Acta Sci. Math., 44 (1982), 23—34

On random censorship from the right

SÁNDOR CSÖRGŐ and LAJOS HORVÁTH

1. Introduction. BURKE *et al.* [4] introduced the following censorship model. Let X be a real random variable with distribution function $F(t)=\mathrm{pr}\,\{X<t\}$. For a fixed integer $k\geqq 1$ let $A^1, \ldots, A^k$ be pairwise disjoint random events, and define the sub-distribution function $F^i(t)=\mathrm{pr}\,\{X<t, A^i\}, i=1, \ldots, k$. We are interested in the joint behaviour of the pairs (X, A^i) as expressed by

$$S^i(t) = \exp\left(-\Lambda^i(t)\right), \quad i = 1, \ldots, k,$$

where Λ^i is the i-th hazard function $\left(\int\limits_{-\infty}^{t} = \int\limits_{(-\infty,t)}\right)$

$$\Lambda^i(t)= \int\limits_{-\infty}^{t} (1-F(s))^{-1}\, dF^i(s).$$

So let $\{X_n, A_n^1, \ldots, A_n^k\}$ be a sequence of independent replicas of $\{X, A^1, \ldots, A^k\}$, $n=1,2, \ldots$. *We assume throughout that the functions* $F, F^1, \ldots, F^k$ *are continuous.* Define the product-limit estimates

$$\tilde{S}_n^i(t) = 1-\tilde{F}_n^i(t) = \begin{cases} \prod\limits_{\{1\leqq j\leqq n:\, X_j<t\}} \left(\dfrac{n-R_j^i}{n-R_j^i+1}\right)^{\delta_j^i}, & \text{if} \quad t < \max(X_1, \ldots, X_n) \\ 0, & \text{otherwise,} \end{cases}$$

$i=1, \ldots, k$, where δ_j^i is the indicator of A_j^i, and R_j^i is the rank of $(X_j, 1-\delta_j^i)$ in the lexicographic ordering of the sequence $(X_1, 1-\delta_1^i), \ldots, (X_n, 1-\delta_n^i)$. Finally, introduce the i-th product-limit process

$$Z_n^i(t) = n^{1/2}\left(S^i(t)-\tilde{S}_n^i(t)\right),$$

and, for $x=(x_1, \ldots, x_k)$, the corresponding vector process

$$Z_n(x) = \left(Z_n^1(x_1), \ldots, Z_n^k(x_k)\right).$$

Received July 1, 1980.

However general is this model, the most important special cases are a) and b) below. By working in this generality we merely would like to emphasize the fact that the asymptotic theory of random censorship on the right requires only the above structure. When dealing with random censorship on the left the basic ingredients S^i and $\tilde{S}_n^i$ should be accordingly modified. This is done in [9].

a) Let $X_1^0, X_2^0, \ldots$ be a sequence of independent random variables with common continuous distribution function F^0. These are censored on the right by $Y_1, Y_2, \ldots$ a sequence of independent random variables, independent of the X^0 sequence, with common continuous distribution function H. One can only observe the sequence $(X_n = \min(X_n^0, Y_n), \delta_n)$, where $\delta_n = \delta_n^1$ is the indicator of $A_n = A_n^1 = \{X_n = X_n^0\}$. In this case $k=2$, $1-F=(1-F^0)(1-H)$, $F^1(t) = \int_{-\infty}^{t} (1-H)\,dF^0$, thus $S^1(t) = S(t) = 1-F^0(t)$, and $\tilde{S}_n^1 = \tilde{S}_n$ reduces to the usual product-limit estimate. This is the Kaplan—Meier [15] model as defined by Efron [12]. It was investigated by Breslow and Crowley [3], Meier [19], Hall and Wellner [14], Burke *et al.* [4] and others. The useful special case when $1-H=(1-F^0)^\beta$, $\beta>0$, was considered by Koziol and Green [18], and their model was investigated by Csörgő and Horváth [7] and Koziol [17].

b) For $k>1$ consider k independent sequences $Y_1^i, Y_2^i, \ldots$ $(i=1, \ldots, k)$ of independent random variables with common continuous distribution function H^i, and let $X_n = \min(Y_n^1, \ldots, Y_n^k)$. One observes the sequences (X_n, δ_n^i), $i=1, \ldots, k$, where δ_n^i is the indicator of the event $A_n^i = \{X_n = Y_n^i\}$. This is the competing risks model (giving back the above Kaplan—Meier model for $k=2$) considered by many authors, notably, from the present viewpoint, by Yang [22] and Burke *et al.* [4]. Here, as Berman [2] proved, the above S^i reduces to $S^i(t) = 1-H^i(t)$.

