Z. Wahrscheinlichkeitstheorie verw. Gebiete
56, 75–86 (1981)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

A LIL Type Result for the Product Limit Estimator

A. Földes and L. Rejtő

Mathematical Institute of the Hungarian Academy of Sciences,
Reáltanoda u. 13–15, Budapest, 1053, Hungary

Summary. Let $X_1, X_2, \ldots, X_n$ be i.i.d. r.v.'-s with $P(X>u)=F(u)$ and $Y_1, Y_2, \ldots, Y_n$ be i.i.d. $P(Y>u)=G(u)$ where both F and G are unknown continuous survival functions. For $i=1,2,\ldots,n$ set $\delta_i=1$ if $X_i \leqq Y_i$ and 0 if $X_i>Y_i$, and $Z_i=\min\{X_i, Y_i\}$. One way to estimate F from the observations (Z_i, δ_i) $i=1,\ldots,n$ is by means of the product limit (P.L.) estimator F_n^* (Kaplan-Meier, 1958 [6]).

In this paper it is shown that F_n^* is uniformly almost sure consistent with rate $O(\sqrt{\log\log n}/\sqrt{n})$, that is

$$P\Big(\sup_{-\infty<u<+\infty} |F_n^*(u)-F(u)| = O(\sqrt{\log\log n/n}\Big)=1$$

if $G(T_F)>0$, where $T_F=\sup\{x\colon F(x)>0\}$.

A similar result is proved for the Bayesian estimator [9] of F. Moreover a sharpening of the exponential bound of [3] is given.

1. Introduction

Let $X_1, \ldots, X_n$, resp. $Y_1, \ldots, Y_n$ be i.i.d. sequences of random variables with fixed unknown continuous survival function F, resp. G. Suppose that the two sequences are independent of each other. Set

$$\delta_i=[X_i \leqq Y_i] \quad \text{and} \quad Z_i=\min\{X_i, Y_i\}$$

for $i=1,2,\ldots,n$, where $[A]$ denotes the indicator function of the set A. One way to estimate F from the sample $\{Z_i, \delta_i\}$ $i=1,2,\ldots,n$ is by means of the *product limit* (PL) estimator [6]. The properties of this estimator are similar to the ones of the empirical distribution function. One reason of this fact is that in the uncensored case the PL is equal to the empirical d.f. Before giving the definition of the PL we introduce some notations.

0044-3719/81/0056/0075/$02.40

Let $P(Z>u)=H(u)$. Then $H(u)=F(u)G(u)$. Let us denote by

$$N^+(u,n)=N^+(u)=\#Z_j\text{-s greater than } u.$$
$$\beta_j(u)=[\delta_j=1,\ Z_j\leqq u].$$

The usual definition of the product limit estimator of the survival function of X, in case of continuous F and G is

$$F_n^*(u)=\begin{cases}\prod_{j=1}^{n}\left(\frac{N^+(Z_j)}{N^+(Z_j)+1}\right)^{\beta_j(u)} & \text{if } u\leqq\max\{Z_1,\dots,Z_n\}\\ 0 & \text{if } u>\max\{Z_1,\dots,Z_n\}.\end{cases}$$

In course of the proof we need the *modified product limit estimator*;

$$\bar{F}_n(u)=\begin{cases}\prod_{j=1}^{n}\left(\frac{N^+(Z_j)+1}{N^+(Z_j)+2}\right)^{\beta_j(u)} & \text{if } u\leqq\max\{Z_1,\dots,Z_n\}\\ 0 & \text{if } u>\max\{Z_1,\dots,Z_n\}.\end{cases}$$

Frequently the sub-distribution function

$$\tilde{F}(u)=P(Z\leqq u,\ \delta=1)=P(X=u,\ X-Y\leqq 0)$$
$$=\int_{-\infty}^{u} G(s)\,d(1-F(s))$$

will be used and its empirical distribution (not survival) is defined by

$$\tilde{F}_n(u)=\frac{1}{n}\sum_{i=1}^{n}[Z_i\leqq u,\ \delta_i=1]=\frac{1}{n}\sum_{i=1}^{n}\beta_i(u)$$
$$=\frac{1}{n}\sum_{i=1}^{n}[X_i\leqq u,\ X_i-Y_i\leqq 0].$$

The empirical survival function of H is

$$H_n(u)=\frac{1}{n}N^+(u).$$

$F^{-1}(u)$ denotes the inverse function of F. Further let

$$T_F=\sup\{u,\ F(u)>0\}.$$

T_G and T_H are defined similarly.

