The Annals of Statistics
1986, Vol. 14, No. 2, 619–637

EMPIRICAL PROCESSES ASSOCIATED WITH V-STATISTICS AND A CLASS OF ESTIMATORS UNDER RANDOM CENSORING

BY MICHAEL G. AKRITAS

The Pennsylvania State University

A class of empirical processes associated with V-statistics (V-empirical process) under random censoring, and a class of nonparametric estimators based on the corresponding quantile process are defined. The V-empirical process is the censored data analogue of the U-empirical process considered by Silverman (1976, 1983). The class of estimators is the analogue of the class of generalized L-statistics introduced by Serfling (1984) and it includes the results of Sander (1975). The weak convergence of the V-empirical process and the corresponding quantile process is obtained and, through that, the asymptotic behavior of the estimators is studied. Linear bounds for the Kaplan–Meier estimator near the origin are established. A number of examples are given, including the generalization of the Hodges–Lehmann estimator for estimating the treatment effect in the two-sample problem under random censoring. A measure of spread, a procedure for estimation in the two-way ANOVA model, and a modified version of the two-sample Hodges–Lehmann estimator, all of which are new even in the uncensored case, are proposed.

1. Introduction. For each s, $s = 1, \ldots, k$, let $X_{s1}^0, \ldots, X_{sN_s}^0$ be a sample of independent identically distributed observations with distribution function F_s (i.i.d. F_s) and $Y_{s1}, \ldots, Y_{sN_s}$ be i.i.d. G_s. Assume that X_{sj}^0, Y_{mi} are independent for all s, $m = 1, \ldots, k$, $j = 1, \ldots, N_s$, $i = 1, \ldots, N_m$. For each $s = 1, \ldots, k$ we observe

$$X_{si} = \min(X_{si}^0, Y_{si}) \quad \text{and} \quad \delta_{si} = I[X_{si} = X_{si}^0], \qquad i = 1, \ldots, N_s. \tag{1.1}$$

Clearly $X_{si}, \ldots, X_{sN_s}$ are i.i.d. H_s where $(1 - H_s) = (1 - F_s)(1 - G_s)$. For $r_s \le N_s$, $s = 1, \ldots, k$, let

$$h_{k;\,r} \equiv h(x_{11}, \ldots, x_{1r_1}; \ldots; x_{k1}, \ldots, x_{kr_k}) \tag{1.2}$$

be a real valued kernel where $r = (r_1, \ldots, r_k)$. In this paper we deal with the problem of estimating some functional of the distribution of $h_{k;\,r}$ under $F_1, \ldots, F_k$ (such as the median, or some other linear combination of its quantiles) when the sample is of the form (1.1) (random censorship model).

In the uncensored case the problem of estimating the mean of $h_{k;\,r}$ was initiated by Hoeffding (1948), who introduced the class of U-statistics and triggered a long sequence of interesting research. See Serfling (1980) for a modern treatment and references. However, the problem of estimating other functionals of the distribution of $h_{k;\,r}$ did not receive any attention until very recently

Received November 1982; revised October 1985.

AMS 1980 *subject classifications.* Primary 62G05, 62G30; secondary 62E20.

Key words and phrases. V-statistics, empirical processes, random censoring, L-statistics, Kaplan–Meier estimator, Hodges–Lehmann estimator.

(Serfling, 1984). This is even more surprising in view of the fact that the lack of robustness of averages was recognized a long time ago.

The class of estimators to be studied includes such statistics as

$$\operatorname{med}\left\{ \frac{X_{1i_1} + \cdots + X_{1i_r}}{r} - \frac{X_{2j_1} + \cdots + X_{2j_r}}{r}; \right.$$
$$\left. i_1, \ldots, i_r = 1, \ldots, N_1,\ j_1, \ldots j_r = 1, \ldots, N_2 \right\}$$

(and its censored data analogue), which can be thought of as a Hodges–Lehmann type statistic for estimating the shift in the two-sample case. A similar extension of the one-sample Hodges–Lehmann estimator, namely

$$\operatorname{med}\left\{ \frac{X_{i_1} + \cdots + X_{i_r}}{r};\ i_1, \ldots, i_r = 1, \ldots, N \right\}$$

was considered (in the uncensored case) by Serfling and Thornton (1982); it was found that using $r = 3$ (instead of the usual $r = 2$) increases the asymptotic relative efficiency from 0.95 to 0.98, while $r = 4$ increases it to 0.99.

Consider for simplicity the case $k = 1$. In the case of uncensored observations, the problem of estimating the distribution of $h_{1;r}$ has been treated by constructing the empirical distribution function corresponding to the set of $N(N-1)\cdots(N-r+1)$ random variables $h(X_{1i_1}, \ldots, X_{1i_r})$ obtained by every possible choice of ordered sets of r distinct integers drawn from $1, \ldots, N$. Such empirical processes were considered by Silverman (1976, 1983). See also Serfling and Thornton (1982). The problem of estimating functionals of the distribution of $h_{1;r}$ was initiated by Serfling (1984), who defined a class of generalized L-, M-, and R-statistics essentially by placing the above mentioned empirical distribution function into the functional form of the usual L-, M-, and R-statistics.

In this paper we consider the random censoring model and deal with the problem of estimating the distribution of $h_{k;r}$ as well as functionals thereof, thus extending the results of Serfling (1984) and Silverman (1976, 1983) in this case. For reasons that will become apparent, we consider instead the empirical distribution function corresponding to V-statistics [cf. Serfling (1980), page 174]. Section 2 presents, for illustrative purposes, the generalization of the Hodges–Lehmann estimator for estimating the treatment effect in the two sample problem under random censoring. In Section 3 we present the empirical distribution function corresponding to the general kernel (1.2), establish its weak convergence to a Gaussian process and do similarly for the corresponding quantile process. The proof uses a Skorokhod construction; in the absence of censoring the results we obtain are identical with that of Silverman (1976). Incidentally, we show that in the uncensored case the V-process is asymptotically equivalent to the U-process so that our method provides a simpler proof for the weak convergence of the process considered by Silverman. In Section 4 we follow Serfling (1984) in defining generalized L-estimators. This does not only extend Serfling's results to the case of censored survival data, it also generalizes the results of

Sander (1975). Serfling's approach of differentiable statistical functionals could also be applied in our case; however, we chose to present a proof adapted from Shorack (1972). A number of results concerning the behavior of the ratio of the V-empirical process to the true distribution of the kernel that are required for such a proof are formulated and proved in the appendix. In particular, we establish linear bounds for the Kaplan–Meier estimator near the origin. A number of examples including a modified version of the two-sample Hodges–Lehmann estimator, a measure of spread, and a procedure of estimation in two-way ANOVA, all of which are new even in the uncensored case, are presented in Section 5.