On the basis of the Efron-transformed variant of the Breslow—Crowley weak convergence theorem, Gillespie and Fisher [13] constructed asymptotic confidence bands for the survival curve $1-F^0$ in the Kaplan—Meier model. However, their Monte Carlo study has shown that sample size $n=200$ is not large enough to apply the asymptotic bands with high precision. Their results were a strong motivation for us to work out a strong approximation theory in [4] for the above general Z_n and related processes. A variant of one of the main approximation theorems is formulated in the next section. This result enabled us to build the approximation rates into the construction of the Gillespie—Fisher type bands, i.e., we could construct "exact" confidence bands ([4]) for the general survival functions S^i under the "i-th risk A^i". We also indicated that these constructions should give reasonable bands for much less sample sizes than the asymptotic ones of Gillespie and Fisher.

Hall and Wellner [14] utilized Doob's transformation of the Brownian motion into the Brownian bridge, and hence proposed the corresponding transformation of the product-limit process in the Kaplan—Meier model. The resulting process converges weakly to a transformed Brownian bridge. Although Doob's transformation belongs to the statistical folklore nowadays, its use by Hall and Wellner in the present context is a remarkable step in the asymptotic theory of censored empirical processes. The resulting asymptotic confidence bands constructed by HALL and WELLNER [14] enjoy many advantages over those of GILLESPIE and FISHER [13] as they explain it in detail. For example, they reduce in the uncensored case to the classical Kolmogorov bands. Following HALL and WELLNER [14], KOZIOL [17] considered Kolmogorov, Kuiper and Cramér—von Mises statistics corresponding to the transformed product-limit process in the Kaplan—Meier model for testing goodness of fit (cf. Section 3 here).

Following AALEN [1], NAIR [20] proposed another clever transformation of the product-limit process in the Kaplan—Meier model. It is a modification of Efron's transformation, where the limit process is a scale-changed Wiener process. The rescaling depends on censoring, but the Kolmogorov—Smirnov, Kuiper and Cramér—von Mises functionals of this process are distribution-free.

The aim of the present note is to develop strong approximation theorems corresponding to the transformations of Hall and Wellner and of Aalen and Nair in the general setting of the first paragraph of this section. This is done in Sections 3 and 4, respectively, after some preliminaries from Burke *et al.* Convergence rates are deduced from these theorems for the above mentioned statistics in Sections 3 and 4. Using the approximation rates, we show in Section 5 a possibility for making exact the asymptotic bands of Hall and Wellner. This is done again in the general setting. Comulative hazard processes are investigated in a similar manner by CSÖRGŐ and HORVÁTH [8].

2. Preliminaries. Let $T_F = \inf\{t : F(t)=1\}$ and define

$$d^i(t) = \begin{cases} \int\limits_{-\infty}^{t} (1-F(s))^{-2}\, dF^i(s), & t < T_F \\ \int\limits_{-\infty}^{T_F} (1-F(s))^{-2}\, dF^i(s), & t \geqq T_F, \end{cases} \tag{2.1}$$

$i=1, \ldots, k$. The last integral $\int\limits_{-\infty}^{T_F}$ here is either finite or infinite. Consider

$$\tilde{Z}_n^i(t) = \exp(\Lambda^i(t)) Z_n^i(t),$$

$i=1, \ldots, k$, and for $x=(x_1, \ldots, x_k)$ the corresponding vector process

$$\tilde{Z}_n(x) = (\tilde{Z}_n^1(x_1), \ldots, \tilde{Z}_n^k(x_k)).$$

If a^i denotes the inverse of d^i, then the vector process with components $\tilde{Z}_n^i(a^i(x_i))$,

$x \in (0, \infty)^k$, is the one which was called by BURKE *et al.* [4] as the Efron transform of Z_n. All of our approximations will take place on the infinite cube $(-\infty, T_n]^k$ where T_n is a sequence of numbers satisfying first the condition:

$$T_n < T_F \quad \text{and} \quad 1-F(T_n) \geqq (2\varepsilon n^{-1} \log n)^{1/2} \tag{2.2}$$

where, throughout this note, ε is some arbitrarily fixed positive number. Let

$$b_n = \big(1-F(T_n)\big)^{-1}, \tag{2.3}$$

and introduce the following (rather messy) rate-sequence

$$r(n) = v(n) + \frac{1}{2} n^{-1/2}\{v(n)+3(\varepsilon/2)^{1/2} b_n^2(\log n)^{1/2}\}^2 \tag{2.4}$$

where

$$v(n) = \big[b_n^2\{(2k+1)5A_1+\big(2+(5(2k+1)/A_3)\big)\varepsilon+\big((2/3)\varepsilon+\varepsilon^2\big)^{1/2}\}+b_n^3 4\varepsilon\big] n^{-1/2}\log n +$$
$$+ b_n^2(12\varepsilon)^{1/2} n^{-1/3}(\log n)^{1/2} + b_n^2(2k+1)\{A_1+(\varepsilon/A_3)\}\varepsilon^{1/2} n^{-1/3}(\log n)^{3/2} + b_n 2n^{-1/2}.$$

