim_{x \to 0^+} F(x)/G(x) = 0$ at the expense that the remainder term is only of order $o(n^{-1/2})$. Under the stronger condition $\int_0^\infty dF/G < \infty$ [the same condition given in Woodroofe (1985)], $(\hat{F} - F)$ is represented as iid means of random variables with remainder terms of order $o(n^{-1/2})$ a.s. uniformly on $[0, b]$ (Theorem 2). One can then obtain the results of weak convergence with covariance structures as well as the results on functional LIL. Before explaining the meaning and implication of our findings, we state the results in two theorems. Woodroofe's notation is followed whenever possible.

Define $H_*(x, y) = \alpha^{-1}\int_0^x G(y \wedge z)\, dF(z)$, $F_*(x) = H_*(x, \infty)$, $G_*(y) = H_*(\infty, y)$, $0 \le y \le x < \infty$, where $\alpha = \int G\, dF$ is assumed positive. For any distribution function $K(x)$ on $[0, \infty)$, define $a_K = \inf\{z > 0: K(z) > 0\}$, $b_K = \sup\{z > 0: K(z) < 1\}$. Let $\mathscr{H} = \{(F, G): F(0) = G(0) = 0,\ \alpha(F, G) > 0\}$, $\mathscr{H}_0 = \{(F, G) \in \mathscr{H}: a_G \le a_F,\ b_G \le b_F\}$. Woodroofe (1985) pointed out that one can only estimate the conditional distribution F_0 of X given $X \ge a_G$ and $Y \le b_F$. Therefore we shall assume $(F, G) \in \mathscr{H}_0$ throughout this paper.

For $z \ge 0$, let

$$\xi(x, y, z) = \alpha^{-1}[g(x, y, z) - I(x \le z)/G(x)S(x)],$$

where

$$g(x, y, z) = \int_0^z I(y \le s \le x)\, dF(s)/g(s)S(s)^2.$$

For $z > a > 0$, write $\xi_a(x, y, z) = \xi(x, y, z) - \xi(x, y, a)$. The first theorem deals with the representation of the empirical cumulative hazard function $-\log \hat{S}(z)$.

THEOREM 1. *If F is continuous and $a > a_G$, $b < b_F$, then*

$$\text{(i)} \quad \begin{aligned}[-\log \hat{S}(z) + \log S(z)] - [-\log \hat{S}(a) + \log S(a)] \\ = -n^{-1}\sum_{i=1}^n \xi_a(x_i, y_i, z) + R'_{na}(z),\end{aligned}$$

where $\sup_{a \le z \le b}|R'_{na}(z)| = O(n^{-3/4}l_n^{3/4})$ *a.s. Furthermore, if $a_G < a_F$, then* (i) *reduces to*

$$\text{(ii)} \qquad -\log \hat{S}(z) + \log S(z) = -n^{-1}\sum_{i=1}^n \xi(x_i, y_i, z) + R'_n(z),$$

where $\sup_{a \le z \le b}|R'_n(z)| = O(n^{-3/4}l_n^{3/4})$ *a.s., and one can also write*

$$\text{(iii)} \qquad \hat{F}(z) - F(z) = (S(z)/n)\sum_{i=1}^n \xi(x_i, y_i, z) + R_n(z),$$

where $\sup_{0 \le z \le b}|R_n(z)| = O(n^{-3/4}l_n^{3/4})$ *a.s.*

The next theorem deals with the case $a_F = a_G = 0$.

THEOREM 2. *Assume F is continuous and* $\lim_{x \to 0^+} F(x)/G(x) = 0$. *Then*

$$\text{(i)} \qquad -\log \hat{S}(z) + \log S(z) = -n^{-1} \sum_{i=1}^{n} \xi(x_i, y_i, z) + R'_n(z),$$

where $\sup_{0 \le z \le b} |R'_n(z)| = o(n^{-1/2})$ *a.s. Furthermore, if* $\int_0^\infty dF/G < \infty$, *then*

$$\text{(ii)} \qquad \hat{F}(z) - F(z) = (S(z)/n) \sum_{i=1}^{n} \xi(x_i, y_i, z) + R_n(z),$$

where $\sup_{0 \le z \le b} |R_n(z)| = o(n^{-1/2})$ *a.s.*

Write $\xi_{ai}(z) = \xi_a(x_i, y_i, z)$ and $\xi(z) = \xi(x, y, z)$. It can be shown that

$$\Gamma(z_1, z_2) \equiv \operatorname{Cov}(\xi(z_1), \xi(z_2)) = \alpha \int_0^{z_1 \wedge z_2} (GS^2)^{-1}\, dF.$$

The following corollaries are easily obtained from Theorem 2.

