stichting
mathematisch
centrum

AFDELING MATHEMATISCHE STATISTIEK SW 74/81 AUGUSTUS
(DEPARTMENT OF MATHEMATICAL STATISTICS)

R.D. GILL

LARGE SAMPLE BEHAVIOUR OF THE PRODUCT-LIMIT ESTIMATOR ON THE WHOLE LINE

Preprint

kruislaan 413 1098 SJ amsterdam

Printed at the Mathematical Centre, 413 Kruislaan, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

1980 Mathematics subject classification: 62G05,62G15,60F05,62N05,62P10

Large sample behaviour of the product-limit estimator on the whole line*)

by

Richard Gill

ABSTRACT

Weak convergence results are proved for the product-limit estimator on the whole line. Applications are given to confidence band construction, estimation of mean lifetime, and to the theory of q-functions. The results are obtained using stochastic calculus and in probability linear bounds for empirical processes.

KEY WORDS & PHRASES: *Product-limit estimator, Kaplan-Meier estimator, random censorship; survival data, confidence bands, mean life-time, couting processes, martingales, stochastic integrals, weak convergence, in probability linear bounds, q-functions*

---

*) This report will be submitted for publication elsewhere.

1. INTRODUCTION

Let $X_1,\ldots,X_n$ be independent positive random variables with common continuous distribution function F. Independent of the X_i's, let $U_1,\ldots,U_n$ be also independent positive random variables with possibly non-continuous and defective common distribution function G. The problem at hand is to make nonparametric inference on F based on the *censored observations* $(\tilde{X}_i,\delta_i)$, $i = 1,\ldots,n$, defined by

$$\tilde{X}_i = X_i \wedge U_i, \quad \delta_i = I\{X_i \le U_i\},$$

where $\wedge$ denotes minimum and $I\{\cdot\}$ is the indicator random variable of the specified event. Classically, F is estimated by the product-limit estimator $\hat{F}$, introduced by KAPLAN & MEIER (1958). This estimator is in a reasonable sense the maximum likelihood estimator of F (see JOHANSEN (1978), SCHOLZ (1980) and WELLNER (1981)) and reduces to the ordinary empirical distribution function when there is no censoring. Defining processes N and Y on $[0,\infty)$ by

$$N(t) = \#\{i: \tilde{X}_i \le t, \delta_i = 1\},$$

$$Y(t) = \#\{i: \tilde{X}_i \ge t\},$$

then $\hat{F}$ is given by

$$1 - \hat{F}(t) = \prod_0^t \left(1 - \frac{dN(s)}{Y(s)}\right),$$

or in more abbreviated notation

$$1 - \hat{F} = \prod\left(1 - \frac{dN}{Y}\right).$$

By $\prod_0^t$ and $\int_0^t$ we mean product integration (see JOHANSEN (1978)) and ordinary integration over the interval $(0,t]$. When we drop the limits of integration, we implicitly define a function or process $t \to \prod_0^t(\cdot)$ or $t \to \int_0^t(\cdot)$. The estimator $\hat{F}$ is seen to be an intuitively reasonable one when we note that

(whether F is continuous or not)

$$1 - F = \prod(1 - \frac{dF}{1 - F_-}),$$

while the cumulative hazard function $\int dF/(1-F_-)$ is naturally estimated by $\int dN/Y$.

Define also the random time T by

$$T = \max_i \tilde{X}_i$$

and for any process W define the *stopped process* W^T by

$$W^T(t) = W(t \wedge T).$$

Note that $\hat{F}^T = \hat{F}$, and that if $\Delta N(T) = N(T) - N(T-) = 0$, i.e. the largest observation is censored, then $\hat{F}(T) < 1$ almost surely. Some authors prefer in this case to set $\hat{F}$ equal to 1 on (T,∞) while others leave it undefined there.

Let H be the distribution function of the $\tilde{X}_i$'s, given by

$$(1 - H) = (1 - F)(1 - G),$$

and define (possibly infinite) times τ_F, τ_G and τ_H by

$$\tau_F = \sup\{t: F(t) < 1\}$$

and similarly for τ_G and τ_H. Of course

$$\tau_H = \tau_F \wedge \tau_G.$$

Define also some continuous, nonnegative, nondecreasing functions Λ, C and K by

$$\Lambda(t) = \int_0^t \frac{dF(s)}{1-F(s-)},$$

$$C(t) = \int_0^t \frac{dF(s)}{(1-F(s-))^2(1-G(s-))} = \int_0^t \frac{d\Lambda(s)}{(1-H(s-))}$$

and

$$K(t) = \frac{C(t)}{1+C(t)},$$

where $K(t) = 1$ if $C(t) = \infty$. Note that $\Lambda(\tau_F) = \infty$, and that $C(\tau_H) = \infty$ and $K(\tau_H) = 1$ if $\tau_G \geq \tau_F$. When $\tau_G < \tau_F$ it is both possible that $C(\tau_H) = \infty$ and $C(\tau_H) < \infty$. (We write e.g. $1-F(s-)$ even though F is continuous to indicate the right extension for non-continuous F.)

Let B be a standard Brownian motion on $[0,\infty)$ and let B^0 be a Brownian bridge on $[0,1]$. Assuming G to be continuous (we shall see later that this condition is unnecessary), BRESLOW & CROWLEY (1974) proved a result on weak convergence of $n^{\frac{1}{2}}(\hat{F}-F)$ equivalent to the following theorem. The two ways in which we state it derive from EFRON (1967) and HALL & WELLNER (1980).