For $x=(x_1, \ldots, x_k)$ let $\|x\| = \max(|x_1|, \ldots, |x_k|)$ denote the maximum norm. In [4] we constructed a special probability space $(\Omega, \mathcal{A}, P)$ carrying k sequences $\{W_n^i\}$ of Wiener processes such that for the vector process

$$W_n^d(x) = \big(W_n^1(d^1(x_1)), \ldots, W_n^k(d^k(x_k))\big)$$

we have

Theorem A (BURKE, CSÖRGŐ, HORVÁTH [4]). *If T_n satisfies condition (2.2), then*

$$P\{\sup_{x\in(-\infty, T_n]^k} \|\tilde{Z}_n(x) - W_n^d(x)\| > r(n)b_n\} \leqq kQn^{-\varepsilon},$$

where $Q = 10A_2(2k+1)+100+16D$.

The constants A_1, A_2 and A_3 in $r(n)$ (A_2 also in Q) are the C, K, λ constants of Theorem 3 of KOMLÓS, MAJOR and TUSNÁDY [16] (quoted as Theorem 2. A in [4]), respectively. According to TUSNÁDY [21] (cf. also M. CSÖRGŐ and P. RÉVÉSZ [5]) A_1, A_2 and A_3 can respectively be taken as 100, 10 and 1/50. Constant D (in Q) is the absolute constant of Lemma 2 of DVORETZKY, KIEFER and WOLFOWITZ [11]. The smallest available value for D known to us at present is $2\{1+32/(6\pi)^{1/2} + 8/3^{1/2} + 2^{1/2} 4 \exp(71/18)\} \leqq 611$ due to DEVROYE and WISE [10]. But in practice one can probably use without harm the well-known conjecture (which was empirically verified in a number of situations) that D is 2.

We should also point out here that originally (Theorem 5.6 of [4]) we had the factor $r_E(n) = \max_{1\leqq i\leqq k} \exp\big(\Lambda^i(T_n)\big)$ of $kr(n)$ instead of b_n. But it is not hard to see that $r_E(n) \leqq b_n$ and the above form of Theorem A is more fortunate since the whole rate-sequence $r(n)b_n$ depends on the censoring only through b_n of (2.3).

3. Approximation theorems for the Hall—Wellner transformation. Goodness of fit. Introduce (with d^i of (2.1))

$$K^i(t) = d^i(t)/(1+d^i(t)), \quad -\infty < t < \infty, \quad i = 1, \ldots, k.$$

$K^i(t)$ is a sub-distribution function in general for each i. It is a distribution function (as HALL and WELLNER [14] point out) in the Kaplan—Meiner model ($k=2$). In the competing risks model K^i is a distribution function for those i for which $T_F = T_{H^i} \leqq \min(T_{H^1}, \ldots, T_{H^k})$, where T_{H^i} is defined analogously to T_F. The empirical counterpart of $d^i(t)$ was considered by BURKE *et al.* [4] as

$$d_n^i(t) = \int_{-\infty}^{t} (1-F_n(s))^{-2}\, dF_n^i(s), \quad i=1, \ldots, k,$$

where F_n is the (left continuous) empirical distribution function of $X_1, \ldots, X_n$ and F_n^i is the empirical sub-distribution function defined as

$$F_n^i(t) = n^{-1} \# \{m\colon 1 \leqq m \leqq n, X_m < t \text{ and } A_m^i \text{ occurs}\}, \quad i = 1, \ldots, k.$$

Independently of us but earlier, HALL and WELLNER [14] have also considered d_n^i (in the Kaplan—Meier model) but pointed out that it fails to satisfy their reduction property. Instead they proposed the following modification of it:

$$c_n^i(t) = \int_{-\infty}^{t} (1-F_n(s))^{-1}(1-F_n^+(s))^{-1}\, dF_n^i(s) = n \sum_{\{j\colon X_j<t\}} (n-j)^{-1}(n-j+1)^{-1}\delta_j^i$$

where F_n^+ is the right-continuous version of F_n. Although we could have worked with d_n^i, we adopted this modification for the sake of accordance.