COROLLARY 1. *Under the condition that* $\int_0^\infty dF/G < \infty$, *the process* $n^{1/2}(\hat{F} - F)$ *converges weakly to a zero mean Gaussian process with covariance structure* $S(z_1)S(z_2)\Gamma(z_1, z_2)$ *on* $D[0, b]$.

COROLLARY 2. *Under the condition that* $\int_0^\infty dF/G < \infty$, *the process* $n^{1/2}(\hat{F} - F)$ *obeys the functional* LIL *on* $[0, b]$.

Note that $a_G < a_F$ implies $\int_0^\infty dF/G < \infty$. Therefore the weak convergence of $n^{1/2}(\hat{F} - F)$ [and hence of $n^{1/2}(-\log \hat{S} + \log S)$] holds automatically in the case of Theorem 1.

For further applications, we can use these representations to obtain asymptotic properties of density estimators and hazard rate estimators. To estimate α and N, Woodroofe (1985) came up with estimators $\hat{\alpha} = \int \hat{G}\, d\hat{F}$ and $\hat{N} = n/\hat{\alpha}$. Under further conditions, one can extend the representations given previously to $[0, \infty)$ and obtain asymptotic properties of $\hat{\alpha}$ and $\hat{N}$. These problems will not be studied here, see Chao (1987).

2. Proofs. In this section we assume $(F, G) \in \mathscr{H}_0$, F, G continuous and a, b are two fixed values such that $a_G \le a \le b \le b_F$. The following arguments extend, with some additional effort, to the case when F is continuous (and hence F_*) and G is arbitrary; however, in order to keep the proofs simple, it is assumed here that both F and G (and hence G_* and C) are continuous. We shall omit the dual problem of estimating G. Define $\hat{\Lambda}(z) = \int_0^z dF_n/C_n = \sum_{x_i \le z} [nC_n(x_i)]^{-1}$. The counterpart of C_n is C, defined by $C(z) = G_*(z) - F_*(z^-) = \alpha^{-1} G(z) S(z^-)$. The proofs of both theorems depend on the following basic decomposition and the technical estimates of the error terms. For any $z > a$, with a little algebraic manipulation, write

$$
\begin{aligned}
&[-\log \hat{S}(z) + \log S(z)] - [-\log \hat{S}(a) + \log S(a)] \\
&\quad = -\int_a^z (C_n/C^2)\, dF_* + \int_a^z (1/C)\, dF_n \\
(1)\qquad &\quad\quad + \int_a^z (1/C_n - 1/C)\, d(F_n - F_*) + \int_a^z \left[(C - C_n)^2/C_n C^2\right] dF_* \\
&\quad\quad + \{[-\log \hat{S}(z) - \hat{\Lambda}_n(z)] - [-\log \hat{S}(a) - \hat{\Lambda}_n(a)]\} \\
&\quad \equiv -n^{-1} \sum_{i=1}^{n} \xi_a(x_i, y_i, z) + \mathrm{III}_n(a, z) + \mathrm{IV}_n(a, z) + \mathrm{V}_n(a, z), \quad \text{say.}
\end{aligned}
$$

The following proposition provides estimates of the remainders.

PROPOSITION 1. *For a and b given previously, one can write*

$$
\text{(i)} \qquad \sup_{a \le z \le b} |\mathrm{III}_n(a, z)| = O\left(n^{-3/4} l_n^{3/4}\right) \quad a.s.,
$$

$$
\text{(ii)} \qquad \sup_{a \le z \le b} |\mathrm{IV}_n(a, z)| = O\left(n^{-1} l_n\right) \quad a.s.,
$$

$$
\text{(iii)} \qquad \sup_{a \le z \le b} |\mathrm{V}_n(a, z)| = O\left(n^{-1} l_n\right) \quad a.s.
$$

The proof of this proposition is given in the Appendix.