In [3] it was proved that

$$P\left(\sup_{0\leqq u\leqq T}|F_n^*(u)-F(u)|=O\left(\sqrt{\frac{\log n}{n}}\right)\right)=1,$$

where $H(T)>0$. This paper contains somewhat more general model, namely the r.v.-s X'-s, and Y'-s are not necessarily nonnegative, furthermore the convergence rate is considered on an interval $(-\infty, T_F)$. Supposing that $T_F<T_G$

($\leqq +\infty$) the upper part of the law of iterated logarithm is proved (Theorem 1), namely

$$P\left(\sup_{-\infty<u<+\infty} |F_n^*(u)-F(u)| = O\left(\sqrt{\frac{\log\log n}{n}}\right)\right)=1.$$

In [1] Burke, Csörgő and Horváth consider a whole class of product limit type statistics and prove a weaker rate result on $(-\infty, T_F)$ without the above mentioned condition. They also prove the $O\left(\sqrt{\frac{\log\log n}{n}}\right)$ rate on the interval $(-\infty, T]$, where $H(T)>0$, giving some information on the constant too, using the embedding theorem of [11].

Furthermore we give an exponential bound of the probability $P(\sup_{-\infty<u\leqq T} |F_n^*(u)-F(u)|>\varepsilon)$ where $H(T)>\delta>0$. This result is the sharpening of the one of paper [3] and the analog of Lemma 2 of [2].

The structure of the paper is the following:

Section 2 contains Theorem 1 which is a law of iterated logarithm type result. The exponential bound of the probability $P(\sup|F_n^*(u)-F(u)|>\varepsilon)$ is given in Sect. 3. Section 4 deals with the Bayesian estimator of F. The properties of this estimator were investigated e.g. in [9]. In Theorem 3 it is proved that the sup distance of the two estimators is $O\left(\sqrt{\frac{\log\log n}{n}}\right)$ and Theorem 1 and 2 are valid for the Bayesian estimator too.

2. The LIL Type Result

Theorem 1. *Suppose that F and G are continuous survival functions, further suppose that $T_F<T_G\leqq +\infty$. Then*

$$P\left(\sup_{-\infty<u<+\infty} |F_n^*(u)-F(u)| = O\left(\sqrt{\frac{\log\log n}{n}}\right)\right)=1.$$

The real meaning of Theorem 1 is that the sup can be taken on $(-\infty, T_H)$ if $T_H=T_F<+\infty$.

Theorem 1 has a corollary the proof of which easily follows from that of Theorem 1.

Corollary 1. *Suppose that F and G are continuous survival functions and $G(T)>0$. Then*

$$P\left(\sup_{-\infty<u\leqq T^*} |F_n^*(u)-F(u)| = O\left(\sqrt{\frac{\log\log n}{n}}\right)\right)=1,$$

where $T^*=\min(T, T_F)$.

For the proofs of the theorems we need some basic remarks and lemmas. First we give a decomposition of $F_n^*(u)-F(u)$. Let

$$T_n(u)=\int_{-\infty}^{u}\frac{1}{H_n(t)}\,d\tilde F_n(t),$$

$$T(u)=\int_{-\infty}^{u}\frac{1}{H(t)}\,d\tilde F(t)=-\log F(u)$$

where $u<T_F$. Then

$$|F_n^*(u)-F(u)|\leqq|F_n^*(u)-\bar F_n(u)|+|\bar F_n(u)-F(u)| \tag{2.1}$$

and

$$\bar F_n(u)-F(u)=(e^{\log\bar F_n(u)}-e^{-T_n(u)})+(e^{-T_n(u)}-e^{-T(u)}).$$

Using Taylor expansions of the two differences separately we have

$$\begin{aligned}\bar F_n(u)+F(u)=e^{-T_n^*(u)}(\log\bar F_n(u)+T_n(u))+F(u)\,(T(u)-T_n(u))\\ +\tfrac{1}{2}e^{-T_n^{**}(u)}(T(u)-T_n(u))^2\end{aligned} \tag{2.2}$$

where

$$\begin{aligned}\min\{-\log\bar F_n(u),\,T_n(u)\}\leqq T_n^*(u)\leqq\max\{-\log\bar F_n(u),\,T_n(u)\}\\ \min\{T(u),\,T_n(u)\}\leqq T_n^{**}(u)\leqq\max\{T(u),\,T_n(u)\}.\end{aligned} \tag{2.3}$$

In the lemmas the order of magnitude of the terms of (2.1) and (2.2) are given.

