Ann. Inst. Statist. Math.
Vol. 57, No. 1, 61–69 (2005)

QUANTILE PROCESS FOR LEFT TRUNCATED AND RIGHT CENSORED DATA*

SZEMAN TSE

Department of Applied Mathematics, National Donghua University, Hualien, Taiwan, R.O.C.,
e-mail: tse@server.am.ndhu.edu.tw

(Received January 15, 2003; revised November 19, 2003)

Abstract. In this paper, we consider the product-limit quantile estimator of an unknown quantile function when the data are subject to random left truncation and right censorship. This is a parallel problem to the estimation of the unknown distribution function by the product-limit estimator under the same model. Simultaneous strong Gaussian approximations of the product-limit process and product-limit quantile process are constructed with rate $O(\frac{(\log n)^{3/2}}{n^{1/8}})$. A functional law of the iterated logarithm for the maximal deviation of the estimator from the estimand is derived from the construction.

Key words and phrases: Left truncation, right censorship, product-limit, quantile process, Gaussian approximations.

1. Introduction

Let X, T and S be independent, positive random variables with continuous distribution functions (df) F^0, G^0 and L^0 respectively. Moreover, F^0 is differentiable with derivative f^0. Let $Y = \min(X, S)$ and $\delta = I(X \leq S)$. If $Y \geq T$, one observes (Y, T, δ). If $Y < T$, nothing is observed. We think of X as the variable of interest, the observation of which is subjected to right censorship, S, and left truncation, T, mechanisms. δ indicates whether the observed Y is a censored item or not. This is the left truncation, right censorship (LTRC) model. Denote the df of Y by J. By the independent assumption, we have $1 - J = (1 - F^0)(1 - L^0)$. Let (X_i, T_i, S_i), $i = 1, \ldots, N$ be i.i.d. as (X, T, S), where the population size N is fixed, but unknown. The empirical data are (Y_i, T_i, δ_i), $i = 1, \ldots, n$ where n is the number of observed triplets.

The nonparametric maximum likelihood estimator of F^0 (Wang *et al.* (1986)) is

$$1 - F_n^0(t) = \prod_{i:Y_i \leq t} \left[1 - \frac{1}{nC_n(Y_i)}\right]^{\delta_i}, \tag{1.1}$$

assuming no ties in the data where $nC_n(z) = \sum_{i=1}^n I\{T_i \leq z \leq Y_i\}$. F_n^0 reduces to the Kaplan-Meier product-limit (PL) estimator when $T = 0$ and to the Lynden-Bell (1971) estimator when there is no right censoring. We shall refer to F_n^0 as the product-limit (PL) estimator for the LTRC model.

Gu and Lai (1990), Lai and Ying (1991) obtained a functional law of the iterated logarithm for a slightly modified form of the PL-estimator using martingale theory.

*Work partially supported by NSC Grant 89-2118-M-259-011.

Gijbels and Wang (1993), Zhou (1996) established almost sure representation of PL-estiamtor in terms of sums of normed i.i.d. random processes. Zhou and Yip (1999) initiated and Tse (2003) established strong Gaussian approximation of the PL-process $Z_n(t) = \sqrt{n}[F_n^0(t) - F^0(t)]$ by a two-parameter Gaussian process at the almost sure rate of $O((\log n)^{3/2}/n^{1/8})$, a rate that reflects the two-dimensional nature of the LTRC model. Zhou (2001) examined the asymptotic properties of a smooth quantile estimator.

The quantile function Q and its empirical counterpart Q_n are defined as

$$(1.2)\qquad Q(p) := \inf\{t : F^0(t) \geq p\}, \qquad Q_n(p) := \inf\{t : F_n^0(t) \geq p\}, \qquad 0 < p < 1.$$

The normed PL-quantile process is defined as $\rho_n(p) := \sqrt{n} f^0(Q(p))[Q(p) - Q_n(p)]$. This is the analog of the normed quantile process for i.i.d. data as defined in Csörgő and Révész (1981). The role of the quantile function in statistical data modeling was emphasized by Parzen (1979). In econometrics, Gastwirth (1971) used the quantile function to give a succint definition of the Lorenz curve, which measures inequality in distribution of resources and in size distribution.