<u>THEOREM 1.1</u>. *For any* τ *such that* $H(\tau-) < 1$

$$n^{\frac{1}{2}}\left(\frac{\hat{F}-F}{1-F}\right) \xrightarrow{\mathcal{D}} B(C)$$

in $D[0,\tau]$ *as* $n \to \infty$*, or equivalently*

$$n^{\frac{1}{2}}\,\frac{1-K}{1-F}(\hat{F}-F) \xrightarrow{\mathcal{D}} B^0(K)$$

in $D[0,\tau]$ *as* $n \to \infty$.

Note that B(C) is a continuous Gaussian martingale, zero at time zero, with covariance function

$$\mathrm{cov}[B(C(s)),B(C(t))] = C(s) \wedge C(T) = C(s\wedge t).$$

Since $(1+C)^{-1} = (1-K)$ the equivalence between the two versions of Theorem 1.1 follows from the fact, easy to verify, that

$$(1+C)^{-1}B(C) \stackrel{\mathcal{D}}{=} B^0(K),$$

see DOOB (1949). BRESLOW & CROWLEY's (1974) proof of Theorem 1.1 was based on approximating $n^{\frac{1}{2}}(\hat{F}-F)$ by an expression linear in the empirical processes N/n and Y/n, and then applying standard results on weak convergence of empirical distribution functions. Though this sounds straightforward, the proof was unavoidably complex.*) For instance a Skorohod-Dudley construction was needed in order to take care of certain remainder terms. Also the simple form of the limiting distribution only appears after long calculations in which complicated expressions surprisingly cancel out. We shall later see why $n^{\frac{1}{2}}(\hat{F}-F)$ has the particular limiting distribution it does. But first let us mention another reason why we have stated Theorem 1.1 in the way we have chosen. Since F and also K can be uniformly consistently estimated on $[0,\tau]$ for any τ such that $H(\tau-) < 1$, the theorem gives two obvious ways of constructing confidence bands for F on $[0,\tau]$ based on the known distributions of $\sup_{x \le C(\tau)} |B(x)|$ and $\sup_{x \le K(\tau)} |B^0(x)|$, respectively, see GILL (1980) and HALL & WELLNER (1980). Of course $C(\tau)$ and $K(\tau)$ have to be estimated too.

Clearly there is hope that the Brownian bridge version of Theorem 1.1 could be extended to $[0,\tau_H]$ giving confidence bands for F on the whole line. When $K(\tau_H) = 1$ one would be able to use the distribution of $\sup_{x \le 1} |B^0(x)|$, leading to simpler computations too. Finally, when there is no censoring ($G = 0$), such bands would reduce to the usual Kolmogorov bands for F based on the empirical distribution function $\hat{F}$. Such a result was conjectured to hold by HALL & WELLNER (1980) and motivated the work presented here. We can only partially confirm the conjecture, but the techniques used will turn out to be of wider application.

Let us first define $\hat{K}$ by

$$1-\hat{K} = 1/(1+\hat{C})$$

and

$$\hat{C}(t) = \int_0^t \frac{n\,dN(s)}{Y(s)(Y(s)-1)}$$

---

*) A misprint in the proof of the central theorem, where the expression three lines from below on page 447 should read $2\varepsilon + 2K_{\rho_T}(X(1-F)^{-2},0)\rho_T(\tilde{F}_N^e,\tilde{F})$, has even led BURKE, CSÖRGŐ & HORVATH (1981) to claim the proof was incorrect.

(so $\hat{K}(T) = 1$ and $\hat{C}(T) = \infty$ almost surely if $\Delta N(T) = 1$). Then we shall prove

THEOREM 1.2.

$$n^{\frac{1}{2}}\left(\frac{1-K}{1-F}(\hat{F} - F)\right)^T \xrightarrow{\mathcal{D}} B^0(K)$$

in $D[0,\tau_H]$ *as* $n \to \infty$

and

THEOREM 1.3.

$$n^{\frac{1}{2}}\left(\frac{1-\hat{K}}{1-\hat{F}}(\hat{F}-F)\right)^T \xrightarrow{\mathcal{D}} B^0(K)$$

in $D[0,\tau_H]$ *as* $n \to \infty$, *provided that*

$$(1.1) \qquad \int_0^{\tau_H} \frac{dF(t)}{1 - G(t-)} < \infty.$$

(*If* $\hat{F}(T) = 1$ *we interpret* $(1-\hat{K})/(1-\hat{F})$ *in the point* T *as equal to its value in* T- .)

Whether or not Theorem 1.3 holds without the condition (1.1) we do not know. We make some more comments on this matter in Remark 3.1. Note that

$$(1.2) \qquad 1 - G \le \frac{1-K}{1-F} \le 1$$

and that $\frac{1-K}{1-F}$ is nonincreasing. The same relationship holds between $\hat{K}$, $\hat{F}$ and $\hat{G}$, where $\hat{G}$ is the product-limit estimator of the censoring distribution G. (These facts follow from the equality

$$d((1+C)(1-F)) = \int \frac{dG}{(1-F)(1-G)(1-G_-)}\, dF$$

which is also valid for non-continuous F.) Thus the factors $(1-K)/(1-F)$ and $(1-\hat{K})/(1-\hat{F})$ can be considered as weighting factors which compensate for the greater variability (increasing as t increases) of $n^{\frac{1}{2}}(\hat{F}-F)$ when there is censoring.

When there is no censoring $\hat{F}$ becomes the ordinary empirical distribution function of the X_i's, also $1-K$ becomes $1-F$ and $1-\hat{K}$ becomes $1-\hat{F}$, so apart from being "stopped at T" the theorems reduce to the classical result on weak convergence of $n^{\frac{1}{2}}$ times the centred empirical distribution function to a time transformed Brownian bridge. For practical purposes, stopping at T is of no consequence at all. The value of $\hat{F}(T)$ is the only information about F we have on (T,∞). Theorem 1.3 gives asymptotic confidence bands for F on the random interval $[0,T]$ (provided censoring is not too heavy, if condition (1.1) is really needed).