In all of the lemmas the sup distance is considered on an interval $(-\infty,u_n)$ where

$$u_n=F^{-1}\left(\frac{2A}{G(T_F)}\sqrt{\frac{\log\log n}{2n}}\right)$$

and $A\geqq 7$ is a fixed number. Further it is supposed that $G(T_F)>0$.

Remark 2.1. Suppose that $G(T_F)>0$. Then for almost all ω there exists an $n_0=n_0(\omega)$ such that

$$H_n(u)\geqq\tfrac{1}{2}H(u)\qquad\text{for } -\infty<u\leqq u_n$$

if $n>n_0(\omega)$.

From the condition $G(T_F)>0$, it follows that

$$H(u)=G(u)\,F(u)>G(T_F)\,F(u)\qquad\text{for } u<T_F.$$

By an obvious computation the statement follows from the law of the iterated logarithm for the empirical distribution function [7].

Remark 2.2. Let F and G be continuous, then

$$P\left(\limsup_{n\to\infty}\frac{\sqrt{n}\,\sup|\tilde F_n(u)-\tilde F(u)|}{\sqrt{(1/2)\log\log n}}\leqq 1\right)=1.$$

This remark is the consequence of the fact that $\tilde F_n(u)$ is the empirical distribution of the two-dimensional random vector $(X,X-Y)$ at the point $(u,0)$, and their joint continuous distribution is exactly

$$P(X\leqq u,\,X-Y\leqq 0)=\tilde F(u)$$

hence the LIL theorem for the multidimensional empirical distribution applies (see [7]).

Lemma 2.1. *For almost all ω there exists an $n_0(\omega)$ such that if $n>n_0(\omega)$ then for all $u \leqq u_n$ and*

$$k_1>0, k_2>0 \quad \text{where} \quad k=k_1+k_2>1$$

a) $\displaystyle\int_{-\infty}^{u} \frac{1}{H_n^{k_1}(t)\,H^{k_2}(t)}\,d\tilde{F}_n \leqq \frac{2^{k_1}}{G^{k-1}(T_F)}\,\frac{1}{F^{k-1}(u)}\left(\frac{2}{A}+\frac{1}{k-1}\right),$

b) $\displaystyle\int_{-\infty}^{u} \frac{1}{H_n^{k_1}(t)\,H^{k_2}(t)}\,d\tilde{F}_n = O\left(\frac{n}{\log\log n}\right)^{\frac{k-1}{2}}.$

Proof. By Remarks 2.1, 2.2

$$\int_{-\infty}^{u} \frac{1}{H_n^{k_1}(t)\,H^{k_2}(t)}\,d\tilde{F}_n(t) \leqq \left|\int_{-\infty}^{u} \frac{2^{k_1}}{H^k(t)}\,d\tilde{F}_n(t)\right|$$
$$\leqq \left|\int_{-\infty}^{u} \frac{2^{k_1}}{H^k(t)}\,d(\tilde{F}_n(t)-\tilde{F}(t))\right| + \left|\int_{-\infty}^{u} \frac{2^{k_1}}{H^k(t)}\,d\tilde{F}(t)\right|$$

holds if $n>n_1(\omega)$ and $u \leqq u_n$, where $n_1(\omega)$ is given by Remark 2.1.

Integrating by parts it can be seen that

$$\left|2^{k_1}\int_{-\infty}^{u} \frac{1}{H^k(t)}\,d(\tilde{F}_n(t)-\tilde{F}(t))\right| \leqq \frac{2^{k_1}}{G^k(T_F)}\,\frac{2\sup|\tilde{F}_n-\tilde{F}|}{F^k(u)}$$
$$\leqq \frac{2^{k_1+1}}{G^k(T_F)}\,\frac{\sup|\tilde{F}_n-\tilde{F}|}{F(u_n)}\,\frac{1}{F^{k-1}(u)} \leqq \frac{2^{k_1+1}}{A\cdot G^{k-1}(T_F)\,F^{k-1}(u)} \tag{2.4}$$

if $n>n_2(\omega)$.

On the other hand if $n>n_0(\omega)=\max(n_1(\omega), n_2(\omega))$

$$2^{k_1}\int_{-\infty}^{u} \frac{1}{H^k(t)}\,d\tilde{F}(t) \leqq \frac{2^{k_1}}{G^{k-1}(T_F)}\left(-\int_{-\infty}^{u} \frac{1}{F^k(t)}\,dF(t)\right)$$
$$=\frac{2^{k_1}}{G^{k-1}(T_F)}\left(-\int_{1}^{F(u)} \frac{1}{s^k}\,ds\right) = \frac{2^{k_1}}{G^{k-1}(T_F)}\,\frac{1}{k-1}\left(\frac{1}{F^{k-1}(u)}-1\right). \tag{2.5}$$

Thus a) follows from (2.4) and (2.5), and (b) is an easy consequence of (a).