2. The Hodges–Lehmann estimator. In the notation of Section 1, let $k = 2$ and consider the kernel

$$h_{2;1,1} = h(x_1; x_2) = x_1 - x_2. \tag{2.1}$$

In the uncensored case the Hodges–Lehmann estimator for estimating the treatment effect in the two sample problem is the median of the uniform probability measure that assigns mass $N_1^{-1}N_2^{-1}$ to each of the points $h(X_{1i}; X_{2j})$, $i = 1, \dots, N_1$, $j = 1, \dots, N_2$. Noting that N_s^{-1}, $s = 1, 2$, is the mass assigned to each X_{si}, $i = 1, \dots, N_s$, by the corresponding empirical distribution function, we conclude that an appropriate analogue of the Hodges–Lehmann estimator in the presence of censoring is the median of the probability measure that assigns mass $[\hat{F}_1(X_{1i}) - \hat{F}_1(X_{1i} -)] \cdot [\hat{F}_1(X_{2j}) - \hat{F}_2(X_{2j} -)]$ to each of the points $h(X_{1i}; X_{2j})$. Here $\hat{F}_s \equiv \hat{F}_{s,N_s}$, $s = 1, 2$, is the Kaplan–Meier estimator corresponding to the sth sample (Kaplan and Meier, 1958). Thus, the above weights are nonzero only if both X_{1j} and X_{2j} are uncensored observations.

Formally, let $N = N_1 + N_2$, let $h(x_1; x_2)$ be as in (2.1), and set

$$\hat{V}_N(t) = \int\int I[h(x_1;x_2) \le t]\, d\hat{F}_1(x_1)\, d\hat{F}_2(x_2), \qquad t \in (-\infty, \infty). \tag{2.2}$$

Thus, $\hat{V}_N(t)$ is, for each t, a V-statistic with kernel $I[h(x_1;x_2) \le t]$. Further let

$$\hat{V}_N^{-1}(p) = \inf\{t: \hat{V}_N(t) \ge p\}, \qquad 0 < p < 1. \tag{2.3}$$

Then the generalized Hodges–Lehmann estimator defined above is given by $\hat{V}_N^{-1}(0.5)$. In the absence of censoring, this is the usual Hodges–Lehmann estimator. Thus, the generalized Hodges–Lehmann estimator belongs in the class of statistics considered in Section 4 where its asymptotic distribution is obtained.

Before concluding this section we give a proposition which shows that the above generalization of the Hodges–Lehmann estimator is a reasonable one. First, recall that Efron's generalization of the Mann–Whitney–Wilcoxon statistic is given by $\int(1 - \hat{F}_1)\, d\hat{F}_2$ (Efron, 1967).

PROPOSITION 2.1. *Let* $\hat{V}_N^{-1}(0.5)$ *be the generalized Hodges–Lehmann estimator as defined by* (2.2) *and* (2.3). *Then* $\hat{V}_N^{-1}(0.5)$ *is given by*

$$\inf\left\{t: \int \hat{F}_1(x + t)\, d\hat{F}_2(x) \ge 0.5\right\}. \tag{2.4}$$

PROOF. Clearly the T that satisfies (2.4) is the median of the convolution of X_1 and $-X_2$ when the distribution of X_1 is $\hat{F}_1$ and that of X_2 is $\hat{F}_2$. But this is just $\hat{V}_N^{-1}(0.5)$.

Thus $\hat{V}_N^{-1}(0.5)$ is obtained from Efron's statistic the same way that the usual Hodges–Lehmann estimator is obtained from the Mann–Whitney–Wilcoxon statistic (Hodges and Lehmann, 1963). Clearly we can obtain other estimators for the treatment effect in the two-sample problem under random censoring by inverting appropriate generalizations of other rank statistics. □

REMARK 2.1. Padgett and Wei (1982) derived (from different context, motivation, and method) the same generalization of the Hodges–Lehmann estimator but they only proved a consistency result. Also Wei and Gail (1983) considered inversion of a class of two-sample rank tests in order to obtain rank estimates of the scale ratio; since their method was tailored out of Hodges and Lehmann (1963), the entire class of their estimators was called generalized Hodges–Lehmann estimators. The results of these authors, however, were derived under the additional assumption that the censoring variable in the second sample has undergone the same scale transformation as the "survival" time and thus their applicability may be limited.

3. The V-empirical process. Now let $h_{k;r}$ be the general kernel of relationship (1.2) and consider the problem of estimating the distribution of it under the random censoring model, that is, when the data are of the form (1.1). In the case of uncensored observations the problem has been treated by constructing an empirical distribution function associated with U-statistics corresponding to kernels $I[h_{k;r} \le t]$, $-\infty < t < \infty$. With censored data, however, it is computationally more convenient to consider an empirical distribution function associated with V-statistics. To see why, consider the special case $k = 1$, $r = 2$; then the U-statistic in the uncensored case assigns weight $N^{-1}(N-1)^{-1}$ to each point $I[h_{1;2}(X_{1i}, X_{1j}) \le t]$, $i \ne j$. This is the weight that the empirical distribution function corresponding to the whole sample $X_{11}, \dots, X_{1N}$ assigns to X_{1i} times the weight that the empirical distribution function corresponding to $X_{11}, \dots, X_{1,i-1}, X_{1,i+1}, \dots, X_{1N}$ assigns to X_{1j}. Thus the analogue of a U-statistic for censored data would require, in the general case, computing several Kaplan–Meier estimators.

Consider now the general kernel as in (1.2), and set

$$N = \sum_{s=1}^{k} N_s, \qquad \lambda_s = \lim_{N\to\infty} (N_s/N), \tag{3.1}$$

$$\hat{V}_N(t) = \int_0^{T_1} \dots \int_0^{T_k} I[h_{k;r} \le t] \prod_{(s,i)} d\hat{F}_s(X_{si}), \tag{3.2}$$

$$V(t) = \int_0^{T_1} \dots \int_0^{T_1} I[h_{k;r} \le t] \prod_{(s,i)} dF_s(X_{si}). \tag{3.3}$$

Here $\hat{F}_s \equiv \hat{F}_{s, N_s}$, $s = 1, \dots, k$, is the Kaplan–Meier estimator of F_s (see notation in Section 1), $\prod_{(s,i)}$ denotes the double product $\prod_{s=1}^{k}\prod_{i=1}^{r_s}$, and if we define

$$T_F = \sup\{t\colon F(t) < 1\}, \quad \text{where } F \text{ is a distribution function,} \tag{3.4}$$

then the numbers $T_1, \dots, T_k$ are any numbers satisfying

$$T_s < \min(T_{F_s}, T_{G_s}), \qquad s = 1, \dots, k. \tag{3.5}$$

We are going to study the weak convergence of the process

$$\hat{W}_N(t) = N^{1/2}[\hat{V}_N(t) - V(t)]. \tag{3.6}$$

In order to formulate the first theorem, we need additional notation. Set

$$g_{s,i}(x|t) = \int \dots \int I[h_{k;r} \le t] \prod_{\substack{(s_1, j)\\ (s_1, j) \ne (s, i)}} dF_{s_1}(x_{s_1 j}), \qquad x \ge 0, \qquad s = 1, \dots, k, \tag{3.7}$$

where the domain of integration with respect to F_{s_1} is $[0, T_{s_1}]$.

REMARK 3.1. The x that appears on the left-hand side of (3.7) corresponds to the (s, i)th argument of $h_{k;r}$. Thus in the absence of censoring, $g_{s,i}(x|t) = P[h_{k;r} \le t | X_{si} = x]$.