The result does have some practical importance. One would be tempted to apply Theorem 1.1 after choosing τ such that $Y(\tau)$ is reasonably large. So in fact τ will not be fixed in advance. Moreover, what is felt to be "reasonably large" may well be numbers as small as 5 or 10. However, Theorem 1.1 has the implicit condition $Y(\tau) \xrightarrow{P} \infty$. Thus it is not obvious that Theorem 1.1 will yield accurate approximations when applied in such a way. (See also GILLESPIE & FISHER (1979) and CSÖRGŐ & HORVÁTH (1981) for some indications that the approximations are not too good anyway with heavy censoring and small to medium sample sizes.)

In proving Theorems 1.2 and 1.3 it turns out that a technique is being used which leads to quite general results on *weighted* or *integrated* Kaplan-Meier processes. More specifically, defining

$$Z = n^{\frac{1}{2}} \frac{(\hat{F}-F)}{1-F},$$

then we can prove weak convergence results on the whole line for the processes hZ, $\int Zdh$, and $\int hdZ$, where h is a nonincreasing, nonnegative, continuous function. (Recall that by $\int Zdh$ we mean the process defined by $(\int Zdh)(t) = \int_{(0,t]} Z(s)dh(s)$.) In view of Theorem 1.1, the natural condition for this to be possible is

$$(1.3) \qquad \int_0^{\tau_H} h(s)^2 dC(s) < \infty$$

and that turns out to be all we need. Extensions are possible to random h functions and the restrictions on h may be weakened in other ways too. However (1.3) may not be sufficient then; cf. Theorem 1.3 where condition

(1.1) corresponds to (1.3) with $h = 1 - F$ rather than the natural $h = 1 - K$ (for which (1.3) always holds, giving Theorem 1.2).

In the next section we therefore state and prove such a general theorem. In Section 3 we show that Theorems 1.2 and 1.3 are corollaries of this, and go on to present other applications to estimating mean lifetime and to the theory of q-functions.

The key to our methods is that Z^T is a martingale on $[0,\tau_F)$, a fact which was discovered by AALEN & JOHANSEN (1978), though in the case when there is no censoring it has been known for a long time. We use the theory of counting processes developed by AALEN (1978), which depends on the theory of continuous time martingales and stochastic integration of MEYER (1976). We refer to AALEN (1978) or GILL (1980a) for brief surveys which should be sufficient for our purposes. After showing that Z^T is a martingale, we can derive our results using the martingale central limit theorem of REBOLLEDO (1980) and various martingale inequalities, as well as a few results on empirical distribution functions.

All our results can be easily extended to non-identically distributed censoring variables, and in particular therefore to the model of fixed censorship, which is more appropriate when censoring is due to limited observation times rather than to competing risks. Also the results can be extended to noncontinuous F. However, since the extra technicalities would only obscure the structure of our proofs we refer to GILL (1980a) for such a fuller treatment, where part of the material here has already appeared together with a study of two-sample tests.

2. MAIN THEOREM

<u>Theorem 2.1</u>. *Let* h *be a nonnegative continuous nonincreasing function on* $[0,\tau_H]$ *such that*

$$(2.1) \qquad \int_0^{\tau_H} h(t)^2 dC(t) < \infty.$$

Define

$$(2.2) \qquad Z = n^{\frac{1}{2}}\left(\frac{\hat{F}-F}{1-F}\right).$$

Then the processes $(hZ)^T$, $(\int hdZ)^T$ *and* $(\int Zdh)^T$ *converge jointly in* $D[0,\tau_H]$ *in distribution to processes* $hZ^{(\infty)}$, $\int hdZ^{(\infty)}$ *and* $\int Z^{(\infty)}dh$ *respectively, where*

$$Z^{(\infty)} = B(C) \tag{2.3}$$

and

$$hZ^{(\infty)} = \int hdZ^{(\infty)} + \int Z^{(\infty)}dh. \tag{2.4}$$

REMARK 2.2. When $C(\tau_H) = \infty$, the limiting processes here are interpreted to be equal in the point τ_H to their limits as $\tau \uparrow \tau_H$, which do exist as we shall now show. In fact the limit of $hZ^{(\infty)}$ is zero. Also we must discuss in any case what we mean by the process $\int hdZ^{(\infty)}$, which cannot be defined by pathwise Lebesgue-Stieltjes integration. Taking up the latter point first, we note that $\int hdZ^{(\infty)}$ can be defined on $[0,\tau_H)$ either by (2.4) or as a stochastic integral in the sense of MEYER (1976). By GILL (1980b) Lemma 5, 2nd part, the two definitions coincide. By (2.1), $\int hdZ^{(\infty)}$ is a square integrable martingale on $[0,\tau_H)$ which can be extended by taking limits to $[0,\tau_H]$. So it remains to show that either $hZ^{(\infty)}$ or $\int Z^{(\infty)}dh$ also has a limit almost surely as $t \to \tau_H$ (in which case both processes do). Now we can write

$$h(t)^2C(t) = \int_0^{\tau_H} h^2(t)I_{[0,t]}(s)dC(s).$$

So when $C(\tau_H) = \infty$ and consequently by (2.1) $h(t) \downarrow 0$ as $t \uparrow \tau_H$, by dominated convergence

$$h(t)^2C(t) \to 0 \quad \text{as } t \uparrow \tau_H \tag{2.5}$$

(this is essentially the Kronecker lemma with integrals instead of sums). It follows immediately that when $C(\tau_H) = \infty$

$$h(t)Z^{(\infty)}(t) \xrightarrow{P} 0 \quad \text{as } t \uparrow \tau_H; \tag{2.6}$$

we must extend this to an a.s. result. By the Birnbaum-Marshall inequality

(BIRNBAUM & MARSHALL (1961)) applied to the submartingale $(Z^{(\infty)}-Z^{(\infty)}(t))^2$ and the nonincreasing function h^2 on $[t,\tau_H)$, we have

$$P\left[\sup_{s\in[t,\tau_H)} ((Z^{(\infty)}(s)-Z^{(\infty)}(t))h(s))^2 \geq \varepsilon\right] \leq \frac{1}{\varepsilon}\int_t^{\tau_H} h(s)^2 dC(s).$$