Lemma 2.2.

$$\sup_{-\infty<u\leqq u_n} |F_n^*(u)-\bar{F}_n(u)| = O\left(\sqrt{\frac{1}{n\log\log n}}\right)$$

almost surely.

Proof. Using that

$$\left|\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i\right| \leqq \sum_{i=1}^{n} |a_i-b_i|$$

for $0 \leqq a_i \leqq 1$, $0 \leqq b_i \leqq 1$, $i = 1, 2, \ldots n$, we get

$$|F_n^*(u) - \bar{F}_n(u)| \leqq \sum_{j=1}^{n} \left| \left(\frac{N^+(Z_j)}{N^+(Z_j)+1}\right)^{\beta_j(u)} - \left(\frac{N^+(Z_j)+1}{N^+(Z_j)+2}\right)^{\beta_j(u)} \right|$$

$$\leqq \sum_{j=1}^{n} \frac{\beta_j(u)}{(N^+(Z_j)+1)^2}.$$

Observe that

$$\sum_{j=1}^{n} \frac{\beta_j(u)}{(N^+(Z_j)+1)^2} = \int_{-\infty}^{u} \frac{n}{(nH_n(t)+1)^2} d\tilde{F}_n(t) \leqq \frac{1}{n} \int_{-\infty}^{u} \frac{1}{H_n^2(t)} d\tilde{F}_n(t).$$

Hence

$$\sup_{-\infty < u \leqq u_n} |F_n^*(u) - \bar{F}_n(u)| \leqq \frac{1}{n} \int_{-\infty}^{u} \frac{1}{H_n^2(t)} d\tilde{F}_n(t)$$

and the statement follows from Lemma 2.1.

Lemma 2.3.

$$\sup_{-\infty < u \leqq u_n} |\log \bar{F}_n(u) + T_n(u)| = O\left(\sqrt{\frac{1}{n \log\log n}}\right)$$

almost surely.

Proof. Observe that

$$\log \bar{F}_n(u) = \int_{-\infty}^{u} n \log\left(1 - \frac{1}{nH_n(t)+2}\right) d\tilde{F}_n(t). \tag{2.6}$$

Using the logarithmic expansion we have

$$|\log \bar{F}_n(u) + T_n(u)|$$

$$= \left| \int_{-\infty}^{u} \left\{ n\left[-\sum_{l=1}^{\infty} \frac{1}{l} (2 + nH_n(t))^{-l} \right] + \frac{1}{H_n(t)} \right\} d\tilde{F}_n(t) \right|$$

$$\leqq \int_{-\infty}^{u} \left| \frac{1}{H_n(t)} - \frac{1}{\frac{2}{n} + H_n(t)} \right| d\tilde{F}_n(t) + \int_{-\infty}^{u} \frac{n}{(2 + nH_n(t))^2} d\tilde{F}_n(t)$$

$$\leqq 3 \int_{-\infty}^{u} \frac{1}{nH_n^2(t)} d\tilde{F}_n(t). \tag{2.7}$$

Applying Lemma 2.1 the statement follows.

Lemma 2.4.

a) $\sup_{-\infty < u \leqq u_n} |T_n(u) - T(u)| \leqq \frac{2}{3}$ *almost surely.*

b) $\sup_{-\infty < u \leqq u_n} F(u) |T_n(u) - T(u)| = O\left(\sqrt{\frac{\log\log n}{n}}\right)$ *almost surely.*

Proof. Using partial integration by Remarks 2.1, 2.2, Lemma 2.1 it follows that

$$|T_n(u)-T(u)| \leqq \int_{-\infty}^{u} \frac{|H_n(t)-H(t)|}{H_n(t)\,H(t)}\, d\tilde{F}_n(t) + \left|\int_{-\infty}^{u} \frac{1}{H(t)}\, d(\tilde{F}_n(t)-\tilde{F}(t))\right|$$

$$\leqq 2\sqrt{\frac{\log\log n}{2n}}\left(\frac{2(1+\frac{2}{A})}{F(u)\,G(T_F)} + \frac{2}{F(u)\,G(T_F)}\right). \tag{2.8}$$

Consequently

$$\sup_{-\infty<u\leqq u_n} |T_n(u)-T(u)| \leqq \frac{4(1+\frac{1}{A})}{A}.$$

Choosing $A \geqq 7$ we have part (a) of the lemma.