Lemma 6.2 of [4] estimates the distance of d_n^i and d^i. Using Lemma 4.1 of that paper, it is trivial that if T_n satisfies condition (2.2), then

$$\mathrm{pr}\Big\{\sup_{-\infty<t\leqq T_n} |d_n^i(t)-c_n^i(t)| > 8b_n^3 n^{-1}\Big\} \leqq 2Dn^{-\varepsilon}.$$

Let

$$K_n^i(t) = c_n^i(t)/(1+c_n^i(t)), \quad i = 1, \ldots, k.$$

Evidently

$$|K_n^i(t)-K^i(t)| \leqq |c_n^i(t)-d^i(t)|,$$

and hence, putting together Lemma 6.2 of [4] and the last probability inequality, we obtain

Lemma 3.1. *If T_n satisfies (2.2), then for each $i=1, \ldots, k$*

$$\mathrm{pr}\Big\{\sup_{-\infty<t\leqq T_n} |K_n^i(t)-K^i(t)| > r_1(n)\Big\} \leqq 8Dn^{-\varepsilon},$$

where $r_1(n) = 12(\varepsilon/2)^{1/2} n^{-1/2} b_n^4 (\log n)^{1/2} + 8b_n^3 n^{-1}$.

Consider now

$$\hat{Z}_n^i(t) = \big(1-K_n^i(t)\big)\exp\big(\Lambda^i(t)\big)Z_n^i(t), \quad i = 1, \dots, k,$$

and let $B_n^i(t)=(1-t)W_n^i\big(t/(1-t)\big)$ be the sequence of Brownian bridges supplied by $\{W_n^i\}$ of Theorem A. For $x=(x_1, \dots, x_k)$ let

$$\hat{Z}_n(x) = \big(\hat{Z}_n^1(x_1), \dots, \hat{Z}_n^k(x_k)\big)$$

and

$$B_n^K(x) = \big(B_n^1(K^1(x_1)), \dots, B_n^k(K^k(x_k))\big).$$

Theorem 3.2. *If T_n satisfies (2.2), then*

$$P\Big\{\sup_{x\in(-\infty,T_n]^k} \|\hat{Z}_n(x)-B_n^K(x)\| > q_1(n)\Big\} \leqq kR_1 n^{-\varepsilon},$$

where $q_1(n)=r(n)b_n+r_2(n)$ *with* $r_2(n)=2\varepsilon^{1/2}r_1(n)\,b_n(\log n)^{1/2}$, *and* $R_1=Q+8D+2=$ $=10A_2(2k+1)+102+24D$.

Proof. It is enough to show that

$$P\big\{\sup\big|\hat{Z}_n^i(t)-B_n^i(K^i(t))\big| > q_1(n)\big\} \leqq R_1 n^{-\varepsilon}$$

for each $i=1, \dots, k$, where unspecified sup means $\sup\limits_{-\infty<t\leqq T_n}$. The last probability is not greater than

$$\begin{aligned}
&P\big\{\sup\big(1-K_n^i(t)\big)\big|\tilde{Z}_n^i(t)-W_n^i(d(t))\big| > r(n)\,b_n\big\}+\\
&+P\big\{\sup|K_n^i(t)-K^i(t)|\,\big|W_n^i(d^i(t))\big| > r_2(n)\big\} \leqq\\
&\leqq (Q+8D)n^{-\varepsilon}+P\big\{\sup\big|W_n^i(d^i(t))\big| > 2\varepsilon^{1/2}b_n(\log n)^{1/2}\big\} \leqq\\
&\leqq (Q+8D)n^{-\varepsilon}+2P\big\{|W_n^i(b_n^2)/b_n| > 2\varepsilon^{1/2}(\log n)^{1/2}\big\}
\end{aligned}$$

by Theorem A, Lemma 3.1, and the fact that $b_n^2\geqq d^i(T_n)$. The last probability is less than or equal to $n^{-\varepsilon}$, and hence the theorem.

The components $\hat{Z}_n^i$ of our vector (-vector) process are in fact weighted processes, the weight being $\exp\big(\Lambda^i(t)\big)$. It is then natural to replace this weight with an empirical counterpart of it and investigate the convergence of the resulting "twice estimated" product-limit process. In principle there are two empirical candidates for doing this. One is the exponential empirical hazard function $\exp\big(\Lambda_n^i(t)\big)$ (cf.[4]) and the other is the product limit estimate itself. The latter being more natural here, consider

$$\hat{\hat{Z}}_n^i(t) = \big(1-K_n^i(t)\big)Z_n^i(t)/\tilde{S}_n^i(t) = \big(1-K_n^i(t)\big)\big(\exp(-\Lambda^i(t))-\tilde{S}_n^i(t)\big)/\tilde{S}_n^i(t)$$

$i=1, \dots, k$, and the corresponding vector process

$$\hat{\hat{Z}}_n(x) = \big(\hat{\hat{Z}}_n^1(x_1), \dots, \hat{\hat{Z}}_n^k(x_k)\big)$$

for $x=(x_1, \dots, x_k)$.