In this paper, we build on the result of Tse (2003) to construct strong approximations of the PL-process and the PL-quantile process by the same two-parameter Kiefer type process at the rate of $O((\log n)^{3/2}/n^{1/8})$. The basis of our work is Borisov's (1982) extension of Komlós, Major and Tusnády's (KMT) (1975) theorem for the univariate empirical process to higher dimensions. In our case, the appropriate dimension is two. The approximation rate is not as fast as that in the KMT's theorem, but is still good enough to let us deduce almost sure statements like the law of the iterated logarithm from that of the corresponding Gaussian processes.

In the absence of censorship, we get the corresponding statements for the random truncation model at the same rate, which complement the strong approximation results of the PL-process in Tse (2000). On the other hand, when there is no truncation mechanism, the LTRC model reduces to the one-dimensional random censorship model for which the optimal rate of $O(\log^2 n/\sqrt{n})$ has been obtained by Burke, S. Csörgő and Horváth (BCH) (1981, 1988) for the PL-process and by Aly *et al.* (1985) for the PL-quantile process. It follows that any model that includes the general truncation mechanism is at least of Dimension 2, in contrast to the one-dimensional nature pertaining to the censoring mechanism.

In Section 2, we introduce the notation and present the main results. Auxiliary results and proofs are relegated to Section 3. Examples of applications can be found in Csörgő and Révész (1981). In particular, our results can form the basis for an asymptotic theory of the empirical Lorenz curve under the same model.

2. Notation and main theorems

As a consequence of truncation, the number of observed pairs, n, is a $\mathrm{Bin}(N, \alpha)$ random variable, with $\alpha := P(T \leq Y)$. By the strong law of large numbers, $n/N \to \alpha$ almost surely as $N \to \infty$. Conditional on the value of n, (Y_i, T_i, δ_i), $i = 1, \ldots, n$ are still i.i.d. but with the joint conditional distribution of (Y, T) becomes

$$\begin{aligned} H(y,t) &= P\{Y \leq y, T \leq t \mid T \leq Y\} \\ &= \alpha^{-1} \int_0^y G^0(t \wedge z) dJ(z) \end{aligned}$$

for $y, t > 0$. The marginal distribution functions are denoted by

$$F(y) := H(y, \infty) = \alpha^{-1} \int_0^y G^0(z) dJ(z),$$
$$G(t) := H(\infty, t) = \alpha^{-1} \int_0^\infty G^0(t \wedge z) dJ(z).$$

Here and in the following, $\int_a^b = \int_{(a,b]}$ for $0 \le a < b \le \infty$. Empirical counterparts of these distribution functions are denoted by $H_n(y,t)$, $F_n(y)$ and $G_n(t)$ respectively. For $0 \le z < \infty$, let

$$C(z) = G(z) - F(z-) = \frac{1}{\alpha} P(T \le z \le S)[1 - F^0(z-)].$$

$C(z)$ is consistently estimated by $C_n(z) = G_n(z) - F_n(z-)$. Note that F_n and G_n are the empirical distribution functions for the observed Y's and T's respectively. To take into account the information from the indicator variable. Let $F_{n1}(z) = \frac{1}{n}\sum_{i=1}^n I(Y_i \le z, \delta_i = 1)$ which is a consistent estimator of $F_1(z) = P(Y \le z, \delta = 1 \mid T \le Y)$. Note that F_1 is a sub-distribution of F. The triplets F_n, G_n and F_{n1} contain all the relevant information from the data for the estimation of F^0 and Q. The corresponding empirical processes are defined as:

$$\alpha_{n1}(z_1) := \sqrt{n}[F_n(z_1) - F(z_1)], \qquad \alpha_{n2}(z_2) := \sqrt{n}[G_n(z_2) - G(z_2)],$$
$$\beta_{n1}(u) := \sqrt{n}[F_{n1}(u) - F_1(u)].$$

The notation is intended to remind us that β_{n1} is a sub-process of α_{n1}, whereas α_{n1} and α_{n2} together form a two components random process with covariance depending on F^0 and G^0.