Therefore

$$P\left[\sup_{[t,\tau_H)} (hZ^{(\infty)} - h(t)Z^{(\infty)}(t))^2 \geq 2\varepsilon\right]$$

$$\leq \frac{1}{\varepsilon}\int_t^{\tau_H} h^2 dC + P[h(t)^2 Z^{(\infty)}(t)^2 \geq \varepsilon]$$

$$\leq \frac{1}{\varepsilon}\left(\int_t^{\tau_H} h^2 dC + h(t)^2 C(t)\right).$$

Now let $\varepsilon_m > 0$ and $\delta_m > 0$ satisfy $\varepsilon_m \downarrow 0$ and $\sum_m \delta_m < \infty$. For each m we can now by (2.1) and (2.5) find a t_m such that

$$P\left[\sup_{[t_m,\tau_H)} (hZ^{(\infty)} - h(t_m)Z^{(\infty)}(t_m))^2 \geq 2\varepsilon_m\right] \leq \delta_m.$$

Thus by the Borel-Cantelli lemma $hZ^{(\infty)}$ converges almost surely; by (2.6) the limit is zero.

In order to prove Theorem 2.1 we first present a sequence of lemmas including a proof of Theorem 1.1.

LEMMA 2.3. *Define* M *by*

$$M(t) = N(t) - \int_0^t Y(s)d\Lambda(s).$$

Then M *is a square integrable martingale on* $[0,\infty]$ *with predictable variation process* $\langle M,M\rangle$ *given by*

$$\langle M,M\rangle(t) = \int_0^t Y(s)d\Lambda(s).$$

PROOF. See AALEN (1976), Section 5C, or GILL (1980a). □

This result can be interpreted as stating that the rate at which $N(t)$ jumps, given the past up to time t, is equal to $Y(t)$ times the hazard rate $d\Lambda(t)/dt$ belonging to F. When F is not continuous, a factor $1-\Delta\Lambda(s)$ must be inserted in the formula for $\langle M,M\rangle$.

LEMMA 2.4. *For all* t

$$\frac{1-\hat{F}(t)}{1-F^T(t)} = 1 - \int_0^t \frac{1-\hat{F}(s-)}{1-F^T(s)} \frac{dM(s)}{Y(s)}.$$

PROOF. See AALEN & JOHANSEN (1978), or GILL (1980a). □

Note that since $M = M^T$ the fact that $Y = 0$ on (T,∞) is of no consequence here. Defining $1/Y = 0$ on (T,∞), the integrand in Lemma 2.4, $(1-\hat{F}_-)/(1-F^T)Y$, is a bounded predictable process on $[0,\tau]$ for any $\tau < \tau_F$. This gives us:

LEMMA 2.5. $(\frac{1-\hat{F}}{1-F})^T$ *and* $Z^T = n^{\frac{1}{2}}(1 - (\frac{1-\hat{F}}{1-F})^T)$ *are square integrable martingales on* $[0,\tau]$ *for any* $\tau < \tau_F$; $\langle Z^T,Z^T\rangle(t) = \int_0^{t\wedge T} \frac{(1-\hat{F}_-)^2}{(1-F)^2} \frac{n}{Y} d\Lambda$ *for all* t.

LEMMA 2.6. *For any* $\beta \in (0,1)$,

$$P[1-\hat{F}(t) \le \beta^{-1}(1-F(t)) \ \forall t \le T] \ge 1-\beta.$$

PROOF. Apply Doob's submartingale version of the Kolmogorov inequality to the nonnegative martingale $(1-\hat{F})/(1-F^T)$ on $[0,\tau]$ for each $\tau < \tau_F$; then take limits as $\tau \uparrow \tau_F$. □

(Lemma 2.6 also holds, taking a little more care at τ_F, when F is noncontinuous.)

The remarkable thing about this result is that when there is no censoring the inequality is actually an equality due to DANIELS (1945). For an elegant proof see PITMAN (1979). Essentially the same idea is also present in VINCZE (1970).

LEMMA 2.7. *For any* $\beta \in (0,1)$,

$$P[Y(t)/n \ge \beta(1-H(t-)) \ \forall t \le T] \ge 1 - e(1/\beta)e^{-1/\beta}.$$

PROOF. See WELLNER (1978) Remark 1(ii). (The fact that H need not be continuous only increases the probability on the left-hand side of this inequality.) □

LEMMA 2.8. *For any* τ *such that* $H(\tau-) < 1$,

$$\sup_{t\le\tau} |\hat{F}(t) - F(t)| \xrightarrow{P} 0,$$

as $n \to \infty$.