By (2.8) it follows that for all $u \leqq u_n$

$$F(u)\,|T_n(u)-T(u)| \leqq \frac{64}{7G(T_F)}\sqrt{\frac{\log\log n}{2n}}$$

which proves (b).

Lemma 2.5. *For* $-\infty < u \leqq u_n$

$$|\bar{F}_n(u)-F(u)| \leqq |\log \bar{F}_n(u) + T_n(u)| + 2F(u)\,|T_n(u)-T(u)|$$

almost surely.

Proof. Using decomposition (2.2) of $|\bar{F}_n(u)-F(u)|$

$$e^{-T_n^*(u)}\,|\log \bar{F}_n(u) + T_n(u)| \leqq |\log \bar{F}_n(u) + T_n(u)| \tag{2.9}$$

follows from (2.3). Furthermore

$$\tfrac{1}{2} e^{-T_n^{**}(u)} |T_n(u)-T(u)|^2 \leqq \tfrac{1}{2} F(u)\, e^{|T_n(u)-T(u)|} |T_n(u)-T(u)|^2. \tag{2.10}$$

By Lemma 2.4 $|T_n(u)-T(u)| \leqq \frac{2}{3}$, therefore the inequality $\frac{1}{2}e^x x^2 < x$ for $0 < x < \frac{2}{3}$ is applicable. Hence

$$\tfrac{1}{2} e^{-T_n^{**}(u)} |T_n(u)-T(u)|^2 \leqq F(u)\,|T_n(u)-T(u)| \tag{2.11}$$

if $n > n_0(\omega)$. These statement follows from (2.2), (2.9) and (2.11).

Proof of Theorem 1.

$$\sup_{-\infty<u<+\infty} |F_n^*(u)-F(u)| \leqq \sup_{-\infty<u\leqq u_n} |F_n^*(u)-F(u)| + \sup_{u_n<u<+\infty} |F_n^*(u)-F(u)|.$$

Clearly

$$\sup_{u_n<u<+\infty} |F_n^*(u)-F(u)| \leqq F(u_n) + |F_n^*(u_n)-F(u_n)|. \tag{2.12}$$

From inequalities (2.1), (2.2) and Lemma (2.5) follows that

$$|F_n^*(u)-F(u)| \leqq |F_n^*(u)-\bar{F}_n(u)| + |\log \bar{F}_n(u) + T_n(u)| + 2F(u)\,|T_n(u)-T(u)|.$$

Using Lemmas 2.2–2.4 we have

$$\sup_{-\infty<u\leqq u_n} |F_n^*(u)-F(u)| = O\left(\sqrt{\frac{\log\log n}{n}}\right), \tag{2.13}$$

if $n>n_0(\omega)$. Now the theorem follows from (2.12) and (2.13).

3. The Exponential Bound

In this section an exponential bound is given to the probability $P(\sup_{-\infty<u\leqq T} |F_n^*(u)-F(u)|>\varepsilon)$. It is supposed that at the point T, $H(T)>\delta>0$.

The result of Theorem 2 is the analog of Lemma 2 of paper [2]. Although we have no such type of result on the whole line, it is conjectured that it is valid on $(-\infty, +\infty)$ as well.

Theorem 2. *Suppose that $H(T)>\delta>0$ and $\varepsilon>\frac{2^7}{n\cdot\delta^2}$. Then*

$$P(\sup_{-\infty<u\leqq T} |F_n^*(u)-F(u)|>\varepsilon)\leqq d_0 \exp\{-n\varepsilon^2\delta^4 d_1\}$$

where d_0 and d_1 are absolute constants, which do not depend on F, G, H.

The following Remark is a trivial consequence of Lemma 1 of Wellner [10].

Remark 3.1. Suppose that $H(T)>\delta>0$. Set

$$D_n=\left\{\omega:\ \sup_{-\infty<u\leqq T}\frac{H(u)}{H_n(u)}\geqq 2\right\}.$$

Then

$$P(D_n)\leqq \exp\{-n\delta c_1\}. \tag{3.1}$$

The proof of Theorem 2 is similar to that of Theorem 1 and based on Remark 3.1, and the multidimensional exponential bound of Kiefer [7], hence we only indicate the necessary steps by stating all the lemmas.