Next set

$$g_s(x|t) = \sum_{i=1}^{r_s} g_{s,i}(x|t), \qquad x \ge 0, \qquad s = 1, \dots, k, \tag{3.8}$$

and

$$\hat{L}_s(x) \equiv \hat{L}_{s, N_s}(x) = N_s^{1/2}[\hat{F}_s(x) - F_s(x)], \qquad s = 1, \dots, k. \tag{3.9}$$

It is then well known that there exists a version of $\hat{L}_s$ and a Gaussian process L_s such that

$$\|\hat{L}_s - L_s\|_0^{T_s} \to 0 \quad \text{a.s.,} \qquad s = 1, \dots, k,$$

where $\|\cdot\|$ denotes the sup-norm. The process $L_s(x)$, $x \ge 0$, is equivalent in law to

$$B^0[K_s(x)]\frac{1 - F_s(x)}{1 - K_s(x)}, \qquad x \ge 0,$$

where B^0 is the Brownian bridge process on $[0, 1]$, and $K_s(x) = C_s(x)/(1 + C_s(x))$ with $C_s(x) = \int_0^x (1 - F_s)^{-2}(1 - G_s)^{-1}\, dF_s$.

Finally set

$$W(t) = \sum_{s=1}^{k} \lambda_s^{-1/2} \int_0^{T_s} g_s(x|t)\, dL_s(x), \tag{3.10}$$

where T_s is defined by (3.4) and (3.5).

THEOREM 3.1. *Assume that* λ_s, *as defined in* (3.1), *is positive for all* $s = 1, \dots, k$, *and let* $g_s(\cdot|t)$ *be of bounded variation in* $[0, \infty)$ *uniformly in* $t \in (-\infty, \infty)$ *for each* $s = 1, \dots, k$ *(see Proposition* 3.2*). Let* $\hat{W}_N(t)$ *and* $W(t)$ *be the processes defined by* (3.6) *and* (3.10), *respectively. Then there exists a version of the process* $\hat{W}_N$ *such that*

$$\|\hat{W}_N(t) - W(t)\|_{-\infty}^{\infty} \to 0 \quad \text{as } N \to \infty$$

almost surely, where $\|\cdot\|$ *denotes sup-norm.*

PROOF. Write

$$\hat{W}_N(t) = N^{1/2} \int_0^{T_1} \cdots \int_0^{T_k} I[h_{k;r} \le t] \left[\prod_{(s,i)} d\hat{F}(x_{si}) - \prod_{(s,i)} dF_s(x_{si}) \right]$$

and use the formula

$$\begin{aligned}
\prod_{i=1}^{N} a_i - \prod_{i=1}^{N} b_i &= \sum_{k=1}^{N} (a_k - b_k) \prod_{i=k+1}^{N} b_i \prod_{i=1}^{k-1} [b_i + (a_i - b_i)] \\
&= \sum_{k=1}^{N} (a_k - b_k) \prod_{i=k+1}^{N} b_i \left\{ \prod_{i=1}^{k-1} b_i + \text{terms involving } (a_i - b_i) \right\} \\
&= \sum_{k=1}^{N} (a_k - b_k) \prod_{i \ne k} b_i + \text{terms of at least second order in } (a_i - b_i).
\end{aligned}$$

We get

$$\begin{aligned}
&\prod_{(s,i)} d\hat{F}_s(x_{si}) - \prod_{(s,i)} dF_s(x_{si}) \\
&\quad = \sum_{(s,i)} \prod_{\substack{(s_1,j) \\ (s_1,j) \ne (s,i)}} dF_{s_1}(x_{s_1 j})\, d[\hat{F}_s(x_{s1}) - F(x_{si})] \\
&\qquad + \text{terms of at least second order in } d[\hat{F}_s(x_{si}) - F_s(x_{si})].
\end{aligned}$$

From this and a simple argument it follows that

$$\hat{W}_N(t) = N^{1/2} \sum_{s=1}^{k} \int_0^{T_s} g_s(x|t)\, d[\hat{F}_s(x) - F_s(x)] + O_N(t), \tag{3.11}$$

where the process $O_N(\cdot)$ is such that $\|O_N\|_{-\infty}^{\infty} \to 0$ in probability. Using this and

Natanson (1961), page 232, we have

$$\|\hat{W}_N(t) - W(t)\|_{-\infty}^{\infty} < \|O_N(t)\|_{-\infty}^{\infty} + \sum_{s=1}^{k} \left| \left(\frac{N}{N_s}\right)^{1/2} \hat{L}_s(T_s) - \lambda_s^{-1/2} L(T_s) \right| \|g_s(T_s| \cdot)\|_{-\infty}^{\infty} + \sum_{s=1}^{k} \left\| \left(\frac{N}{N_s}\right)^{1/2} \hat{L}_s - \lambda_s^{-1/2} L_s \right\|_0^{T_s} \sup_t TV_{[0,T_s]}[g_s(\cdot|t)],$$

where $TV_{[a,b]}$ denotes the total variation in $[a, b]$. So it suffices to show that each of the terms above converges to zero a.s. This is true for the last term since $TV_{[0,\infty)}[g_s(\cdot|t)] \le M < \infty$ for all t by assumption; noting that, by (3.7) and (3.8), $\sup_t[g_s(T_s|t)] \le r_s$, the second term is easily seen to converge to zero, and this completes the proof of the theorem. □

Next, in order to find the covariance $\rho(v, t)$ of the process $W(t)$, note that the k terms in (3.10) are independent, so that $\rho(v, t) = \sum_{s=1}^{k} \lambda_s^{-1} \rho_s(v, t)$, where $\rho_s(v, t)$ is the covariance function of

$$W_s(t) = \int_0^{T_s} g_s(x|t)\, dL_s(x) =_L \int_0^{T_s} g(x|t)\, d\left\{ B^0[K_s(x)] \frac{1 - F_s(x)}{1 - K_s(x)} \right\}.$$

But $B^0(u) =_L B(u) - uB(1)$, $u \in [0,1]$, where B is the standard Brownian motion. Thus,

$$\begin{aligned} W_s(t) =_L & \int_0^{T_s} g_s(x|t) \frac{1 - F_s(x)}{1 - K_s(x)}\, dB[K_s(x)] \\ & + \int_0^{T_s} g_s(x|t) B[K_s(x)]\, d\frac{1 - F_s(x)}{1 - K_s(x)} \\ & - B(1) \int_0^{T_s} g_s(x|t)\, dK_s(x) \frac{1 - F_s(x)}{1 - K_s(x)} \\ \equiv\ & A_{s1}(t) + A_{s2}(t) - A_{s3}(t), \quad \text{say.} \end{aligned} \tag{3.12}$$