For T_n we introduce a slightly stronger regularity condition instead of (2.2):

(3.1) $$T_n < T_F \quad \text{and} \quad 1-F(T_n) \geqq 2n^{-1/2} r_3(n),$$

where

$$r_3(n) = r(n)+3(\varepsilon/2)^{1/2} b_n^2 (\log n)^{1/2}.$$

By definition (2.4) of $r(n)$ it can be shown that a rough sufficient condition for condition (3.1) to be satisfied is that ($T_n \nearrow T_F$ so slowly that)

(3.2) $$b_n = \left(1-F(T_n)\right)^{-1} \leqq M_\varepsilon (n/\log n)^{1/6}$$

with some constant M_ε depending only on ε, which can be computed from $r_3(n)$.

Just as Lemma 5.1 of [4] was deduced form an approximation theorem, the first statement of the next lemma easily results from Theorem 5.5 of [4] which is the original Breslow—Crowley-type variant of the Efron-type theorem cited here. When deducing it, one also should apply the already mentioned fact that $\exp\left(-\Lambda^i(T_n)\right) \geqq 1-F(T_n)$. The second statement of the lemma follows from the first just as Lemma 4.1 of [4] followed from the Dvoretzky—Kiefer—Wolfowitz bound.

Lemma 3.3. *If T_n satisfies (3.1), then*

$$\mathrm{pr}\left\{\sup_{-\infty<t\leqq T_n} |Z_n^i(t)| > r_3(n)\right\} \leqq (Q+6)\, n^{-\varepsilon}$$

and

$$\mathrm{pr}\left\{\sup_{-\infty<t\leqq T_n} \frac{1}{\tilde{S}_n^i(t)} > \frac{2}{\exp\left(-\Lambda^i(T_n)\right)}\right\} \leqq (Q+6) n^{-\varepsilon}.$$

Theorem 3.4. *If T_n satisfies (2.2), then*

$$P\left\{\sup_{x\in(-\infty,\, T_n]^k} \|\hat{\hat{Z}}_n(x)-B_n^K(x)\| > q_2(n)\right\} \leqq kR_2 n^{-\varepsilon},$$

where $q_2(n) = q_1(n)+2n^{-1/2} b_n^2 \left(r_3(n)\right)^2$ *and* $R_2 = R_1+2Q+12 = 3Q+14+8D = 30A_2(2k+1)+314+56D.$

Proof.

$$\left|\hat{\hat{Z}}_n^i(t)-B_n^i(K^i(t))\right| \leqq \left|\hat{Z}_n^i(t)-B_n^i(K^i(t))\right| + |\hat{\hat{Z}}_n^i(t)-\hat{Z}_n^i(t)| \leqq$$
$$\leqq \left|\hat{Z}_n^i(t)-B_n^i(K^i(t))\right| + n^{-1/2}|Z_n^i(t)|^2/\{\tilde{S}_n^i(t)\exp\left(-\Lambda^i(t)\right)\},$$

and the theorem follows from Theorem 3.2 and Lemma 3.3.

As to the order of our rate sequences $q_1(n)$ and $q_2(n)$, we note that since

$$r(n) = O\left(\max\{b_n^2 n^{-1/3}(\log n)^{3/2},\ b_n^4 n^{-1/2}\log n\}\right),$$

we have

$$q_1(n) = O\left(\max\{b_n^3 n^{-1/3}(\log n)^{3/2},\ b_n^5 n^{-1/2}\log n\}\right),$$
$$q_2(n) = O\left(\max\{b_n^3 n^{-1/3}(\log n)^{3/2},\ b_n^6 n^{-1/2}\log n\}\right).$$

Now we formulate the corresponding consequences for approximation on the fixed cube $(-\infty, T]^k$ with $T<T_F$. These consequences follow from Theorems 3.2 and 3.4 in the same way as Corollary 5.7 of [4] did. Note that $q_1(n)$, $q_2(n)$ and $r_3(n)$ are understood from now on with b_n replaced in them by the constant

$$b = (1-F(T))^{-1}.$$

Corollary 3.5. *If* $n/\log n \geqq 2\varepsilon b^2$, *then*

$$P\{\sup_{x\in(-\infty, T]^k} \|\hat{Z}_n(x)-B_n^K(x)\| > q_1(n)\} \leqq kR_1 n^{-\varepsilon},$$

and if $n^{1/2}/r_3(n) \geqq 2b$, *then*

$$P\{\sup_{x\in(-\infty, T]^k} \|\hat{\hat{Z}}_n(x)-B_n^K(x)\| > q_2(n)\} \leqq kR_2 n^{-\varepsilon}.$$

The rough sufficient condition for the second statement here is (3.2) with b in place of b_n.