Lemma 3.1. *Suppose that $H(T)>\delta>0$, and*

$$\varepsilon>\frac{2^{k_1+1}}{\delta^k}.$$

Then

$$P\left(\int_{-\infty}^{T}\frac{1}{H_n^{k_1}(t)\,H^{k_2}(t)}\,d\tilde{F}_n(t)>\varepsilon\right)\leqq \exp\{-n\delta c_1\}+c_3\exp\{-n\varepsilon^2\delta^{2k}c_2 2^{-2(k_1+2)}\}$$

where $k_1\geqq 0$, $k_2\geqq 0$, $k_1+k_2=k$.

Lemma 3.2. *Suppose that $H(T)>\delta$ and $\varepsilon>\frac{8}{n\delta^2}$. Then*

$$P(\sup_{-\infty<u\leqq T}|F_n^*(u)-\tilde{F}_n(u)|>\varepsilon)\leqq \exp\{-n\delta c_1\}+c_3\exp\{-n^3\varepsilon^2\delta^4 2^{-6}c_2\}.$$

Lemma 3.3. *Suppose that* $H(T)>\delta$ *and* $\varepsilon>\frac{32}{n\delta^2}$. *Then*

$$P(\sup_{-\infty<u\leqq T}|\log \bar{F}_n(u)+T_n(u)|>\varepsilon)\leqq \exp\{-n\delta c_1\}+c_3\exp\{-n^3\varepsilon^2\delta^4 2^{-12}c_2\}.$$

Lemma 3.4. *Suppose that* $H(T)>\delta$. *Then*

$$P(\sup_{-\infty<u\leqq T}|T_n(u)-T(u)|>\varepsilon)$$
$$\leqq \exp\{-n\delta c_1\}+2\exp\{-n\varepsilon^2\delta^4 2^{-4}c_4\}+c_3\exp\{-n\varepsilon^2\delta^2 2^{-4}c_2\}.$$

Proof of Theorem 2. Using (2.1), (2.2), (2.3) we have

$$\begin{aligned} P(\sup_{-\infty<u\leqq T}|F_n^*(u)-F(u)|>\varepsilon)&\leqq P\left(\sup_{-\infty<u\leqq T}|F_n^*(u)-\bar{F}_n(u)|>\frac{\varepsilon}{4}\right)\\ &+P\left(\sup_{-\infty<u\leqq T}|\log \bar{F}_n(u)+T_n(u)|>\frac{\varepsilon}{4}\right)\\ &+P\left(\sup_{-\infty<u\leqq T}F(u)|T_n(u)-T(u)|>\frac{\varepsilon}{4}\right)\\ &+P\left(\sup_{-\infty<u\leqq T}|T_n(u)-T(u)|^2>\frac{\varepsilon}{4}\right). \end{aligned} \tag{3.2}$$

Lemmas 3.2 and 3.3 give exponential bounds of the first two terms of the right-hand side. Observe that $F(u)\leqq 1$, and

$$P\left(\sup_{-\infty<u\leqq T}\tfrac{1}{2}|T_n(u)-T(u)|^2>\frac{\varepsilon}{4}\right)$$
$$=P\left(\sup_{-\infty<u\leqq T}|T_n(u)-T(u)|>\sqrt{\frac{\varepsilon}{2}}\right).$$

Hence the last two terms of (3.2) can be estimated by Lemma 3.4. This proves the theorem.

Remark 4.2. c_1 is constant coming from Wellner's lemma. c_2 and c_3 are the constants of [7] and c_4 is the constant of [2].

4. Results for the Bayesian Estimator

It is well-known that instead of the PL estimator the so-called Bayesian estimator can also be applied in case of nonnegative random variables X and Y. The definition of this estimator is the following:

Let α be an arbitrary non-null positive measure on the Borel σ-field of $(0,\infty)$. If $1-F$ is assumed to be a random continuous distribution function with Dirichlet process prior with parameter α, the Bayes estimator of the survival function $F(u)$ of the nonnegative random variable X is

$$F_n^\alpha(u)=\frac{N^+(u)+\alpha(u)}{n+\alpha(0)}\prod_{i=1}^{n}\left(\frac{N^+(Z_i)+\alpha(Z_i)+1}{N^+(Z_i)+\alpha(Z_i)}\right)^{[\delta_i=0,\ Z_i\leqq u]} \tag{4.1}$$

where $\alpha(u)=\alpha([u;\ +\infty))$ $\alpha(Z_i)=\alpha([Z_i;\ +\infty))$ and the censoring random variable is also assumed to be nonnegative.