Direct computation gives

COROLLARY 3.1. *Let* K_s, $s = 1, \dots, k$, *be defined in connection with* (3.9) *and* A_{si}, $i = 1, 2, 3$, $s = 1, \dots, k$, *be defined by* (3.12). *Then under the notation and assumptions of Theorem* 3.1, *the process* $\hat{W}_N(t)$ *converges weakly to a mean zero Gaussian process with covariance function given by*

$$\rho(v, t) = \sum_{s=1}^{k} \lambda_s^{-1} \rho_s(v, t), \tag{3.13}$$

where

$$\rho_s(v,t) = EA_{s1}(v)A_{s1}(t) + EA_{s1}(v)A_{s2}(t) - EA_{s1}(v)A_{s3}(t)$$
$$(3.14) \qquad +EA_{s2}(v)A_{s1}(t) + EA_{s2}(v)A_{s2}(t) - EA_{s2}(v)A_{s3}(t)$$
$$-EA_{s3}(v)A_{s1}(t) - EA_{s3}(v)A_{s2}(t) + EA_{s3}(v)A_{s3}(t)$$

and

$$EA_{s1}(v)A_{s1}(t) = \int_0^{T_s} g_s(x|t)g_s(x|v)D_s^2(x)\,dK_s(x),$$

$$EA_{s2}(v)A_{s2}(t) = \int_0^{T_s}\int_0^{T_s} g_s(x|v)g_x(y|t)[K_s(x) \wedge K_s(y)]\,dD_s(x)\,dD_s(y),$$

$$EA_{s3}(v)A_{s3}(t) = \int_0^{T_s} g_s(x|v)\,d[K_s(x)D_s(x)] \cdot \int_0^{T_s} g_s(x|t)\,d[K_s(x)D_s(x)],$$

$$EA_{s1}(v)A_{s2}(t) = \int_0^{T_s} g_s(x|t) \cdot \int_0^{T_s \wedge x} g_s(y|v)\,D_s(y)\,dK_s(y)\,dD_s(x),$$

$$EA_{s1}(v)A_{s3}(t) = \int_0^{T_s} g_s(x|t)\,d[K_s(x)D_s(x)] \cdot \int_0^{T_s} g_s(x|v)D_s(x)\,dK_s(x),$$

$$EA_{s2}(v)A_{s3}(t) = \int_0^{T_s} g_s(x|v)K_s(x)\,dD_s(x) \cdot \int_0^{T_s} g_s(x|t)\,d[K_s(x)D_s(x)],$$

where $D_s = (1 - F_s)/(1 - K_s)$ *and* $a \wedge b = \min(a, b)$.

COROLLARY 3.2. *Let* $\hat{V}_N^{-1}(p)$, $0 < p < 1$, *be the empirical quantile process, where* $\hat{V}_N^{-1}(p) = \inf\{t\colon \hat{V}_N(t) \geq p\}$, *let* $V^{-1}(p)$ *be similarly defined, and consider the notation and assumptions of Corollary* 3.1. *Then*

(a) $N^{1/2}[V \circ \hat{V}_N^{-1}(p) - p]$, $0 < p < \hat{V}_N(\infty)$, *converges weakly to a mean zero Gaussian process* $Z(p)$ *with covariance* $\rho(V^{-1}(p), V^{-1}(q))$.

(b) $N^{1/2}[\hat{V}_N^{-1}(p) - V^{-1}(p)]$, $0 < p < \hat{V}_N(\infty)$, *converges weakly to* $Z(p)/V'(V^{-1}(p))$ *provided the derivative* V' *of* V *exists and is continuous on* $(0, V(\infty))$.

PROOF. The proof follows from Corollary 3.1 and the results of Vervaat (1972).

Note that in the uncensored case $D_s \equiv 1$ so that, for $k = 1$ and for $T_1 = \infty$, formula (3.14) reduces to the formula of Theorem B, Silverman (1976), or formula (5) of Silverman (1983). This, however, does not constitute yet an alternative proof of the weak convergence of the U-empirical process $\hat{G}_N(t)$ considered by Silverman. In order to obtain such a proof, set $N_1 = N$, $r_1 = r$ and note that since $\hat{G}_N(t)$ is, for each fixed t, the U-statistic with kernel $I[h_{1;r} \leq t]$ we have

$$\hat{V}_N(t) = \frac{N_{(r)}}{N^r}\hat{G}_N(t) + \left(1 - \frac{N_{(r)}}{N^r}\right)\hat{H}_N(t).$$

Here $\hat{V}_N(t)$ is given by (3.2) with $h_{1;r}$ and no censoring, $\hat{H}_N(t)$ is the average of all terms $I[h_{1;r}(X_{1j_1}, \ldots, X_{1j_r}) \leq t]$ with at least one equality $i_\alpha = i_\beta$, $\alpha \neq \beta$, and $N_{(r)} = N(N-1)\dots(N-r+1)$. It follows that

$$\begin{aligned}\|N^{1/2}(\hat{V}_N - \hat{G}_N)\| &= N^{1/2}\left(1 - \frac{N_{(r)}}{N^r}\right)\|\hat{H}_N - \hat{G}\| \\ &\le N^{1/2}\frac{N^r - N_{(r)}}{N^r}\left[\|\hat{H}_N\| + \|\hat{G}_N\|\right] \\ &= 2N^{1/2}\frac{N^r - N_{(r)}}{N^r} \to 0.\end{aligned}$$

Thus we have established

PROPOSITION 3.1. *In the uncensored case the U-empirical process $\hat{G}_N(t)$ is asymptotically equivalent to the V-empirical process $\hat{V}_N(t)$.*

The next result provides a sufficient condition for $g_s(\cdot\,|t)$ to be of bounded variation in $[0, \infty)$ uniformly in t (see Theorem 3.1). Assumption 3.1 below is also used in the appendix. Consider

$$h_{s,i}(x_{si}) \equiv h(X_{11}, \dots, X_{1r_1}; \dots; x_{s1}, \dots, X_{si}, \dots, X_{sr_s}; \dots; X_{k1}, \dots, X_{kr_k})$$

as a stochastic process in x_{si}, $i = 1, \dots, r_s$, $s = 1, \dots, k$.

ASSUMPTION 3.1. *Almost surely $[P_s]$, where $P_s = F_1^{r_1} \times \cdots \times F_s^{r_s - 1} \times \cdots \times F_k^{r_k}$ there exists a partition of $[0, \infty)$ such that the function $h_{s,i}(y)$, $0 \le y < \infty$ is monotonic within each interval of the partition for all $i = 1, \dots, r_s$, $s = 1, \dots, k$. Moreover there exists a positive number $M_h < \infty$ such that the number of intervals in each of the above partitions is $\le M_h$ almost surely $[P_s]$, $i = 1, \dots, r_s$, $s = 1, \dots, k$.*

PROPOSITION 3.2. *Under Assumption 3.1, the function $g_s(\cdot\,|t)$ is of bounded variation in $[0, \infty)$ uniformly in $t \in (-\infty, \infty)$ for each $s = 1, \dots, k$.*

PROOF. From (3.8) it follows that

$$TV_{[0,\infty)}[g_s(\cdot\,|t)] \le \sum_{i=1}^{r_s} TV_{[0,\infty)}[g_{s,i}(\cdot\,|t)]. \tag{3.15}$$

By definition,

$$\begin{aligned}TV_{[0,a]}[g_{s,i}(\cdot\,|t)] &= \sup \sum_j |g_{s,i}(y_j|t) - g_{s,i}(y_{j-1}|t)| \\ &\le \sup \sum_j \int \cdots \int \big| I[h_{s,i}(y_j) \le t] \\ &\qquad - I[h_{s,i}(y_{j-1}) \le t] \big| \prod_{\substack{(s_1, i_1) \\ (s_1, i_1) \ne (s, i)}} dF_{s_1}(x_{s_1, i_1}),\end{aligned}$$

where the supremum is taken over all partitions of $[0, a]$. But under Assumption