The joint weak convergence of the components of $\hat{Z}_n$ and $\hat{\hat{Z}}_n$ follows from this corollary together with rate-of-convergence results. Namely, for many functionals ψ (cf. Corollary of Komlós *et al.* [16] and Corollary 1 of Csörgő [6]) on the space of functions defined on $(-\infty, T]^k$ we obtain

$$\text{(3.2)} \qquad \sup_{-\infty<y<\infty} |\mathrm{pr}\{\psi(\overset{*}{Z}_n(\cdot)) < y\} - \mathrm{pr}\{\psi(B^K(\cdot)) < y\}| = O(n^{-1/3}(\log n)^{3/2}),$$

where $*=\hat{}, \hat{\hat{}}$ and B^K is a copy of B_n^K since, if T $(<T_F)$ is fixed, then

$$q_1(n) = O(n^{-1/3}(\log n)^{3/2}) = q_2(n).$$

For example, (3.2) holds for the Kolmogorov, Smirnov and Kuiper statistics considered by Koziol [17].

4. An approximation theorem for the Aalen—Nair transformation. Goodness of fit. Let T be a number such that the inequalities

$$\text{(4.1)} \qquad T < T_F, \quad F^i(T) > 0, \quad i = 1, \ldots, k,$$

hold, and consider the processes

$$\check{Z}_n^i(t) = \tilde{Z}_n^i(t)/(d_n^i(T))^{1/2}, \quad i = 1, \ldots, k,$$

proposed by Aalen [1] and Nair [20] in the Kaplan—Meier case $(k=2)$ where $\tilde{Z}_n^i$ and d_n^i are of Sections 2 and 3 respectively. Also, with W_n^i of Theorem A, introduce

$$W_n^{(i)}(t) = W_n^i(d^i(t))/(d^i(T))^{1/2}, \quad i = 1, \ldots, k,$$

and for $x=(x_1, \ldots, x_k)$ set

$$\check{Z}_n(x) = (\check{Z}_n^1(x_1), \ldots, \check{Z}_k^n(x_k)), \quad \check{W}_n(x) = (W_n^{(1)}(x_1), \ldots, W_n^{(k)}(x_k)).$$

We note that, $i=1, \ldots, k$,

$$\{W_n^{(i)}(t):\ -\infty < t \leqq T\} =_{\mathcal{D}} \{W(d^i(t)/d^i(T)):\ -\infty < t \leqq T\} \tag{4.2}$$

and this equality in distribution is in fact the main advantage of the Aalen—Nair transformation. Introduce the notation (in addition to those of the preceding sections)

$$a = \max_{1 \leqq i \leqq k} 1/F^i(T).$$

Theorem 4.1. *If* $n/\log n \geqq \max(2\varepsilon b^2, 8\varepsilon a^2)$, *then*

$$P\{\sup_{x \in (-\infty, T]^k} \|\check{Z}_n(x) - \check{W}_n(x)\| > q_3(n)\} \leqq kR_3 n^{-\varepsilon},$$

where $q_3(n) = b(2a)^{1/2} r(n) + 12(2)^{1/2} \varepsilon b^5 a^{3/2} n^{-1/2} \log n$ *and* $R_3 = Q + 9D + 1 = 10A_2(2k+1) + 25D + 101$.

Proof.

$$P\{\sup_{-\infty < t \leqq T} |\check{Z}_n^i(t) - W_n^{(i)}(t)| > q_3(n)\} \leqq$$

$$\leqq P\{\sup_{-\infty < t \leqq T} (d_n^i(T))^{-1/2} |\tilde{Z}_n^i(t) - W_n^i(d^i(t))| > b(2a)^{1/2} r(n)\} +$$

$$+ P\{\sup_{-\infty < t \leqq T} |W_n^i(d^i(t))| > 2\varepsilon^{1/2} b(\log n)^{1/2}\} + P\{(d^i(T)\, d_n^i(T))^{-1/2} > (2a^2)^{1/2}\} +$$

$$+ P\{|(d^i(T))^{1/2} - (d_n^i(T))^{1/2}| > 12(\varepsilon/2)^{1/2} b^4 (a/2)^{1/2} n^{-1/2} (\log n)^{1/2}\}.$$

Since $d_n^i(T) \geqq F_n^i(T)$ and by an obvious analogue of Lemma 4.1 of [4]

$$P\left\{\frac{1}{F_n^i(T)} \geqq \frac{2}{F^i(T)}\right\} \leqq Dn^{-\varepsilon}, \tag{4.3}$$

provided that $n/\log n \geqq 8\varepsilon a^2$, we obtain, using also Theorem A, that the first term of the above sum is not greater than $(Q+D)n^{-\varepsilon}$. We saw in the proof of Theorem 3.2 that the second term is not greater than $n^{-\varepsilon}$. By (4.3) the third term is majorized by $Dn^{-\varepsilon}$. Using again (4.3), the fourth probability is majorized by

$$Dn^{-\varepsilon} + P\{|d^i(T) - d_n^i(T)| > 12(\varepsilon/2)^{1/2} b^4 n^{-1/2} (\log n)^{1/2}\} \leqq 7Dn^{-\varepsilon},$$

where we used Lemma 6.2 of [4] in the last step. This proves the theorem.