For any df K, let $a_K = \inf\{z : K(z) > 0\}$ and $b_K = \sup\{z : K(z) < 1\}$ denote the left and right end points of its support. As in the random truncation model (Woodroofe (1985)), F^0 can be reconstructed only when $a_{G^0} \le a_J$ and $b_{G^0} \le b_J$. The borderline case is $a_{G^0} = a_J$. For the sake of simplicity, we assume that $a_{G^0} = a_J = 0$ and $b_{G^0} \le b_J$ throughout. The cumulative hazard function associated with F^0 is

$$\Lambda^0(t) := \int_0^t \frac{dF^0(z)}{1 - F^0(z-)}, \qquad 0 \le t < \infty,$$
$$= \int_0^t \frac{dF_1(z)}{C(z)},$$

which is consistently estimated by

$$\Lambda_n^0(t) := \int_0^t \frac{dF_{1n}(z)}{C_n(z)}, \qquad 0 \le t < \infty.$$

The corresponding cumulative hazard process is $\hat{Z}_n(t) := \sqrt{n}[\Lambda_n^0(t) - \Lambda^0(t)]$. For the theorems below, we assume that F^0, G^0 and L^0 satisfy the condition

$$\int_0^\infty \frac{dF_1(z)}{C^3(z)} < \infty. \tag{2.1}$$

The condition, while not optimal, serves to keep the variances of the limiting processes finite near the lower end point and simplify the proof a great deal.

For $0 < t < b < b_J$, let

$$l(t) := \int_0^t \frac{dF_1(u)}{C^2(u)}. \tag{2.2}$$

THEOREM 2.1. *Suppose condition* (2.1) *is satisfied and* $0 < p_0 \le p_1$. *Assume that* F^0 *is Lipschitz continuous and that* F^0 *is twice continuously differentiable on* $[Q(p_0) - \delta, Q(p_1) + \delta]$ *for some* $\delta > 0$ *such that* f^0 *is bounded away from zero there. On a rich enough probability space, one can define a sequence of independent and identically distributed mean zero Gaussian processes* $\{B_n(t), 0 < t < b\}$, *for* $b < b_J$, *with* $\mathrm{Cov}[B_n(s), B_n(t)] = l(\min(s,t))$, *for* $0 < s, t < b < b_J$ *such that, almost surely,*

$$\sup_{0 \le t \le b} |Z_n(t) - [1 - F^0(t)]B_n(t)| = O\left(\frac{\log n}{n^{1/6}}\right),$$

$$\sup_{p_0 \le p \le p_1} |\rho_n(p) - (1-p)B_n(Q(p))| = O\left(\frac{\log n}{n^{1/6}}\right),$$

where $0 < Q(p_0) < Q(p_1) < b$.

The statements above are conditional on n, the observed sample size. The results may be formulated in terms of the non-random population size N with considerable complicated notation as in Tse (2000). To keep things simple, we choose to present the results in terms of n.

Theorem 2.1 approximates Z_n and ρ_n by sequences of copies of their Gaussian limits. Weak convergence results follow immediately. However, almost sure statements cannot be obtained from them since the covariances between different members in the sequences are not specifed. In the next theorem, these sequences are replaced by single two-parameter Gaussian processes.

THEOREM 2.2. *Suppose condition* (2.1) *is satisfied and* $0 < p_0 \le p_1$. *Assume that* F^0 *is Lipschitz continuous and that* F^0 *is twice continuously differentiable on* $[Q(p_0) - \delta, Q(p_1) + \delta]$ *for some* $\delta > 0$ *such that* f^0 *is bounded away from zero there. On a rich enough probability space, one can construct a two-parameter mean zero Gaussian process* $B(t,u)$ *for* $t \ge 0$ *and* $u \ge 0$ *with* $\mathrm{Cov}[B(s,n), B(t,m)] = \sqrt{n/m}\,l(s)$, *for* $n \le m$, $s < t$ *such that, almost surely,*

$$\sup_{0 \le t \le b} |Z_n(t) - [1 - F^0(t)]B(t,n)| = O\left(\frac{(\log n)^{3/2}}{n^{1/8}}\right),$$

$$\sup_{p_0 \le p \le p_1} |\rho_n(p) - (1-p)B(Q(p),n)| = O\left(\frac{(\log n)^{3/2}}{n^{1/8}}\right),$$

where $0 < Q(p_0) < Q(p_1) < b$.

As a consequence of the second statement of Theorem 2.2 for ρ_n, we obtain the next theorem for the uniform consistency rate of Q_n for the LTRC model. The counterpart for the PL-estimator F_n^0, which follows from the first statement of Theorem 2.2, has been established in Tse (2003).