PROOF. Since $P[T < \tau] \to 0$ as $n \to \infty$, it suffices by Lemma 2.4 to show that for any $\varepsilon > 0$

$$P\left[\sup_{t\le\tau} \left|\int_0^t \frac{1-\hat{F}(s-)}{1-F^T(s)} \frac{dM(s)}{Y(s)}\right| > \varepsilon\right] \to 0 \quad \text{as } n \to \infty. \tag{2.7}$$

By Lemma 2.3, $n^{-\frac{1}{2}}Z^T = \int \frac{1-\hat{F}_-}{1-F}\frac{dM}{Y}$ is a square integrable martingale on $[0,\tau]$ with predictable variation process equal at time t to

$$\int_0^{t\wedge T} \frac{(1-\hat{F}(s-))^2}{(1-F(s))^2} \frac{d\Lambda(s)}{Y(s)}.$$

Therefore, by LENGLART's (1977) inequality applied to $(n^{-\frac{1}{2}}Z^T)^2$ and $n^{-1}\langle Z^T,Z^T\rangle$, it follows that for any $\eta > 0$, the left-hand side of (2.7) is bounded by

$$\frac{\eta}{\varepsilon^2} + P\left[\int_0^{\tau\wedge T} \frac{(1-\hat{F}(s-))^2}{(1-F(s))^2} \frac{1}{Y(s)}\, d\Lambda(s) > \eta\right]$$

$$\le \frac{\eta}{\varepsilon^2} + P\left[\frac{\Lambda(\tau)}{(1-F(\tau))^2} \frac{1}{Y(\tau\wedge T)} > \eta\right].$$

Since $\frac{Y(\tau\wedge T)}{n} \xrightarrow{P} 1 - H(\tau-) > 0$ as $n \to \infty$, we now easily see that (2.7) holds. □

REMARK 2.9. A strong consistency result can also be simply derived from the representation of $\hat{F} - F$ given in Lemma 2.4; see Lemma 2 and the remarks following Theorem 1 in GILL (1980b). Also if $\tau_H = \tau_F$ it is easy by monotonicity to show that Lemma 2.8 also holds for $\tau = \tau_H$. However, it is an open question as to whether

$$\sup_{t\leq T} |\hat{F}(t) - F(t)| \xrightarrow{P} 0 \quad \text{as } n \to \infty$$

always holds. Note that a corollary of Theorem 1.2, which we could have derived directly, is that we always have

$$\sup_{t\leq T} \frac{1-K(t)}{1-F(t)} |\hat{F}(t) - F(t)| \xrightarrow{P} 0 \quad \text{as } n \to \infty.$$

LEMMA 2.10. *Let* h *be a continuous, nonnegative, and nonincreasing function and let* Z *be a semimartingale, zero at time zero. Then for all* τ

$$\sup_{0\leq t\leq\tau} h(t)|Z(t)| \leq 2 \sup_{0\leq t\leq\tau} \left|\int_0^t h(s)dZ(s)\right|.$$

PROOF. See WELLNER (1977) Lemma 1 for a discrete version of this result. Note that $\int hdZ$ can equivalently be interpreted here as a stochastic integral, a pathwise integral when it exists as such, and by formally integrating by parts; cf. Remark 2.2. Define

$$W(t) = \int_0^t h(s)dZ(s)$$

so that for t such that $h(t) > 0$,

$$Z(t) = \int_0^t \frac{dU(s)}{h(s)} = \frac{U(t)}{h(t)} - \int_0^t U(s-)d(\frac{1}{h(s)})$$

$$= \int_0^t (U(t)-U(s-))d(\frac{1}{h(s)}).$$

Thus

$$|h(t)Z(t)| = \left|\int_0^t (U(t)-U(s-))d(\frac{h(t)}{h(s)})\right|$$

$$\leq 2 \sup_{0\leq s\leq t} |U(s)|(1-\frac{h(t)}{h(0)})$$

giving the required result. □

PROOF OF THEOREM 1.1. Recall that τ satisfies $H(\tau-) < 1$. Since $P[T < \tau] \to 1$ as $n \to \infty$, it suffices to show that

$$Z^T = n^{\frac{1}{2}}\left(\frac{(\hat{F} - F)}{1 - F}\right)^T = \int \frac{1 - \hat{F}_-}{1 - F^T} \frac{n^{\frac{1}{2}} dM}{Y} \xrightarrow{\mathcal{D}} Z^{(\infty)}$$

in $D[0,\tau]$ as $n \to \infty$. By Corollary 2.5 above and Theorem V.1 of REBOLLEDO (1980) this is the case if for each $t < \tau$

$$\langle Z^T, Z^T\rangle(t) = \int_0^{t\wedge T} \frac{(1 - \hat{F}(s-))^2}{(1 - F^T(s))^2} \frac{n}{Y(s)} d\Lambda(s) \xrightarrow{P} C(t)$$

and for each $\varepsilon > 0$

$$\int_0^{\tau\wedge T} \frac{(1 - \hat{F}(s-))^2}{(1 - F^T(s))^2} \frac{n}{Y(s)} I\left\{\frac{1 - \hat{F}(s-)}{1 - F^T(s)} \frac{n^{\frac{1}{2}}}{Y(s)} > \varepsilon\right\} d\Lambda(s) \xrightarrow{P} 0$$

(note that by continuity of F we are in the quasi-left-continuous case in which REBOLLEDO's (1980) strong and weak ARJ(2) conditions coincide). By Lemma 2.8 and the Glivenko-Cantelli theorem for Y/n both conditions are easily seen to hold. □

PROOF OF THEOREM 2.1. By Theorem 1.1 we certainly have weak convergence on $[0,\tau]$ for any τ such that $H(\tau-) < 1$. Also by Remark 2.2 the limiting processes do exist on $[0,\tau_H]$ and are continuous in τ_H. Thus (see BILLINGSLEY (1968) Theorem 4.2) it suffices to prove "tightness at τ_H", i.e. we must show

(2.8) $$\lim_{\tau\uparrow\tau_H} \limsup_{n\to\infty} P[\sup_{\tau\le t\le T} |h(t)Z(t) - h(\tau)Z(\tau)| > \varepsilon] = 0 \qquad \forall\varepsilon > 0$$

and

(2.9) $$\lim_{\tau\uparrow\tau_H} \limsup_{n\to\infty} P\left[\sup_{\tau\le t\le T} \left|\int_\tau^t h(s)dZ(s)\right| > \varepsilon\right] = 0 \qquad \forall\varepsilon > 0.$$

(Note that by the equality $\int Zdh = hZ - \int hdZ$ the corresponding result for $\int Zdh$ does not need to be explicitly verified.)