3.1, the number of times that the process $h_{s,i}(y)$, $0 \le y < \infty$, will cross the number t is $\le M_h$ almost surely P_s. Thus

$$\sum_j |I[h_{s,i}(y_j) \le t] - I[h_{s,i}(y_{j-1}) \le t]| \le M_h$$

for all partitions of $[0, a]$ almost surely $[P_s]$. Since this is true for all $a > 0$, the result follows from (3.15). □

REMARK 3.2. If $T_{F_s} < T_{G_s}$ then the sth domain of integration in (3.2) can be from 0 to $\hat{T}_s = \max\{X_{si};\ i = 1, \dots, N_s\}$ while in relation (3.3) it can be from 0 to $T_s = T_{F_s}$. Indeed, in this case $(1 - K_s)/(1 - F_s)$ remains bounded away from zero and thus Theorem 1.2 in Gill (1983) implies that $\|\hat{L}_s - L_s\|_0^{\hat{T}_s} \to 0$ a.s.; the rest of the arguments in Theorem 3.1 follow with minor adjustments.

4. Generalized *L*-statistics. In this section we extend the notion of generalized L-statistics, as introduced by Serfling (1984), to the case of censored survival data. Recall that if $X_1, \dots, X_N$ are i.i.d. F and $\hat{F}_N$ denotes the corresponding empirical distribution function, an L-statistic $N^{-1}\Sigma_i C_{Ni}\tilde{g}(\hat{F}_N^{-1}(i/N))$ may be written as

$$\int_0^1 J_N(s)\tilde{g}(\hat{F}_N^{-1}(s))\,ds, \tag{4.1}$$

where $J_N(s) = C_{Ni}$ for $s \in ((i-1)/N, i/N]$, $i = 1, \dots, N$. This functional form of an L-statistic lends itself to generalization. In particular, if we substitute the empirical process considered by Silverman (1976, 1983) instead of $\hat{F}_N$ in (4.1) we obtain the class of generalized L-statistics considered by Serfling (1984). And if we substitute the process $\hat{V}_N$ considered in Section 3 we obtain the extension of the class of generalized L-statistics to the case of censored survival data, which will be the object of study in this section. But now $J_N(s)$ is not suitable as defined above. Due to the fact that $\hat{V}_N$ has jumps of random size, J_N will have to be replaced by a function $\hat{J}_N$, say, which is constant over random intervals. Also $\hat{V}_N(\infty)$ is not necessarily equal to one. Thus the statistic we will study is of the form

$$T_N = \int_0^{\hat{V}_N(\infty)} \hat{J}_N(s)\tilde{g}(\hat{V}_N^{-1}(s))\,ds. \tag{4.2}$$

To illustrate this point further, consider for simplicity the case $k = 1$ (one sample) and $h(x) = x$, so that $\hat{V}_N = \hat{F}_N$, the Kaplan–Meier estimator. Due to the fact that under random censoring we end up with a random number of $\tilde{N} \le N$ of uncensored observations, the construction of linear combinations of order statistics consists, in addition to choosing the weights $C_{N1}, \dots, C_{NN}$, in deciding what weight corresponds to each uncensored observation. If we define the rank of the uncensored observation X_i as $N\hat{F}_N(X_i)$, then we may assign to X_i the weight C_{Nj} with $j = [N\hat{F}_N(X_i)]$ ($[\cdot]$ denotes integer part). Note that if $J_N(s) = C_{Ni}$ for $s \in ((i-1)/N, i/N]$, as before, then the above assignment of weights corresponds to $\hat{J}_N(s) = J_N(\hat{F}_N(\hat{F}_N^{-1}(s)))$ in (4.2) with $\hat{V}_N$ replaced by $\hat{F}_N$; since in the uncensored case (i.e., when $\hat{F}_N$ is the usual empirical distribution function)

$J_N(s) = J_N(\hat{F}_N(\hat{F}_N^{-1}(s)))$ it follows that the above choice of weights is a reasonable one. Some other assignments of weights are discussed in Lemma 4.1.

The purpose of this section is to study the asymptotic distribution of T_N given by (4.2). This will be done by adapting the method of Shorack (1972) which allows unbounded "scores." Let

$$\hat{C}_0 = \hat{V}_N^{-1}(0+), \qquad \hat{C}_1 = \hat{V}_N^{-1}(\hat{V}_N(\infty)),$$
$$C_0 = V^{-1}(0), \qquad C_1 = V^{-1}(V(\infty)) \tag{4.3}$$

and for $\varepsilon \geq 0$, $\beta = \beta(\varepsilon) > 0$, set

$$Q_{N\varepsilon} = \big[\hat{V}_N(t) \leq \beta^{-1}V(t), -\infty < t < \infty;\ \hat{V}_N(t) \geq \beta V(t), \hat{C}_0 \leq t < \infty;$$
$$1 - \hat{V}_N(t) \leq \beta^{-1}[1 - V(t)], -\infty < t \leq \hat{C}_1; \tag{4.4}$$
$$1 - \hat{V}_N(t) \geq \beta[1 - V(t)], -\infty < t \leq \hat{C}_1\big].$$

PROPOSITION 4.1. *Let Assumption* 3.1 *hold with* $M_h = 1$ (*see Remark* A.1 *in the Appendix*). *Then for any* $\varepsilon > 0$ *there exists* $\beta = \beta(\varepsilon) > 0$ *such that*

$$P(Q_{N\varepsilon}) \geq 1 - \varepsilon$$

holds for all N, *where* Q_N *is defined in* (4.4).

PROOF. It follows directly from Theorem A.2 of the Appendix. □

ASSUMPTION 4.1. The function $g = \tilde{g} \circ V^{-1}$ is of bounded variation on $(\theta, 1 - \theta)$ for all $\theta > 0$.

For fixed numbers b_1, b_2 and $K > 0$ define a "scores bounding" function SB by

$$\mathrm{SB}(s) = Ks^{-b_1}(1 - s)^{-b_2} \quad \text{for } 0 < s < 1 \tag{4.5}$$

and for fixed $\delta > 0$ define

$$D(s) = Ks^{-1/2 + b_1 + \delta}(1 - s)^{-1/2 + b_2 + \delta} \quad \text{for } 0 < s < 1.$$

Further, let J be a fixed measurable function on $(0, 1)$.

ASSUMPTION 4.2 (Boundedness). Assume $|g| \leq D$, $|J| \leq \mathrm{SB}$ and for all N, $|\hat{J}_N| \leq \mathrm{SB}$ almost surely.

ASSUMPTION 4.3 (Smoothness). Except on a set of s's of $|g|$-measure zero, we have both that J is continuous at s and $\hat{J}_N \to J$ almost surely uniformly in some small neighborhood of s as $N \to \infty$.

ASSUMPTION 4.4. The function $\mathrm{SB}[V(t)]$ is $\tilde{g}$-integrable on $[C_0, C_1]$ for all $s = 1, \dots, k$.