THEOREM 2.3. *Under the same assumption as in Theorem 2.2, the sequence*

$$\left\{ \left(\frac{1}{2 \log \log n} \right)^{1/2} \rho_n(p) \right\}$$

is almost surely relatively compact in the supremum norm of functions over $[p_0, p_1]$, *and its set of limit points is* $\{l(Q(p_1))^{1/2}(1-p)g(\frac{l(Q(p))}{l(Q(p_1))}) : g \in S\}$ *where* S *is Strassen's set of absolutely continuous functions*

$$S = \left\{ g \mid g : [0,1] \to \boldsymbol{R}, g(0) = 0, \int_0^1 \left(\frac{dg(x)}{dx} \right)^2 dx \leq 1 \right\}.$$

Consequently, with $v^2(p) = (1-p)^2 l(Q(p))$,

$$\limsup_{n \to \infty} \left(\frac{n}{2 \log \log n} \right)^{1/2} \sup_{p_0 \leq p \leq p_1} |f^0(Q(p))[Q(p) - Q_n(p)]| = \sup_{p_0 \leq p \leq p_1} v(p),$$

$$\liminf_{n \to \infty} (n \log \log n)^{1/2} \sup_{p_0 \leq p \leq p_1} \frac{f^0(Q(p)) |F_n^0(Q(p)) - p|}{1-p} = \frac{\pi}{8^{1/2}} (l(Q(p_1)))^{1/2},$$

implying,

$$\begin{aligned} \frac{\pi}{8^{1/2}} v(Q(p_1)) &\leq \liminf_{n \to \infty} (n \log \log n)^{1/2} \sup_{p_0 \leq p \leq p_1} f^0(Q(p)) |F_n^0(Q(p)) - p| \\ &\leq \frac{\pi}{8^{1/2}} (l(b))^{1/2}. \end{aligned}$$

Note that in the absence of censorship, Theorems 2.1, 2.2 and 2.3 give corresponding results for the random truncation model. In the absence of truncation, the model reduced to the one-dimensional censorship model for which BCH's (1981, 1988) result is optimal.

3. Auxiliary results and proofs

To achieve our goal of simultaneous Gaussian approximations of Z_n over $[0, b]$ and Q_n over $[p_0, p_1]$ where $0 < Q(p_0) < Q(p_1) < b$, we shall obtained a representation of the PL-quantile process in terms of the PL process. This is the analogue of Bahadur's (1966) result for i.i.d. data. Similar representations for randomly censored data have been obtained by Cheng (1984), Aly *et al.* (1985), Lo and Singh (1986), and for randomly truncated data by Guler *et al.* (1993) and Zhou (2001). Then we apply the strong Gaussian approximation results for the PL process Tse (2003) to obtain Theorems 2.1 and 2.2. Finally, Theorem 2.3 follow from the its analogue for the PL process.

To prepare for the Proof of Theorem 2.1, we start with a few lemmas. Lemma 3.1 shows the strong consistency of $Q_n(p)$ as an estimator of $Q(p)$. Lemma 3.2 shows that F_n^0 composed with Q_n is an approximate identity up to order $O(1/n)$. Lemmas 3.3 and 3.4 give global and local bounds for the deviation between Q_n and Q.

LEMMA 3.1. *Suppose F^0 is continous, $a_{G^0} \le a_J$, $b_{G^0} \le b_J$ and condition (2.1) is satisfied. If $Q(p)$ is the unique solution of $F^0(y) = p$, then $Q_n(p)$ converges to $Q(p)$ with probability one.*

The proof is parallel to that for i.i.d. data, we omit it here. (See Serfling (1980).)

LEMMA 3.2. *If condition (2.1) is satisfied, then $\sup_{p_0 \le p \le p_1} |F_n^0(Q_n(p)) - p| = O(\frac{1}{n})$ with probability one.*