PROOF OF THEOREM 1. Parts (i) and (ii) of the theorem follow immediately from Proposition 1 and (1). Part (iii) of the theorem follows from (ii) and a two-term Taylor expansion of $\log \hat{S} - \log S$. □

PROOF OF THEOREM 2. To prove Theorem 2, we need to take a proper $a = \alpha_n \to 0$ in (1). For any fixed $\delta > 0$, let $\alpha_n = F_*^{-1}(n^{-3/2} l_n^{-\delta})$. It can be shown that $-\log \hat{S}(\alpha_n) - \Lambda(\alpha_n) = o(n^{-1/2})$ a.s. Write $C_n^*(t) = \max\{C_n(t), n^{-1}\}$. Since $C_n(x_i) \ge n^{-1}$ we see that in (1) we can replace C_n by C_n^* at will if integration is with respect to F_n. On the other hand, $C_n(t) = 0$ for $t < y_{(1)}$ = the smallest-order statistics of the y's. Using these observations, it can be shown that

$$
\limsup |\mathrm{III}_n(0, \alpha_n) + \mathrm{IV}_n(0, \alpha_n) + \mathrm{V}_n(0, \alpha_n)| = o(n^{-1/2}) \quad \text{a.s.}
$$

Letting $z = \alpha_n$ and $a = 0$ in (1), we have

$$
n^{-1} \sum_{i=1}^{n} \xi(x_i, y_i, \alpha_n) = o(n^{-1/2}) \quad \text{a.s.}
$$

and (1) reduces to

$$
\begin{aligned}
(2)\qquad - \log \hat{S}(z) + \log S(z) = &- n^{-1} \sum_{i=1}^{n} \xi(x_i, y_i, z) + \mathrm{III}_n(\alpha_n, z) \\
&+ \mathrm{IV}_n(\alpha_n, z) + \mathrm{V}_n(\alpha_n, z) + o(n^{-1/2}) \quad \text{a.s.}
\end{aligned}
$$

Part (i) follows if the orders of II, IV and V can be controlled properly. Part (ii) follows by a two-term Taylor expansion. It remains to estimate the orders of III, IV and V. We summarize these estimates in

PROPOSITION 2. *Assuming* $\lim_{x \to 0^+} F(x)/G(x) = 0$, *one can write*

(i) $$\sup_{\alpha_n \le z \le b} |\mathrm{III}_n(\alpha_n, z)| = o(n^{-1/2}) \quad a.s.,$$

(ii) $$\sup_{\alpha_n \le z \le b} |\mathrm{IV}_n(\alpha_n, z)| = o(n^{-1/2}) \quad a.s.,$$

(iii) $$\sup_{\alpha_n \le z \le b} |\mathrm{V}_n(\alpha_n, z)| = o(n^{-1/2}) \quad a.s.$$

The proof of this proposition is also given in the Appendix.

PROOF OF COROLLARY 1. It suffices to show that the process $n^{1/2}(\hat{F}_n(t) - F(t))$ is tight in $D[0, b]$. From Theorem 2 and some easy calculations, it can be shown that for all $s, t \in [0, b]$,

$$\mathrm{Var}(S(s)\xi(s) - S(t)\xi(t)) \le |W(s) - W(t)|,$$

for a continuous nondecreasing function W, e.g., $W(t) = \text{const.}(\int_0^t dF/GS^2 + F(t))$. It follows that for any $0 \le t_1 \le t \le t_2 \le b$,

$$E\left\{n\left(S(t_1)\bar{\xi}(t_1) - S(t)\bar{\xi}(t)\right)^2 n\left(S(t)\bar{\xi}(t) - S(t_2)\bar{\xi}(t_2)\right)^2\right\} \le 3\{W(t_2) - W(t_1)\}^2, \tag{3}$$

where $\bar{\xi}(t) = n^{-1}\sum_{i=1}^n \xi(x_i, y_i, t)$. The tightness of the process follows from page 128 of Billingsley (1968). One can also deduce the same result by first proving that the hazard process $n^{1/2}(-\log \hat{S}(z) + \log S(z))$ converges weakly (using Proposition 2 and the similar argument given previously) to a zero mean Gaussian process with covariance structure Γ on $D[0, b]$ under the condition $\int dF/G < \infty$. The desired result can be obtained via Taylor expansion and the continuous mapping principle. □

PROOF OF COROLLARY 2. To see this, one can either use a theorem of Dudley and Philipp (1983) or Proposition 2.1 of Philipp (1977). In the latter approach, one needs finite-dimensional LIL's for $\{n^{1/2}\bar{\xi}(t_i)/ll_n^{1/2}\}_{i=1}^k$ and this is not hard to establish in the present case. In addition, one needs a bound for the fluctuations of the process $n^{1/2}\bar{\xi}(t)/ll_n^{1/2}$ and this is provided by (3). □

While this manuscript was under preparation, we came to know about the preprint by Wang, Jewell and Tsai (1986), a work giving the covariance structure of the process $n^{1/2}\bar{\xi}_a(t)$, where $\bar{\xi}_a(t) = n^{-1}\sum_{i=1}^n \xi_a(x_i, y_i, t)$; however, we think the overlap, if any, is not substantial.