By (4.2) the limit distributions of the Kolmogorov, Smirnov and Kuiper statistics based on the processes $\check{Z}_n^i$ coincide with the distributions of the corresponding functionals of $\{W(s) : 0 \leqq s \leqq 1\}$. These distributions are well known, one of them is tabulated in [7]. If $\psi(\check{Z}_n^i)$ denotes any of these three statistics and $\psi(W)$ denotes its distribution-free limiting random variable, then we have (3.2) for their distribution functions by Theorem 4.1.

Since the Aalen—Nair modified Efron transformation leads to asymptotically distribution-free statistics, this transformation is more advantageous than those of

Hall and Wellner when testing goodness of fit. However, the latter seems much better when constructing confidence bands. This is why we do not spell out the exact probability inequalities in the next section corresponding to the confidence bands arising from the transformation of Aalen and Nair.

The two-sample processes, or, more generally, their vector-process generalization (for the general competing risks model) can be similarly approximated as the one-sample processes in Theorems 3.2, 3.4 and 4.1.

5. Conficence bands. If G is a continuous distribution function and G_n is the n-stage empirical distribution function of a sample corresponding to G, then it follows from Theorem 3 of Komlós *et al.* (1975) that for any $\lambda, \varepsilon>0$ we have

$$-A_2 n^{-\varepsilon}+M\big(\lambda-(A_1+(\varepsilon/A_3))n^{-1/2}\log n\big) \leqq$$
$$\leqq \mathrm{pr}\,\{G_n(t)-\lambda/n^{1/2} \leqq G(t) \leqq G_n(t)+\lambda/n^{1/2},\ -\infty<t<\infty\} \leqq$$
$$\leqq M\big(\lambda+(A_1+(\varepsilon/A_3))n^{-1/2}\log n\big)+A_2 n^{-\varepsilon},$$

where

$$M(y) = \mathrm{pr}\big\{\sup_{0\leqq s\leqq 1} |B(s)| < y\big\}.$$

As we have already noted, A_1, A_2 and A_3 can be taken by Tusnády [21] as 100, 10 and 1/50, respectively. (It would be interesting to search for smaller A_1, A_2 and larger A_3 by Monte Carlo through the above inequalities.) By Remark 1 of [6] and the fact that $\sup_{-\infty<y<\infty} \big|B(G(y))\big| \leqq \sup_{0\leqq s\leqq 1} |B(s)|$, the lower half of the above inequality remains valid for discontinuous G as well.

For $0<a<1$ set

$$M_a(y) = \mathrm{pr}\big\{\sup_{0\leqq s\leqq a} |B(s)| < y\big\}.$$

The analogues of the above inequalities for the general right censorship model are the following consequences of Corollary 3.5.

Corollary 5.1. *Let $T<T_F$. If $n/\log n \geqq 2\varepsilon b^2$, then for any $\lambda>0$ and $i=1, \dots, k$ we have*

$$-R_1 n^{-\varepsilon}+M_{K^i(T)}\big(\lambda-q_1(n)\big) \leqq$$
$$\leqq \mathrm{pr}\left\{\frac{\tilde{S}_n^i(t)}{1+\dfrac{\lambda}{n^{1/2}(1-K_n^i(t))}} \leqq S^i(t) \leqq \frac{\tilde{S}_n^i(t)}{1-\dfrac{\lambda}{n^{1/2}(1-K_n(t))}},\ -\infty<t\leqq T\right\} \leqq$$
$$\leqq M_{K^i(T)}\big(\lambda+q_1(n)\big)+R_1 n^{-\varepsilon}.$$

If $n^{1/2}/r_3(n) \geqq 2b$, *then for any* $\lambda > 0$ *and* $i = 1, ..., k$ *we have*

$$-R_2 n^{-\varepsilon} + M_{K^i(T)}(\lambda - q_2(n)) \leqq \mathrm{pr}\left\{\tilde{S}_n^i(t) - \lambda \frac{\tilde{S}_n^i(t)}{n^{1/2}(1-K_n^i(t))} \leqq \right.$$

$$\left. \leqq S^i(t) \leqq \tilde{S}_n^i(t) + \lambda \frac{\tilde{S}_n^i(t)}{n^{1/2}(1-K_n^i(t))}, -\infty < t \leqq T \right\} \leqq M_{K^i(T)}(\lambda + q_2(n)) + R_2 n^{-\varepsilon}.$$

Since $K^i(T) \leqq b^2/(1+b^2)$, $i=1, ..., k$, we have $M(\lambda - q_j(n)) < M_{b^2/(1+b^2)}(\lambda - q_j(n)) \leqq$ $\leqq M_{K^i(T)}(\lambda - q_j(n))$, $i=1, ..., k$; $j=1, 2$, thus $M_{K^i(T)}$ can be replaced by either M (as noted by Hall and Wellner [14]), or $M_{b^2/(1+b^2)}$ in the lower bounds. Since the choice of ε is ours, the only unknown quantity in the lower bounds $R_j n^{-\varepsilon} +$ $+ M_{b^2/(1+b^2)}(\lambda - q_j(n))$, $j=1, 2$, is b, and this can be estimated by $(1-F_n(T))^{-1}$.