PROOF. Noting that $Q_n(p) = Y_i$ for some observed item Y_i, we have

$$|F_n^0(Q_n(p)) - p| = F_n^0(Q_n(p)) - p \le F_n^0(Y_i) - F_n^0(Y_i-).$$

Taking supremum over $[p_0, p_1]$ and using the definition of the PL estimator (1.1), we have, almost surely,

$$\begin{aligned}
\sup_{p_0 \le p \le p_1} |F_n^0(Q_n(p)) - p| &\le \sup_{Q_n(p_0) \le Y_i \le Q_n(p_1)} |F_n^0(Y_i) - F_n^0(Y_i-)| \\
&\le \sup_{Q_n(p_0) \le Y_i \le Q_n(p_1)} [1 - F_n^0(Y_i-)] \left[\frac{1}{nC_n(Y_i)}\right]^{\delta_i} \\
&\le \sup_{Q_n(p_0) \le Y_i \le Q_n(p_1)} \left|\frac{1}{nC_n(Y_i)}\right|
\end{aligned}$$

where $nC_n(Y_i) \ge 1$ for $1 \le i \le n$. For large enough n and small enough ϵ, the last expression is bounded by $\sup_{Q(p_0)-\epsilon \le Y_i \le Q(p_1)+\epsilon} |\frac{1}{nC_n(Y_i)}|$ by Lemma 3.1. Finally, since C_n converges uniformly to C and $\inf_{Q(p_0)-\epsilon \le y \le Q(p_1)+\epsilon} C(y) > 0$, we obtained the desired result. □

LEMMA 3.3. *Suppose condition (2.1) is satisfied and F^0 is continuously differentiable with derivative f^0 bounded away from zero on $[Q(p_0) - \delta, Q(p_1) + \delta]$ for some $\delta > 0$. Then we have, almost surely, $\sup_{p_0 \le p \le p_1} \sqrt{n}|Q_n(p) - Q(p)| = O(\sqrt{\log\log n})$.*

PROOF. Using the strong approximation results for the PL-process in Theorems 2.1 and 2.2 in Tse (2003), and the law of the iterated logarithm (LIL) for the Wiener process, we have, almost surely, $\sup_{0 \le t \le b} \sqrt{n}|F_n^0(t) - F^0(t)| = O(\sqrt{\log\log n})$. With $t = Q_n(p)$, we have, almost surely,

$$\sqrt{n}F_n^0(Q_n(p)) = \sqrt{n}F^0(Q_n(p)) + O(\sqrt{\log\log n}).$$

By Lemma 3.1, a Taylor expansion about $Q(p)$ gives, almost surely,

$$\sqrt{n}F_n^0(Q_n(p)) = \sqrt{n}F^0(Q(p)) + \sqrt{n}f^0(\zeta_n(p))[Q_n(p) - Q(p)] + O(\sqrt{\log\log n})$$

for some $\zeta_n(p)$ between $Q(p)$ and $Q_n(p)$. Rearranging and using Lemma 3.2, we have, almost surely,

$$\begin{aligned}
\sqrt{n}f^0(\zeta_n(p))[Q_n(p) - Q(p)] &= \sqrt{n}[F_n^0(Q_n(p)) - F^0(Q(p))] + O(\sqrt{\log\log n}) \\
&= \sqrt{n}[p - F^0(Q(p))] + O(\sqrt{\log\log n}) \\
&= O(\sqrt{\log\log n})
\end{aligned}$$

since F^0 is continuous. Recalling the assumption for f^0, we can divide both sides by $f^0(\zeta_n)$. Taking supremum over $[p_0, p_1]$ gives the desired result. □

LEMMA 3.4. *Let* $\epsilon_n = k_1(\log n/n)^{1/2}$ *and* $\delta_n = k_2(\log n/n)^{3/4}$, *where* k_1, k_2 *are constants. Suppose* F^0 *is Lipschitz continuous on* $[0, b]$, $b < b_J$, *and condition* (2.1) *is satisfied. Then, for any* $0 \le s, t \le b$, *we have, almost surely,*

$$\sup_{|s-t|\le\epsilon_n} \left| \frac{1}{\sqrt{n}}[Z_n(s) - Z_n(t)] \right| = O(\delta_n). \tag{3.1}$$

PROOF. We start with the cumulative hazard process $\hat{Z}_n$, which has the usual decomposition

$$\hat{Z}_n(t) = \int_0^t \frac{d\beta_{n1}}{C} + \int_0^t \frac{\sqrt{n}(C - C_n)}{C^2} dF_1 + R_{n1}(t), \qquad \text{for} \quad 0 \le t \le b,$$

where $\sup_{0\le t\le b} |R_{n1}(t)| = O(\frac{\log\log n}{\sqrt{n}})$ almost surely by Theorem 2.1 of Zhou and Yip (1999). For $0 \le s, t \le b$, we have, almost surely,

$$\begin{aligned} \frac{1}{\sqrt{n}}|\hat{Z}_n(s) - \hat{Z}_n(t)| = {} & \int_s^t \frac{d(F_{n1} - F_1)}{C} + \int_s^t \frac{\sqrt{n}(C - C_n)}{C^2} dF_1 \\ & + O\left(\frac{\log\log n}{n}\right). \end{aligned} \tag{3.2}$$

By the Lipschitz continuity of F^0 and the LIL for empirical process, the second term in the right hand side is $O(\log n/n)$ almost surely.