Now

$$\sup_{\tau \le t \le T} |h(t)Z(t) - h(\tau)Z(\tau)| \le \sup_{\tau \le t \le T} |h(t)(Z(t)-Z(\tau))|$$
$$+ \ (h(\tau)-h(\tau_H))Z(\tau)|.$$

We already know that $Z(\tau) \xrightarrow{\mathcal{D}} Z^{(\infty)}(\tau)$ as $n \to \infty$ so that

$$\limsup_{n\to\infty} P[|(h(\tau_H)-h(\tau))Z(\tau)| > \varepsilon] \le \frac{(h(\tau)-h(\tau_H))^2 C(\tau)}{\varepsilon^2} .$$

If $C(\tau_H) < \infty$ this quantity converges trivially to zero as $\tau \uparrow \tau_H$. However, if $C(\tau_H) = \infty$ we must have $h(\tau_H) = 0$ by (2.1) and convergence to zero follows from (2.5). By Lemma 2.9

$$\sup_{\tau \le t \le T} |h(t)(Z(t)-Z(\tau))| \le 2 \sup_{\tau \le t \le T} \left| \int_\tau^t h(s)dZ(s) \right| .$$

Thus (2.9) implies (2.8).

Since h is, as a process, predictable and bounded, and $Z^T - Z^T(\tau)$ is a square integrable martingale on $[\tau,\tau']$ for each τ' such that $H(\tau'-) < 1$, we have for any $\eta > 0$ by the inequality of Lenglart (1977) (cf. the proof of Lemma 2.8)

$$P\left[\sup_{\tau \le t \le \tau' \wedge T} \left| \int_\tau^t h(s)dZ(s) \right| > \varepsilon \right]$$

$$\le \frac{\eta}{\varepsilon^2} + P\left[\int_\tau^{\tau' \wedge T} \frac{h(s)^2 (1-\hat{F}(s-))^2}{(1-F(s))^2} \frac{n}{Y(s)} d\Lambda(s) > \eta \right]$$

$$\le \frac{\eta}{\varepsilon^2} + \beta + e(1/\beta)e^{-1/\beta} + P\left[\int_\tau^{\tau'} \frac{\beta^{-2} h(s)^2 d\Lambda(s)}{1-H(s-)} > \eta \right]$$

for any $\beta \in (0,1)$ by Lemmas 2.6 and 2.7. Letting $\tau' \uparrow \tau_H$ (or choosing $\tau' = \tau_H$ if $H(\tau_H-) < 1$) and choosing $\eta = \int_\tau^{\tau_H} \beta^{-2} h(s)^2 dC(s)$, we obtain

$$P\left[\sup_{\tau \le t \le T} \left| \int_\tau^t h(s)dZ(s) \right| > \varepsilon \right] \le \beta^{-2}\varepsilon^{-2} \int_\tau^{\tau_H} h(s)^2 dC(s)$$
$$+ \beta + e(1/\beta)e^{-1/\beta}.$$

By (2.1), and since β was arbitrary, this gives us (2.9). □

3. APPLICATIONS

We first prove Theorems 1.2 and 1.3.

PROOF OF THEOREM 1.2. Choose $h = 1 - K$ in Theorem 2.1. Then (2.1) holds because

$$\int_0^{\tau_H} (1 - K(s))^2 dC(s) = \int_0^{\tau_H} \frac{dC(s)}{(1 + C(s))^2} = 1 - \frac{1}{1 + C(\tau_H)} \leq 1 < \infty. \quad \square$$

PROOF OF THEOREM 1.3. Choosing $h = 1 - F$ in Theorem 2.1, we see that

$$n^{\frac{1}{2}}(\hat{F}-F)^T \xrightarrow{\mathcal{D}} (1-F)Z^{(\infty)} \quad \text{in } D[0,\tau_H], \tag{3.1}$$

provided that

$$\int_0^{\tau_H} (1 - F(s))^2 dC(s) = \int_0^{\tau_H} \frac{dF(s)}{1 - G(s-)} < \infty;$$

i.e. provided that (1.1) holds. Now straightforward arguments show that

$$\sup_{0 \leq t \leq \tau} |\hat{K}(t) - K(t)| \xrightarrow{P} 0 \quad \text{as } n \to \infty$$

for any τ such that $H(\tau-) < 1$. Therefore, by Lemma 2.8 we certainly have weak convergence of

$$n^{\frac{1}{2}}\left(\frac{1 - \hat{K}}{1 - \hat{F}}(\hat{F} - F)\right)^T \quad \text{in } D[0,\tau]$$

for any τ such that $H(\tau-) < 1$. Thus to prove Theorem 1.3 it only remains to prove "tightness at τ_H" as in Theorem 2.1. Since $(1-\hat{K})/(1-\hat{F}) \leq 1$ (cf. (1.2)), this follows from (3.1). □

REMARK 3.1. In order to prove Theorem 1.3 *without* the condition (1.1) it would suffice to show that

$$(3.2) \qquad \lim_{\beta\downarrow 0} P\left[\frac{1-\hat{K}(t)}{1-\hat{F}(t)} \Big/ \frac{1-K(t)}{1-F(t)} \le \beta^{-1} \ \forall t \le T\right] = 1$$

uniformly in n. Equivalently, taking the fact that $(1-\hat{F}_-)/(1-\hat{G}_-) = Y/n$ into account and then applying Lemma 2.6 to $1-\hat{G}_-$ and Lemma 2.7 to Y/n, one must show that

$$(3.3) \qquad \lim_{\beta\downarrow 0} P\left[\frac{1+\hat{C}(t)}{1+C(t)} \ge \beta \quad \forall t \le T\right] = 1$$

uniformly in n. However, at present this is an open question.