It turned out that this $F_n^\alpha(u)$ estimator has similar properties to the P.L. estimator. An elegant way to show that a property of one of the above two estimators also possessed by the other one is to estimate the distance $\sup|F_n^*(u)-F_n^\alpha(u)|$ appropriately. The first paper considering the distance of the two estimators was published by Phadia and Van Ryzin [8]. They have proved that

$$E(F_n^*(u)-F_n^\alpha(u))^2=O\left(\frac{1}{n^2}\right),$$

for every fixed $u\,(\alpha(u)>0)$. From this result it follows that the pointwise strong consistency of F_n^* with rate $O\left(\frac{\log n}{\sqrt{n}}\right)$ implies the same property of F_n^α. In our paper [4] it was proved that the sup distance of the two estimators in any fixed interval $[0,T]$ (where $\alpha(T)>0$) is $O\left(\frac{1}{n}\right)$ almost surely. (Corollary 4.1). At the same time an exponential bound was given to the probability

$$P(\sup_{0<u\leqq T}|F_n^*(u)-F_n^\alpha(u)|>\varepsilon)$$

as well. Therefore it follows that the analog of Theorem 2 is valid for the Bayesian estimator. Our present aim is to extend the result of [4] to the interval $(0,T_F)$ which implies the strong uniform consistency of $F_n^\alpha(u)$ in $(0,T_F)$ with rate $O\left(\sqrt{\frac{\log\log n}{n}}\right)$.

Lemma 4.1. *Let X and Y be nonnegative random variables. Suppose that*

$$\alpha(T_F)>0,\qquad G(T_F)>0.$$

Then for $n>n_0(\omega)$

$$\sup_{0\leqq u\leqq u_n}|F_n^*(u)-F_n^\alpha(u)|=O\left(\frac{1}{\sqrt{n}\log\log n}\right)\qquad\text{a.s.}$$

where the O depends only on $G(T_F)$ and $\alpha(0)$.

Proof. We summarize some notations and facts from [8] and [4] (Sect. 6). Denote by $Z_{(1)}\leqq Z_{(2)}\leqq\ldots\leqq Z_{(n)}$ the ordered sample $Z_1,\ldots,Z_n$ and by $\delta_{(i)}$ $(i=1,\ldots n)$ the δ corresponding to $Z_{(i)}$ in the original sample.

For $Z_{(i)}\leqq u<Z_{(i+1)}$ $i=0,1,\ldots,n-1$ with $Z_{(0)}=0$,

$$|F_n^*(u)-F_n^\alpha(u)|\leqq\left|\frac{n-i}{n}\left(\prod_{j=1}^{i}A_j-\prod_{j=1}^{i}B_j\right)\right|+\prod_{j=1}^{i}B_j\left|\frac{n-i}{n}-\frac{n-i+\alpha(u)}{n+\alpha(0)}\right| \tag{4.2}$$

where

$$A_j = \left(\frac{n-j+1}{n-j}\right)^{[\delta_{(j)}=0]} \quad \text{and} \quad B_j = \left(\frac{n-j+\alpha(Z_{(j)})+1}{n-j+\alpha(Z_{(j)})}\right)^{[\delta_j=0]}$$

$$\left|\frac{n-i}{n}\left(\prod_{j=1}^{i} A_j - \prod_{j=1}^{i} B_j\right)\right| \leqq \sum_{j=1}^{i} |A_j - B_j| \tag{4.3}$$

and

$$\prod_{j=1}^{i} B_j \left|\frac{n-i}{n} - \frac{n-i+\alpha(u)}{n+\alpha(0)}\right| \leqq \frac{\alpha(0)}{n-i}. \tag{4.4}$$

Observe that by the definition of A_j and B_j

$$\begin{aligned}\sum_{j=1}^{i} |A_j - B_j| &= \sum_{j=1}^{i} [\delta_{(j)}=0]\,|A_j - B_j| \\ &\leqq \sum_{j=1}^{i} \frac{[\delta_{(j)}=0]\,\alpha(Z_{(j)})}{(n-j)(n-j+\alpha(Z_{(j)}))} \leqq \alpha(0) \sum_{j=1}^{i} \frac{[\delta_{(j)}=0]}{(n-j)^2} \\ &\leqq \alpha(0) \sum_{j=1}^{i} \frac{[\delta_{(j)}=0]}{(n H_n(Z_{(j)}))^2}.\end{aligned}$$

Furthermore for $n > n_0(\omega)$ the largest observation $Z_{(n)} > u_n$ almost surely: as

$$\begin{aligned}\sum_{n=1}^{\infty} P(Z_{(n)} < u_n) &\leqq \sum_{n=1}^{\infty} \prod_{j=1}^{n} P(Z_j < u_n) = \sum_{n=1}^{\infty} (1-H(u_n))^n \\ &\leqq \sum_{n=1}^{\infty} e^{-nH(u_n)} \leqq \sum_{n=1}^{\infty} e^{-nG(T_F)F(u_n)} \leqq \sum_{n=1}^{\infty} e^{-2A\sqrt{\frac{n\log\log n}{2}}} < +\infty.\end{aligned}$$