APPENDIX

LEMMA A1 [Lo and Singh (1986)]. *Let $\eta, \eta_1, \eta_2, \dots, \eta_n$ be iid random variables with mean* $0, |\eta| \le c$ *and* $\operatorname{Var}(\eta) = \sigma^2$. *Then for any positive z and d satisfying $cz \le d$ and $nz\sigma^2 \le d^2$, $P(|\sum_{i=1}^n \eta_i| \ge 3d) \le 2e^{-z}$.*

PROOF. Let $\phi(t)$ be the moment generating function of η. Since $e^x \le 1 + x + 2x^2$ for $|x| \le 1$, it follows that $\phi(t) \le 1 + 2t^2\sigma^2$ if $|t\eta| < 1$. We have

$$\begin{aligned} P\left(\sum_{i=1}^n \eta_i \ge 3d\right) &\le e^{-3t}[\phi(t)]^n \\ &\le e^{-3dt+2nt^2\sigma^2} \\ &\le e^{-z}, \end{aligned}$$

since $z = dt$. Similarly, we can show that $P(\sum_{i=1}^n \eta_i \le -3d) \le e^{-z}$, and the lemma follows easily by combining these two inequalities. □

LEMMA A2. *If* $\lim_{x \to 0^+} F(x)/G(x) = 0$, *then*

$$\sup_{\alpha_n \le z \le b} |[C_n(x) - C(x)]/C^{1/2}(x)| = O\left(n^{-1/2} l_n^{1/2}\right) \quad a.s.,$$

where $\alpha_n = F_*^{-1}(n^{-3/2} l_n^{-\delta})$ *for some* $\delta > 0$.

PROOF. Since $\lim_{x \to 0^+} F(x)/G(x) = 0$, it is easy to see that $C(x)$ for $x \ge \alpha_n$ is bounded below by $O(n^{-3/4} l_n^{-\delta/2})$ a.s. Since C_n is a difference of two empirical distributions, we can use Theorem 5.1.5 of Csörgő and Révész (1981) or Theorem 2 of Földes (1981) to obtain the desired bound. □

LEMMA A3. *Assuming* $\lim_{x \to 0^+} F(x)/G(x) = 0$ *and* $0 < \delta < 1/2$, *then*

$$\sup_{a_n \le z \le b} |\mathrm{III}_n(a_n, x)| = O\left(n^{-3/4+2\delta} l_n^{3/4}\right),$$

where $F_*(a_n) = n^{-\delta}$.

PROOF. A readily adaptable proof is available in the literature, with only minor changes to accommodate our choice of a_n. See Lemma 2 of Lo and Singh (1986). □

LEMMA A4. *Assume* $\lim_{x \to 0^+} F(x)/G(x) = 0$ *and* $\delta > 2$ *and let* $\gamma_n = n^{-3/2} l_n^{-\delta}$, $F_*(a_n) = \gamma_n n^{\varepsilon_1}$, $F_*(b_n) = \gamma_n n^{\varepsilon_2}$. *If* $0 \le \varepsilon_1 < \varepsilon_2 < 1$ *and* $\varepsilon_2 - 3\varepsilon_1/4 = 3/8$, *then*

$$\sup_{a_n \le x \le b_n} |\mathrm{III}_n(a_n, x)| = o(n^{-1/2}) \quad a.s.$$

PROOF. Write $\zeta(z) = (C_n - C)(C_n C)^{-1}(z)$. Since

$$|\mathrm{III}_n(a_n, x)| \le \int_{a_n}^x |\zeta(z)|\, dF_n(z) + \int_{a_n}^x |\zeta(z)|\, dF_*(z),$$

and since the two integrals can be estimated with identical methods, we only evaluate the first term. By Lemma A2,

$$\sup_{a_n \le x \le b_n} |[C_n(x) - C(x)]/C^{1/2}(x)| = O\left(n^{-1/2} l_n^{1/2}\right) \quad \text{a.s.}$$

and $F_n(b_n) = F_*(b_n)(1 + o(1))$ a.s. Hence

$$\int_{a_n}^{x} |\zeta(z)|\, dF_n = O\left(n^{-1/2} l_n^{1/2} G^{-3/2}(a_n)\gamma_n n^{\varepsilon_2}\right) \quad \text{a.s.},$$

uniformly on $x \in [a_n, b_n]$. It is easy to show $G^{-3/2}(a_n) = O(\gamma_n^{-3/4} n^{-3\varepsilon_1/4})$ and the final term found is $O(n^{-1/2} l_n^{1/2-\delta/4}) = o(n^{-1/2})$. The proof of this lemma is complete. □

LEMMA A5. *Under the conditions of Lemma* A4 *but with* $\varepsilon_2 - \varepsilon_1 < 1/2$, *then*

$$\sup_{a_n \le x \le b_n} |\mathrm{IV}_n(a_n, x)| = o(n^{-1/2}) \quad a.s.$$

The proof of this lemma is similar to that of Lemma A4, only simpler. We omit it.