If $k=2$ and we are in the Kaplan—Meier model, then the symmetric bands of the second statement of the above corollary are those of [14] (without rates). Even if we compute with the conjecture $D=2$ but with $A_1=100$, $A_2=10$ and $A_3=1/50$, a practical application of the lower halves of the above inequalities would demand rather astronomic sample sizes. Nevertheless, the above inequalities constitute the only information presently available for the precision of the bands in question, and if one can dream about future values of the A's as A_1, $A_2 \approx 1/10$, $A_3 \approx 10$, then this information is not disappointing at all.

References

[1] O. Aalen, Nonparametric inference in connection with multiple decrement models, *Scand. J. Statist.*, 3 (1976), 15—27.

[2] S. M. Berman, A note on extreme values, competing risks and semi-Markov processes, *Ann. Math. Statist.*, 34 (1963), 1104—1106.

[3] N. Breslow and J. Crowley, A large sample study of the life table and product limit estimates under random censorship, *Ann. Statist.*, 2 (1974), 437—453.

[4] M. D. Burke, S. Csörgő and L. Horváth, Strong approximation of some biometric estimates under random censorship. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 56 (1981), 87—112.

[5] M. Csörgő and P. Révész, *Strong approximation in probability and statistics*, Academic Press (New York, 1981).

[6] S. Csörgő, Limit behaviour of the empirical characteristic function, *Ann. Probability*, 9 (1981), 130—144.

[7] S. Csörgő and L. Horváth, On the Koziol-Green model for random censorship, *Biometrika*, 68 (1981), 391—401.

[8] S. Csörgő and L. Horváth, On cumulative hazard processes under random censorship, *Scand. J. Statist.*, 9 (1982), 13—21.

[9] S. Csörgő and L. Horváth, Random censorship from the left, *Studia Sci. Math. Hung.*, to appear.

[10] L. P. Devroye and G. L. Wise, On the recovery of discrete probability densities from imperfect measurements, *J. Franklin Inst.*, **307** (1979), 1—20.

[11] A. Dvoretzky, J. Kiefer and J. Wolfowitz, Asymptotic minimax character of the sample distribution function and of the multinomial estimator, *Ann. Math. Statist.*, **27** (1956), 642—669.

[12] B. Efron, The two-sample problem with censored data, *Proc. Fifth Berkeley Symp. Math. Statist. Probability*, **4** (1967), 831—853.

[13] M. J. Gillespie and L. Fisher, Confidence bands for the Kaplan—Meier survival curve estimate, *Ann. Statist.*, **7** (1979), 920—924.

[14] W. J. Hall and J. A. Wellner, Confidence bands for a survival curve from censored data, *Biometrika*, **67** (1980), 133—143.

[15] E. L. Kaplan and P. Meier, Nonparametric estimation from incomplete observations, *J. Amer. Statist. Assoc.*, **53** (1958), 457—481.

[16] J. Komlós, P. Major and G. Tusnády, An approximation of partial sums of independent r. v.'s and the sample d. f. I., *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **32** (1975), 111—132.

[17] J. A. Koziol, Goodness-of-fit tests for randomly censored dara, *Biometrika*, **67** (1980), 693—696.

[18] J. A. Koziol and S. B. Green, A Cramér—von Mises statistic for randomly censored data, *Biometrika*, **63** (1976), 465—474.

[19] P. Meier, Estimation of a distribution function from incomplete observations, *Perspectives in Probability and Statistics*, Academic Press (New York, 1975), pp. 67—87.

[20] V. N. Nair, Plots and tests for goodness of fit with randomly censored data, *Biometrika*, **68** (1981), 99—103.

[21] G. Tusnády, *Investigations of statistical hypotheses* (Hungarian), Candidate's dissertation. Hungarian Academy of Sciences (1978).

[22] G. Yang, Life expectancy under random censorship, *Stochastic Processes App.*, **6** (1977), 33—39.

BOLYAI INSTITUTE
SZEGED UNIVERSITY
ARADI VÉRTANÚK TERE 1
H—6720 SZEGED, HUNGARY