Before going into the main result of this section we will provide a result that helps check the assumptions $|\hat{J}_N| \le \mathrm{SB}$ a.s. (Assumption 4.2) assuming that we are willing to accept a definition of $\hat{J}_N$ that depends on the choice of b_1, b_2. In particular, if $\hat{J}_N$ is defined as $\hat{J}_N^{(i)}$, $i = 1, \dots, 4$ depending on the choice of b_1, b_2 (see Lemma 4.1 below), then $|\hat{J}_N| \le \mathrm{SB}$ almost surely holds provided that $J_N \le \mathrm{SB}$ holds. Let $h_{k;\,r}$ $r = (r_1, \dots, r_k)$, be the kernel in question and let $J_N(s) = C_{N\iota}$ for $s \in ((i-1)/M, i/M]$ where $M = N_1^{r_1} \dots N_k^{r_k}$, be the choice of weights that would have been used for the generalized L-statistic in the absence of censoring.

LEMMA 4.1. *Let* SB *be given by* (4.5) *and assume that* $|J_N| \le \mathrm{SB}$ *where* J_N *is given above. Then*

(i) *if* $b_1 > 0$, $b_2 < 0$, $|\hat{J}_N^{(1)}| \le \mathrm{SB}$, *where* $\hat{J}_N^{(1)}(s) = J_N(\hat{V}_N(\hat{V}_N^{-1}(s)))$;

(ii) *if* $b_1 < 0$, $b_2 > 0$, $|\hat{J}_N^{(2)}| \le \mathrm{SB}$, *where* $\hat{J}_N^{(2)}(s) = J_N(\hat{V}_{N-}(\hat{V}_N^{-1}(s)))$, *where* $\hat{V}_{N-}$ *denotes the left-continuous version of* $\hat{V}_N$;

(iii) *if* $b_1 > 0$, $b_2 > 0$, $|\hat{J}_N^{(3)}| < \mathrm{SB}$, *where* $\hat{J}_N^{(3)}(s) = J_N(\hat{V}_N(\hat{V}_N^{-1}(s)))$ *for* $s \in [0, S_0]$ *and* $\hat{J}_N^{(3)}(s) = J_N(\hat{V}_{N-}(\hat{V}_N^{-1}(s)))$ *for* $s \in (S_0, 1)$, *where* S_0 *is the point at which* SB *attains its minimum*;

(iv) *if* $b_1 < 0$, $b_2 < 0$, $|\hat{J}_N^{(4)}| < \mathrm{SB}$, *where* $\hat{J}_N^{(4)}(s) = J_N(\hat{V}_{N-}(\hat{V}_N^{-1}(s)))$ *for* $s \in [0, S_1]$ *and* $\hat{J}_N^{(4)}(s) = J_N(\hat{V}_N(\hat{V}_N^{-1}(s)))$ *for* $s \in (S_1, 1)$, *where* S_1 *is the point at which* SB *attains its maximum.*

PROOF. (i) This follows from $\hat{V}_N(\hat{V}_N^{-1}(s)) \ge (i-1)/M$ when $s \in ((i-1)/M, i/M]$ and the fact that SB is, in this case, decreasing. (ii) This follows from $\hat{V}_{N-}(\hat{V}_N^{-1}(s)) \le (i-1)/M$ and the fact that SB is, in this case, increasing. (iii) and (iv) follow by combining (i) and (ii). □

REMARK 4.1. If $|J_N| \le \mathrm{SB}$, with $b_1 < 0$, there does not exist another SB*, $b_1^* < 0$, such that $|\hat{J}_N| \le \mathrm{SB}^*$ with $\hat{J}_N(s) = J_N(\hat{V}_N(\hat{V}_N^{-1}(s)))$.

THEOREM 4.1. *Under Assumptions* 4.1–4,

$$N^{1/2}(T_N - \mu_N) \to -\int_0^{V(\infty)} J(s) W(V^{-1}(s))\, dg(s)$$

in probability, as $N \to \infty$, *where*

$$\mu_N = \int_0^{V(\infty)} \hat{J}_N(s) \tilde{g}(V^{-1}(s))\, ds.$$

PROOF. Let

$$\Psi_N(s) = -\int_s^{V(\infty)} \hat{J}_N(u)\, du$$

and write

$$T_N = -\tilde{g}(\hat{C}_0)\Psi_N(0) - \int_{\hat{C}_0}^{\hat{C}_1}\Psi_N(\hat{V}_N(t))\,d\tilde{g}(t),$$

$$\mu_N = \tilde{g}(t)\Psi_N(V(t))|_{\hat{C}_0}^{\hat{C}_1} - \int_{\hat{C}_0}^{\hat{C}_1}\Psi_N(V(t))\,d\tilde{g}(t) + \int_{[\hat{C}_0,\hat{C}_1)^c}\tilde{g}(t)\,d\Psi_N(V(t)),$$

where $\hat{C}_0, \hat{C}_1$ are given in (4.3) and $[\hat{C}_0, \hat{C}_1)^c$ denotes the complement of $[\hat{C}_0, \hat{C}_1)$ with respect to $[C_0, C_1]$. Thus,

$$\begin{aligned} S_N &\equiv N^{1/2}(T_N - \mu_N) \\ &= -\int_{C_0}^{C_1} A_N^*(t)\hat{W}_N\,d\tilde{g}(t) - (\gamma_{N1} + \gamma_{N2} + \gamma_{N3}), \end{aligned}$$

where

$$A_N^*(t) = \frac{\int_{V(t)}^{\hat{V}_N(t)}\mathcal{J}_n(s)\,ds}{\hat{V}_N(t) - V(t)} \cdot I_{[\hat{C}_0,\hat{C}_1]}(t),$$

$$\gamma_{N1} = N^{1/2}\tilde{g}(\hat{C}_0)[\Psi_N(0) - \Psi_N(V(\hat{C}_0))], \qquad \gamma_{N2} = N^{1/2}\tilde{g}(\hat{C}_1)\Psi_N(V(\hat{C}_1)),$$

and

$$\gamma_{N3} = N^{1/2}\int_{[\hat{C}_0,\hat{C}_1)^c}\tilde{g}(t)\,d\Psi_N(V(t)).$$

Now fix $\varepsilon > 0$ and let β be as in Proposition 4.1. If $\chi_{N\varepsilon} = I_{Q_{N\varepsilon}}(\omega)$, Assumption 4.2 implies

$$\chi_{N\varepsilon}|A_N^*(t)| \le \left|\frac{\int_{V(t)}^{\hat{V}_N(t)}\mathrm{SB}(s)\,ds}{\hat{V}_N(t) - V(t)}\right| I_{[\hat{C}_0,\hat{C}_1]}(t) \le C\cdot\mathrm{SB}(V(t))\cdot I_{[\hat{C}_0,\hat{C}_1]}(t),$$

where the constant C depends on ε, b_1, b_2. Thus with

$$S = -\int_{C_0}^{C_1} J(V(t))W(t)\,d\tilde{g}(t)$$

we have

$$\begin{aligned} |\chi_{N\varepsilon}S_N - S| \le &\int_{C_0}^{C_1}|\chi_{N\varepsilon}A_N^*(t)\hat{W}_N(t) - J(V(t))W(t)|\,d\tilde{g}(t) \\ &+ |\gamma_{N1} + \gamma_{N2} + \gamma_{N3}|. \end{aligned} \tag{4.6}$$