Next we consider estimation of mean lifetime $\int_0^\infty t dF(t) = \int_0^\infty (1-F(t))dt$ which we suppose here to be finite. Many authors mention this problem but only SUSARLA & VAN RYZIN (1980) achieve any really general result. Even so, they are obliged to work with an estimator

$$\int_0^M (1-\hat{F}(t))dt,$$

where $M = M_n \uparrow \infty$ is a sequence of constants depending on the unknown F and G in a complicated way. We shall consider the estimator $\int_0^T (1-\hat{F}(t))dt$ and obtain a more general result under a natural condition.

Define functions μ and $\bar{\mu}$ and a process $\hat{\mu}$ by

$$\mu(t) = \int_0^t (1-F(s))ds,$$

$$\bar{\mu}(t) = \int_t^\infty (1-F(s))ds,$$

$$\hat{\mu}(t) = \int_0^t (1-\hat{F}(s))ds.$$

Note that

$$\begin{aligned} n^{\frac{1}{2}}(\hat{\mu}-\mu) &= -\int n^{\frac{1}{2}}(\hat{F}-F)dt \\ &= -\int n^{\frac{1}{2}} \frac{\hat{F}-F}{1-F}(1-F)dt \\ &= -\int Z d\bar{\mu}. \end{aligned}$$

Thus we obtain immediately from Theorem 2.1:

THEOREM 3.2. *Suppose*

(3.4) $$\int_0^{\tau_H} \bar{\mu}^{-2} dC < \infty.$$

Then

$$n^{\frac{1}{2}}(\hat{\mu}-\mu)^T \xrightarrow{\mathcal{D}} -\int Z^{(\infty)} d\bar{\mu} \quad \text{in } D[0,\tau_H].$$

COROLLARY 3.3. *Suppose* (3.4) *holds and furthermore*

(3.5) $$n^{\frac{1}{2}}\, \bar{\mu}(T) \xrightarrow{P} 0 \quad \text{as } n \to \infty.$$

Then

$$\begin{aligned} n^{\frac{1}{2}}(\hat{\mu}(T) - \mu(\infty)) &\xrightarrow{\mathcal{D}} -\int_0^{\tau_H} Z^{(\infty)}(s) d\bar{\mu}(s) \\ &= \int_0^{\tau_H} \bar{\mu}(s) dZ^{(\infty)}(s) \\ &\overset{\mathcal{D}}{=} N\left(0, \int_0^{\tau_H} \bar{\mu}(s)^2 dC(s)\right). \end{aligned}$$

PROOF. Under (3.5) we are in the situation $\tau_H = \tau_F$ and $C(\tau_H) = \infty$. Also $T \xrightarrow{P} \tau_H$ as $n \to \infty$. Thus the corollary follows by Remark 2.2. □

REMARK 3.4. Suppose F is a distribution function in the class NBUE (New Better than Used in Expectation), i.e. $\bar{\mu}/(1-F)$ is nonincreasing (in fact we only need that it is bounded). This covers in particular all increasing hazard rate distributions. Then (3.4) is easily seen to hold if (1.1) does, i.e.

$$\int_0^{\tau_H} \frac{dF(t)}{1-G(t-)} < \infty.$$

This is true in particular in the case when

(3.6) $$(1-G) \ge c(1-F)^\beta \quad \text{close to } \tau_F$$

for some constants $c > 0$ and $\beta < 1$; i.e. when the censoring distribution is lighter in the tail than the distribution of interest.

Some straightforward calculations also show that (3.5) holds if for some $0 < \alpha < 2$ and $c' > 0$ we also have

$$(3.7) \qquad (1-H) \geq c'\bar{\mu}^{\alpha} \quad \text{close to } \tau_F.$$

This in turn is implied by (3.6) and F being NBUE (or just $\bar{\mu}/(1-F)$ being bounded).

One would expect to be able to extend Theorem 3.2 to the problem of estimating the residual lifetime function $\bar{\mu}/(1-F)$; we have not attempted this yet however (cf. GHORAI, SUSARLA, SUSARLA & VAN RYZIN (1981)).

In GILL (1980a) we show that the asymptotic variance $\int_0^{\tau_H} \bar{\mu}^2 dC$ may be estimated consistently in the natural way.

Finally, we hope that the conditions of Theorem 3.2 and Corollary 3.3 will discourage actual estimation of mean lifetime in practical applications!

Finally, we sketch an application to q-functions, which appear in the theory of empirical distribution functions (cf. PYKE & SHORACK (1968) Theorem 2.1 and WELLNER (1977)).

THEOREM 3.5. *Suppose* q *is a continuous function on* [0,1] *which is positive on* (0,1), *symmetric about* $\frac{1}{2}$, *nondecreasing on* $[0,\frac{1}{2}]$, *and such that*

$$(3.8) \qquad \int_0^1 \frac{dt}{q(t)^2} < \infty$$

and $(1-t)/q(t)$ *is nonincreasing close to* 1. *Then*

$$n^{\frac{1}{2}}\left(\frac{1}{q(K)}\,\frac{1-K}{1-F}(\hat{F}-F)\right)^T \xrightarrow{\mathcal{D}} \frac{B^0(K)}{q(K)}$$

in $D[0,\tau_H]$ *as* $n \to \infty$.