Hence from (4.2) and (4.4) it follows that for $n > n_0(\omega)$ we have almost surely

$$\sup_{0 \leqq u \leqq u_n} |F_n^*(u) - F_n^\alpha(u)| \leqq \alpha(0) \left\{ \int_0^{u_n} \frac{1}{n H_n^2(u)}\, d\tilde{G}_n(u) + \frac{1}{n H_n(u_n)} \right\} \tag{4.5}$$

where

$$\tilde{G}_n(u) = \frac{1}{n} \sum_{j=1}^{n} [Z_j \leqq u, \delta_j = 0].$$

Using the obvious fact that $\tilde{G}_n(u) = 1 - H_n(u) - \tilde{F}_n(u)$ and similar argument to Lemma 2.1, it can be seen that

$$\int_0^{u_n} \frac{1}{n H_n^2(u)}\, d\tilde{G}_n(u) = O\left(\frac{1}{\sqrt{n \log\log n}}\right). \tag{4.6}$$

Hence applying Remark 2.1

$$\sup_{0 \leqq u \leqq u_n} |F_n^*(u) - F_n^\alpha(u)| = O\left(\frac{1}{\sqrt{n \log\log n}}\right). \tag{4.7}$$

which proves the lemma.

Since $F_n^\alpha(u)$ is decreasing, it is obvious that

$$\sup_{0\leqq u\leqq +\infty} |F_n^\alpha(u)-F(u)| \leqq \sup_{0\leqq u\leqq u_n} |F_n^\alpha(u)-F(u)|+F(u_n)$$
$$\leqq \sup_{0\leqq u\leqq u_n} |F_n^\alpha(u)-F_n^*(u)| + \sup_{0\leqq u\leqq u_n} |F_n^*(u)-F(u)|+F(u_n). \tag{4.8}$$

By (4.8) and Lemma 4.1 we get the following

Theorem 3. *Suppose that F and G are continuous survival functions and $T_F<T_G\leqq+\infty$. Then if $\alpha(T_F)>0$*

$$P\left(\sup_{-\infty<u<+\infty} |F_n^\alpha(u)-F(u)| = O\left(\sqrt{\frac{\log\log n}{n}}\right)\right)=1.$$

References

1. Burke, M.D., Csörgő, S., Horváth, L.: Strong approximation of some biometric estimates under random censorship. Z. Wahrscheinlichkeitstheorie verw. Gebiete **56**, 87–112 (1981)
2. Dvoretzky, A., Kiefer, I., Wolfowitz, I.: Asymptotic minmax character of the sample distribution function and the classical multinomial estimator. Ann. Math. Stat. **27**, 642–669 (1956)
3. Földes, A., Rejtő, L.: Strong uniform consistency for nonparametric survival estimators from randomly censored data. Annals of Statist. **9** (1981) (to appear)
4. Földes, A., Rejtő, L.: Asymptotic properties of the nonparametric survival curve estimators under variable censoring. Proc. of First Pannon Conf. in Statist. Springer-Verlag (To appear)
5. Földes, A., Rejtő, L., Winter, B.B.: Strong consistency properties of nonparametric estimators for randomly censored data. Periodica Math. Hung. **11**, 233–250 (1980)
6. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observation. J. Amer. Statist. Assoc. **58**, 457–481 (1963)
7. Kiefer, I.: On large deviations of the empiric d.f. of vector chance variables and a law of the iterated logarithm. Pacific J. Math. **11**, 649–660 (1961)
8. Phadia, E.G., Van Ryzin, J.: A note on convergence rates for the P.L. estimator. [To appear]
9. Susarla, V., van Ryzin J.: Large sample theory for a Bayesian nonparametric survival curve estimator based on censored samples. Ann. Statist. **6**, 755–768 (1978)
10. Wellner, Jon A.: Limit theorems for the ratios of the true distribution. Z. Wahrscheinlichkeitstheorie verw. Gebiete **45**, 73–88 (1978)
11. Komlós, J., Major, P., Tusnády, G.: An approximation of partial sums of independent random variables and the sample distribution function I. Z. Wahrscheinlichkeitstheorie verw. Gebiete **32**, 111–131 (1975)

Received December 19, 1979