For the first term, introduce the grid

$$y_{nj} = j\delta_n, \qquad j = 0, 1, \ldots, l_n = \left[\frac{b}{\delta_n}\right], \qquad y_{nl_n+1} = b.$$

For each $y \in [0, b]$, choose y_{nj} such that $y_{nj} \le y \le y_{nj+1}$. In particular, s_{nj} and t_{nj_1} satisfy $s_{nj} \le s \le s_{nj+1}$ and $t_{nj} \le t \le t_{nj+1}$. Since $C(y) > 0$ for $y \in [s, t]$, and F^0 is Lipschitz continuous,

$$\int_s^t \frac{d(F_{n1} - F_1)}{C} \le \int_{s_{nj}}^{t_{nj+1}} \frac{d(F_{n1} - F_1)}{C} + O(\delta_n) := A_{nj} + O(\delta_n). \tag{3.3}$$

Each A_{nj} is an average of mean zero i.i.d. random variables with common bound denoted say, by M. Moreover, $n\,\mathrm{Var}(A_{nj}) \le \int_{s_{nj}}^{t_{nj+1}} \frac{dF_1}{C^2} = O(\epsilon_n)$. Bennett's or, Bernstein's inequality (see Pollard (1984), p. 193) gives

$$P(A_{nj} \ge \delta_n) \le 2\exp\left\{ \frac{-Kn\delta_n^2}{2\epsilon_n + \frac{2}{3}M\delta_n} \right\}$$

where $K(k_2)$ increases with k_2. Choose large enough K so that the last expression is $O(n^{-3})$. Since there are at most $O(n^{5/4})$ A_{nj}'s, Borel-Cantelli lemma gives $P(A_{nj} \ge \delta_n$

infinitely often) $= 0$. Or, equivalently, $\max A_{nj} \leq \delta_n$ eventually with probability one. Returning to (3.2), we now have the $O(\delta_n)$ upper bound for the oscillation modulus of $\hat{Z}_n/\sqrt{n}$. The lower bound is obtained in a similar manner.

Finally, Theorem 2.2 of Zhou and Yip (1999) allows us to obtain the corresponding statement for the PL-process. □

PROOF OF THEOREM 2.1. We continue to use the notation ϵ_n and δ_n as in Lemma 3.4. Let $s = Q_n(p)$ and $t = Q(p)$, $p_0 \leq p \leq p_1$, Lemma 3.3 yields $|s - t| = O(\epsilon_n)$. Applying Lemma 3.4 gives, almost surely,

$$F_n^0(Q_n(p)) - F_n^0(Q(p)) = F^0(Q_n(p)) - F^0(Q(p)) + O(\delta_n). \tag{3.4}$$

By Lemma 3.2, $F_n^0(Q_n(p))$ can be replaced by p up to $O(1/n)$. For the right hand side, a Taylor expansion of the first term about $Q(p)$ up to second order term gives, almost surely, $f^0(Q(p))[Q_n(p) - Q(p)] + O([Q_n(p) - Q(p)]^2) + O(\delta_n)$ for $p_0 \leq p \leq p_1$. Invoking Lemma 3.3 and rearranging terms in (3.4), we have, almost surely,

$$\sqrt{n} f^0(Q(p))[Q_n(p) - Q(p)] = \sqrt{n}[p - F_n^0(Q(p))] + O(\sqrt{n}\delta_n)$$

for $p_0 \leq p \leq p_1$. Since F^0 is continuous, $F^0(Q(p)) = p$. Recalling the definitions of the PL process Z_n and normed PL-quantile process ρ_n, we now have, almost surely,

$$\rho_n(p) = Z_n(Q(p)) + O\left(\frac{(\log n)^{3/4}}{n^{1/4}}\right) \tag{3.5}$$

for $p_0 \leq p \leq p_1$. Finally, Theorem 2.1 of Tse (2003) provides the strong Gaussian approximation statement for Z_n, and hence, by (3.5), also for ρ_n. This completes the proof of the theorem. □