PROOF. In Theorem 2.1 we only needed h to be nonincreasing in the neighbourhood of τ_H; and this was only needed in the case $C(\tau_H) = \infty$ ($K(\tau_H) = 1$). Thus by (3.8) and Theorem 2.1 with $h = (1-K)/q(K)$ we have weak convergence on $[\tau,\tau_H]$ for any τ such that $K(\tau) > 0$. So we only need further to prove "tightness near zero". Since $(1-K)/q(K)$ is nonincreasing for t such that $K(t) \le \frac{1}{2}$, exactly the same arguments as in Theorem 2.1 can be used again to give the required result. □

ACKNOWLEDGEMENT. *I am very grateful to Jon Wellner for the stimulating discussions which led to this paper being written.*

REFERENCES

[1] AALEN, O.O. (1976), *Statistical inference for a family of counting processes*, Institute of Mathematical Statistics, University of Copenhagen.

[2] AALEN, O.O. (1978), *Nonparametric inference for a family of counting processes*, Ann. Statist. 6, 701-726.

[3] AALEN, O.O. & S. JOHANSEN (1978), *An empirical transition matrix for nonhomogeneous Markov chains based on censored observations*, Scand. J. Statist. 5, 141-150.

[4] BILLINGSLEY, P. (1968), *Weak convergence of probability measures*, Wiley, New York.

[5] BIRNBAUM, Z.W. & A.W. MARSHALL (1961), *Some multivariate Chebyshev inequalities with extensions to continuous parameter processes*, Ann. Math. Statist. 32, 687-703.

[6] BRESLOW, N. & J. CROWLEY (1974), *A large sample study of the life table and product-limit estimates under random censorship*, Ann. Statist. 2, 437-453.

[7] BURKE, M.D., CSÖRGŐ, S. & L. HORVÁTH (1981), *Strong approximations of some biometric estimates under random censorship*, Zeit. Wahrsch. verw. Geb. 56, 87-112.

[8] CSÖRGŐ, S. & L. HORVÁTH (1981), *On cumulative hazard processes under random censorship*, to appear in Scand. J. Statist.

[9] DANIELS, H.E. (1945), *The statistical theory of the strength of bundles of threads I,* Proc. Roy. Soc. A 183, 405-435.

[10] DOOB, J.L. (1949), *A heuristic approach to the Kolmogorov-Smirnov theorems,* Ann. Math. Statist. 20, 393-403.

[11] EFRON, B. (1967), *The two sample problem with censored data,* Proc. 5th Berkeley Symp. 4, 831-853.

[12] GILL, R.D. (1980a), *Censoring and stochastic integrals*, Mathematical Centre Tracts 124, Mathematisch Centrum, Amsterdam.

[13] GILL, R.D. (1980b), *Nonparametric estimation based on censored observations of a Markov renewal process*, Zeit. Wahrsch. verw. Geb. 53, 97-116.

[14] GILLESPIE, M.J. & L. FISHER (1979), *Confidence bands for the Kaplan-Meier survival curve estimate,* Ann. Statist. 7, 920-924.

[15] GHORAI, J., SUSARLA, A., SUSARLA, V. & J. VAN RYZIN (1981), *Nonparametric estimation of mean residual life-time with censored data,* Preprint, University of Wisconsin-Milwaukee.

[16] HALL, W.J. & J.A. WELLNER (1980), *Confidence bands for a survival curve from censored data*, Biometrika 67, 133-143.

[17] JOHANSEN, S. (1978), *The product-limit estimator as maximum likelihood estimator,* Scand. J. Statist. 5, 195-199.

[18] KAPLAN, E.L. & P. MEIER (1958), *Nonparametric estimation from incomplete observations,* J. Amer. Statist. Assoc. 53, 457-481.

[19] LENGLART, E. (1977), *Relation de domination entre deux processus*, Ann. Inst. Henri Poincaré 13, 171-179.

[20] MEYER, P.A. (1978), *Un cours sur les intégrales stochastiques,* Séminaire de Probabilités VI, Lecture Notes in Mathematics 511, 245-400, Springer-Verlag, Berlin.

[21] PITMAN, E.J.G. (1979), *Some basic theory for statistical inference*, Chapman and Hall, London.

[22] PYKE, R. & G.R. SHORACK (1968), *Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems*, Ann. Math. Statist. 39, 755-771.

[23] REBOLLEDO, R. (1980), *Central limit theorems for local martingales*, Zeit. Wahrsch. verw. Geb. 51, 269-286.

[24] SCHOLZ, F.W. (1980), *Towards a unified definition of maximum likelihood*, Can. J. Statist. 8, 193-203.

[25] SUSARLA, V. & J. VAN RYZIN (1980), *Large sample theory for an estimator of the mean survival time from censored samples*, Ann. Statist. 8, 1002-1016.

[26] VINCZE, I. (1970), *On Kolmogorov-Smirnov type distribution theorems*, p. 385-401 in: Nonparametric techniques in statistical inference, M.L. Puri (ed.), Cambridge University Press.

[27] WELLNER, J.A. (1977), *A martingale inequality for the empirical process*, Ann. Probab. 5, 303-308.

[28] WELLNER, J.A. (1978), *Limit theory for the ratio of the empirical distribution function to the true distribution function*, Zeit. Wahrsch. verw. Geb. 45, 73-88.

[29] WELLNER, J.A. (1981), *Estimating a distribution function with random censorship*, Preprint no. 13, Mathematical Institute, University of Munich.