But

$$\begin{aligned} |\chi_{N\varepsilon}A_N^*(t)\hat{W}_N(t) - J(V(t))W(t)| \le\; &C\cdot\mathrm{SB}(V(t))|\hat{W}_N(t)|I_{[\hat{C}_0,\hat{C}_1]}(t) \\ &+ \mathrm{SB}(V(t))|W(t)|, \end{aligned}$$

and, for N large,

$$|\hat{W}_N(t)| \le 2\sum_{s=1}^{k}\left[g_s(T_s|t)\cdot|\hat{L}_s(T_s)| + TV_{[0,T_s]}[g_s(\cdot|t)]\,\|\hat{L}_s(\cdot)\|_0^{T_s}\right],$$

$$|W(t)| \le \sum_{s=1}^{k} TV_{[0,T_s]}[g_s(\cdot|t)]\,\|L_s(\cdot)\|_0^{T_s}$$

so that

$$|\chi_{N\epsilon} A_N^*(t)\hat{W}_N(t) - J(V(t))W(t)|$$

$$\leq \sum_{s=1}^{k} \mathrm{SB}(V(t)) \cdot TV_{[0,T_s]}[g_s(\cdot|t)]\left[2C\|\hat{L}_s(\cdot)\|_0^{T_s} - \|L_s(\cdot)\|_0^{T_s}\right]$$

$$+2C|\hat{L}_s(T_s)| \sum_{s=1}^{k} \mathrm{SB}(V(t))g_s(T_s|t) \cdot I_{[\hat{C}_0, \hat{C}_1]}(t).$$

Thus, since $A_N^*(t) \to J(V(t))$ almost everywhere $|\tilde{g}|$ (Assumption 4.3) and $\|\hat{W}_N(t) - W(t)\| \to 0$, we may, by Assumption 4.4 and Proposition 3.2, apply Pratt's dominated convergence theorem (Pratt, 1960) to conclude that, for each ω the integral on the right-hand side of (4.6) converges to zero. That $\gamma_{N1}, \gamma_{N2}, \gamma_{N3}$ are asymptotically negligible may be shown as in Shorack (1972). Hence,

$$\chi_{N\epsilon} S_N \to S$$

which implies that $S_N \to S$ in probability. □

5. Some examples. The wide applicability of generalized L-statistics is demonstrated by the following examples.

EXAMPLE 5.1. *Simple L-statistics.* For the kernel $h(x) = x$ ($k = 1$, $r_1 = 1$), we obtain a version of the results of Sander (1975).

EXAMPLE 5.2. *Hodges–Lehmann estimator.* For the kernel given in (2.1) $\tilde{g}(x) = x$, and $\hat{J}_N(s) = (a_1 - a_0)^{-1} I_{(a_0, a_1]}(s)$, where $a_0 = \inf\{p\colon \hat{V}_N^{-1}(p) = \hat{V}_N^{-1}(0.5)\}$, $a_1 = \sup\{p\colon \hat{V}_N^{-1}(p) = \hat{V}_N^{-1}(0.5)\}$, relation (4.2) gives $\hat{V}_N^{-1}(0.5)$. Note that the asymptotic distribution of $\hat{V}_N^{-1}(0.5)$ may also be obtained from Corollary 3.2.

EXAMPLE 5.3. *Modified Hodges–Lehmann estimators.* In the spirit of Serfling and Thornton (1982) we may consider modifications of the Hodges–Lehmann estimator for the shift in the two sample problem corresponding to kernels

$$h_{2;r,r} = h(x_1, \ldots, x_r; y_1, \ldots, y_r) = \frac{x_1 + \cdots + x_r}{r} - \frac{y_1 + \cdots + y_r}{r}.$$

Thus in the uncensored case and for samples $X_1, \ldots, X_{N_1}, Y_1, \ldots, Y_{N_2}$ one estimates the shift by

$$\operatorname{med}\left\{\frac{X_{i_1} + \cdots + X_{i_r}}{r} - \frac{Y_{j_1} + \cdots + Y_{j_r}}{r};\ i_1, \ldots, i_r = 1, \ldots, N_1,\ j_1, \ldots, j_r = 1, \ldots, N_2\right\}.$$

For $r = 1$ we have the usual Hodges–Lehmann estimator.

EXAMPLE 5.4. *A measure of spread.* Bickel and Lehmann (1979) propose as a measure of spread the quantity $\text{med}|X_i - X_j|$. We can extend this measure of spread to the censored data case by taking $h(x_1, x_2) = |x_1 - x_2|$ ($k = 1$, $r_1 = 2$), and $\tilde{g}$, $\hat{J}_N$ as in Example 5.2. Again its asymptotic distribution may also be derived from Corollary 3.2. Moreover by Theorem 4.1 [or the corresponding result of Serfling (1984) for the uncensored case] we may study, as a measure of spread, any other linear combination of the quantiles of $\hat{V}_N$.

EXAMPLE 5.5. *Another measure of spread.* It is well known that the sample variance is the U-statistic corresponding to the kernel $h(x_1, x_2) = (x_1 - x_2)^2$ [cf. Serfling (1980), page 173]. In the spirit of the present paper we may consider, as an alternative measure of spread, the quantity $\text{med}\{(X_i - X_j)^2\}$. Thus, with the above kernel ($k = 1$, $r_1 = 2$), and $\tilde{g}$, $\hat{J}_N$ as in Example 5.2, Theorem 4.1 or Corollary 3.2 will give the asymptotic distribution of this quantity. Again, by Theorem 4.1, we may study the asymptotic distribution of any other linear combination of the quantiles of $\hat{V}_N$.

EXAMPLE 5.6. *A measure of association.* It is easy to see that the sample covariance is the U-statistic corresponding to the kernel $h(x_1, y_1, x_2, y_2) = (x_1 - y_1)(x_2 - y_2)$; again we may form the V-empirical process and consider instead some combination of its quantiles. Here, however, it is the bivariate Kaplan–Meier estimator that is required and until recently the available results were inconclusive [see Campbell and Földes (1980)].

EXAMPLE 5.7. *Two-way ANOVA.* Consider for simplicity the noninteraction model $X_{ijm} = \mu + \alpha_i + \beta_j + e_{ijm}$, $m = 1, \dots, N_{ij}$, $i = 1, \dots, R$, $j = 1, \dots, C$, $\Sigma\alpha_i = 0$, $\Sigma\beta_j = 0$, and assume that R, C remain fixed while N_{ij} tend to ∞. Hall (1982) proposes a method for estimating the parameters that fits our formulation. For each choice of one observation per cell (there are $N_{11} \dots N_{RC}$ such choices) compute the average of the observations and take as an estimate $\hat{\mu}$ of μ the median of these averages. Computing, for each choice of one observation per cell again the average of the ith row minus the total average we obtain, for $i = 1, \dots, R - 1$, an estimate $\hat{\alpha}_i$ of α_i by taking the median of these differences; for $\hat{\alpha}_R$ take $-\Sigma_{i=1}^{R-1}\hat{\alpha}_i$. Estimates $\hat{\beta}_j$ of β_j, $j = 1, \dots, C$ are obtained similarly. With $k = RC$ it is easily seen that the estimators $\hat{\mu}$, $\hat{\alpha}_i$, $\hat{\beta}_j$ are the medians of the V-empirical processes corresponding to kernels

$$h^{\mu}_{k;1,\dots,1} \equiv h^{\mu}_{k;1,\dots,1}(x_{11}; \dots; x_{1C}; \dots; x_{R1}; \dots; x_{RC}) = k^{-1}\sum x_{ij},$$

$$h^{\alpha_i}_{k;1,\dots,1} = C^{-1}\sum_{j=1}^{C} x_{ij} - h^{\mu}_{k;1,\dots,1} \quad \text{and} \quad h^{\beta_j}_{k;1,\dots,1} = R^{-1}\sum_{i=1}^{R} x_{ij} - h^{\mu}_{k;1,\dots,1},$$

respectively. The extension of these estimators to the censored data case is, in the spirit of the present paper, straightforward.