PROOF OF THEOREM 2.2. This is similar to the proof of last theorem with the role of Theorem 2.1 of Tse (2003) replaced by Theorem 2.2 of the same paper. □

PROOF OF THEOREM 2.3. This theorem follows from the representation (3.5) and the uniform consistency rate of the PL-process established in Theorem 2.3 of Tse (2003). □

Acknowledgements

The author would like to thank the editor and the referee for their valuable comments. This work was partially supported by NSC Grant 89-2118-M-259-011, R.O.C.

REFERENCES

Aly, E. E., Csörgő, M. and Horváth, L. (1985). Strong approximations of the quantile process of the product-limit estimator, *Journal of Multivariate Analysis*, **16**, 185–210.

Bahadur, R. R. (1966). A note on quantiles in large samples, *Annals of Mathematical Statistics*, **37**, 577–580.

Borisov, I. S. (1982). Approximation of empirical fields constructed with respect to vector observations with dependent components, *Sibirskii Matematicheskii Zhurnal*, **23**, 31–41.

Burke, M. D., Csörgő, S. and Horváth, L. (1981). Strong approximations of some biometric estimates under random censorship, *Zeitschrift für Wahrscheinlichkeitstheorie Verwandte Gebiete*, **56**, 87–112.

Burke, M. D., Csörgő, S. and Horváth, L. (1988). A correction to and improvement of 'Strong approximations of some biometric estimates under random censorship', *Probability Theory and Related Fields*, **79**, 51–57.

Cheng, K. F. (1984). On almost sure representations for quantiles of the product-limit estimator with applications, *Sankhya A*, **46**, 426–443.

Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*, Academic Press, New York.

Gastwirth, J. L. (1971). A general definition of the Lorenz curve, *Econometrica*, **39**, 1037–1039.

Gijbels, I. and Wang, T. J. (1993). Strong representations of the survival function estimator for truncated and censored data with applications, *Journal of Multivariate Analysis*, **47**, 210–229.

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, *Annals of Probability*, **18**, 160–189.

Gürler, Ü., Stute, W. and Wang, J. L. (1993). Weak and strong quantile representations for randomly truncated data with applications, *Statistics & Probability Letters*, **17**, 139–148.

Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent r.v.'s and the sample d.f., *Zeitschrift für Wahrscheinlichkeitstheorie Verwandte Gebiete*, **32**, 111–132.

Lai, T. L. and Ying, Z. (1991). Estimating a distribution function with truncated and censored data, *Annals of Statistics*, **19**, 417–442.

Lo, S. H. and Singh, K. (1986). The product-limit estimator and the bootstrap: Some asymptotic representations, *Probability Theory and 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 Royal Astronomical Society*, **155**, 95–118.

Parzen, E. (1979). Nonparametric statistical data modeling, *Journal of American Statistical Association*, **74**, 105–131.

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer-Verlag, New York.

Serfling, R. (1980). Approximation theorems of mathematical statistics, *Wiley Series in Probability and Mathematical Statistics*, John Wiley and Sons, New York.

Tse, S. M. (2000). Strong Gaussian approximations in the random truncation model, *Statistica Sinica*, **10**, 281–296.

Tse, S. M. (2003). Strong Gaussian approximations in the left truncated and right censored model, *Statistica Sinica*, **13**, 275–282.

Wang, M. C., Jewel, N. P. and Tsai, W. Y. (1986). Asymptotic properties of the product-limit estimate under random truncation, *Annals of Statistics*, **14**, 1597–1605.

Woodroofe, M. (1985). Estimating a distribution function with truncated data, *Annals of Statistics*, **13**, 163–177.

Zhou, Y. (1996). A note on the TJW estimator for truncated and censored data, *Statistics & Probability Letters*, **26**, 381–387.

Zhou, Y. (2001). Properties of convergence of a smooth quantile estimator for truncated and censored data, *Chinese Journal of Applied Probability & Statistics*, **17**, 351–358.

Zhou, Y. and Yip, S. F. (1999). A strong representation of the product-limit estimator for left truncated and right censored data, *Journal of Multivariate Analysis*, **69**, 261–280.