PROOF OF PROPOSITION 1. Part (i) is a special case of Lemma A3 with $\delta = 0$. Parts (ii) and (iii) follow from a similar, easier argument. See Proposition 2(ii) and (iii). □

PROOF OF PROPOSITION 2. (i) Define $\varepsilon_1 = 0$, $\varepsilon_k = 3\varepsilon_{k-1}/4 + 3/8$. It can be shown that the interval $[\alpha_n, \beta_n]$, where $\beta_n = F_*^{-1}(n^{-\delta})$, is covered by a finite number of intervals of the form specified in Lemma A4. Applying Lemma A4 repeatedly to these intervals, we have

$$\sup_{\alpha_n \le x \le \beta_n} |\mathrm{III}_n(\alpha_n, x)| = o(n^{-1/2}) \quad \text{a.s.}$$

The upper bound for $\mathrm{III}_n(\beta_n, x)$ for x in $[\beta_n, b]$ can be obtained by setting $\delta = 1/9$ in Lemma A3.

(ii) Define $\varepsilon_j = (j-1)/4$. For $i = 1$–4, applying Lemma A5 with ε_1 and ε_2 replaced by ε_i and ε_{i+1}, we obtain

$$\sup_{\alpha_n \le x \le \beta_n} |\mathrm{IV}_n(\alpha_n, x)| = o(n^{-1/2}) \quad \text{a.s.},$$

where $\beta_n = F_*^{-1}(n^{-1/2} l_n^{-\delta})$. The bound for $\mathrm{IV}_n(\beta_n, x)$ over $[\beta_n, b]$ can be easily shown to be $o(n^{-1/2})$. Part (ii) follows.

(iii) Define $Q(a, x) = \int_a^x [nC_n^2(t)]^{-1}\, dF_n(t)$ and note that for $a \ge \alpha_n$, $Q(a, x)$ has essentially the same form as $\mathrm{IV}_n(a, x)$. Hence the estimates of its orders can be carried out in exactly the same way as in part (ii). Part (iii) follows because $V_n(\alpha_n, x)$ is bounded in order by $Q(\alpha_n, x)$ for all x. □

Acknowledgement. The authors would like to thank Dr. Philip E. Cheng for helpful discussions.

REFERENCES

BHATTACHARYA, P. K. (1983). Justification for a K-S type test for the slope of a 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.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

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

CSÖRGŐ, M. and RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics.* Academic, New York.

DUDLEY, R. M. and PHILIPP, W. (1983). Invariance principles for sums of Banach space valued random elements and empirical processes. *Z. Wahrsch. verw. Gebiete* **62** 509–522.

FÖLDES, A. (1981). Strong uniform consistency of the product limit estimator under variable censoring. *Z. Wahrsch. verw. Gebiete* **58** 95–108.

GALAMBOS, J. (1978). *The Asymptotic Theory of Extreme Order Statistics.* Wiley, New York.

KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

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

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

NICOLL, J. F. and SEGAL, I. E. (1980). Nonparametric elimination of the observational cutoff bias. *Astron. and Astrophys.* **82** L3–L6.

PHILIPP, W. (1977). A functional law of the iterated logarithm for empirical distribution functions of weakly dependent random variables. *Ann. Probab.* **5** 319–350.

SEGAL, I. E. (1975). Observational validation of the chronometric cosmology: I. Preliminaries and the red shift–magnitude relation. *Proc. Nat. Acad. Sci. U.S.A.* **72** 2437–2477.

WANG, M.-C. and JEWELL, N. P. (1985). The product limit estimate of a distribution function under random truncation. Technical Report, Program in Biostatistics, Univ. California, Berkeley.

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.

INSTITUTE OF STATISTICS
ACADEMIA SINICA
TAIPEI 11529
REPUBLIC OF CHINA

DEPARTMENT OF STATISTICS
HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS 02138