APPENDIX

Linear bounds for the Kaplan–Meier estimator and for $\hat{V}_N$. The purpose of this appendix is to establish linear bounds for $\hat{V}_N$ (Theorem A.2) which are needed in the proof of Theorem 4.1. This, however, requires linear bounds for the Kaplan–Meier estimator near the origin. This result (Theorem A.1) extends to the censored data case the corresponding result of Shorack (1972) for the usual empirical distribution function. Linear bounds for the upper tail of the Kaplan–Meier estimator have been established in Gill (1980) but the corresponding bounds near the origin remained an open problem (Gill, 1980, page 40).

It should be mentioned that the proof of Theorem A.1 is due to a referee; the original proof of this result (Akritas, 1983) is based on a different argument that yields a (much) lengthier proof. The statement and proof of Lemma A.1 below, however, are contained in the original proof.

For the statement and proof of Lemma A.1 and Theorem A.1 we will let $\hat{F}$ denote the Kaplan–Meier estimator based on a sample $(X_1, \delta_1), \dots, (X_n, \delta_n)$ generated from a "survival" distribution F and a "censoring" distribution G; also we will let $H_0(t) = \int_0^t (1 - G_-)\, dF$, $\hat{H}_1$ denote the empirical c.d.f. based on the random number m ($m = \sum_1^n \delta_i$) of observations from $H_1 = H_0/H_0(\infty)$, and $\hat{H}_0 = m/n\hat{H}_1$.

LEMMA A.1. *In the notation above we have*

$$\hat{H}_0 \le \hat{F} \le \hat{H}_1 \quad \textit{almost surely}. \tag{A.1}$$

PROOF. Let S_n denote the largest uncensored observation and $X_{(n)} = \max\{X_1, \dots, X_n\}$. We will first show the right inequality in (A.1).

CASE 1. $S_n = X_{(n)}$. Note that both $\hat{F}$ and $\hat{H}_1$ assign positive mass only on the uncensored observations and that the jumps of $\hat{F}$ are increasing (that is, if $X_i < X_j$ are both uncensored, the mass that $\hat{F}$ assigns to X_j is greater than or equal to the mass it assigns to X_i). Next it is easy to see that the mass that $\hat{F}$ assigns to S_1 (= smallest uncensored observation) is always less than or equal to the mass that $\hat{H}_1$ assigns to S_1. This means that $\hat{F}(X_i) = \hat{H}_1(X_i)$ can happen only when $X_i = S_n$ or when it happens that the smallest $n - m$ observations are all censored and the m largest are the uncensored observations in which case $\hat{F} \equiv \hat{H}_1$. In all other cases $\hat{F} < \hat{H}_1$.

CASE 2. $S_n < X_{(n)}$. Note first that $\hat{F}(S_n) < 1 = \hat{H}_1(S_n)$. If we now relabel the largest observation as uncensored and S_n as censored, the new Kaplan–Meier estimator will assign the same mass as $\hat{F}$ to all uncensored X's that are less than S_n. Thus by Case 1 $\hat{F} < \hat{H}_1$ on $[0, S_n)$ and thus the proof of the right inequality in relation (A.1) is complete. The left inequality in (A.1) follows easily by noting that $\hat{H}_1/\hat{F}$ is largest (when we interpret $0/0$ as 0) at S_1 and in particular when

S_1 equals $\min\{X_1, \ldots, X_n\}$ in which case $\hat{F} = m/n\hat{H}_1$. This completes the proof of the lemma. □

THEOREM A.1. *Given $\varepsilon > 0$ there exists $\beta = \beta(\varepsilon)$ so that*

$$P\left(\left\|\frac{\hat{F}}{F}\right\|_0^{S_n} \geq \beta^{-1}\right) \leq \varepsilon \tag{A.2}$$

and

$$P\left(\left\|\frac{F}{\hat{F}}\right\|_{S_1}^{S_n} \geq \beta^{-1}\right) \leq \varepsilon, \tag{A.3}$$

where S_n (S_1) is the largest (smallest) uncensored observation.

PROOF. From the definitions of H_0 and H_1 (given right before Lemma A.1) we have

$$[1 - G(t_0 -)]F \leq H_0 \text{ on } [0, t_0], \quad H_0(\infty)H_1 \leq F \text{ on } [0, \infty).$$

Thus, using Lemma A.1 we have $\|\hat{F}/F\|_0^{S_n} \leq \|\hat{H}_1/(H_0(\infty)H_1)\|_0^{S_n}$ which (conditionally on m and hence unconditionally too) is $O_p(1)$ uniformly in n, giving relation (A.2). Similarly, considering only the interval $[0, t_0]$ which is easily shown to be sufficient, $\|F/\hat{F}\|_{S_1}^{t_0} \leq \|(H_0/[1 - G(t_0 -)])/\hat{H}_0\| = O_p(1)$ giving relation (A.3). □

We are now ready to present the linear bounds for $\hat{V}_N$.

THEOREM A.2. *Let Assumption* 3.1 *hold with $M_h = 1$ (see the remark following the proof). Then for any $\varepsilon > 0$ there exists a $\beta = \beta(\varepsilon) > 0$ such that*

$$P[1 - \hat{V}_N \leq \beta^{-1}(1 - V) \text{ on } (-\infty, \hat{C}_1]] \geq 1 - \varepsilon, \tag{A.4}$$

$$P[1 - \hat{V}_N \geq \beta(1 - V) \text{ on } (-\infty, \hat{C}_1]] \geq 1 - \varepsilon, \tag{A.5}$$

$$P[\hat{V}_N \leq \beta^{-1}V \text{ on } (-\infty, \infty)] \geq 1 - \varepsilon, \tag{A.6}$$

and

$$P[\hat{V}_N \geq \beta V \text{ on } [\hat{C}_0, \infty)] \geq 1 - \varepsilon \tag{A.7}$$

hold for all N, where $\hat{C}_0, \hat{C}_1$ are given in (4.3).

PROOF. We will show only relation (A.4); the other relations are established similarly. In what follows $\prod^*$ will denote the product $\prod_{(s_1, j)}$ with $(s_1, j) \neq (s, i)$. Note that if $V(\infty) < 1$ the result holds trivially so we will assume $V(\infty) = 1$ where V is defined in (3.3) with $T_s = \min(T_{F_s}, T_{G_s}) \equiv T_{F_s}$, $s = 1, \ldots, k$ (see Remark 3.2); also we will set $\hat{T}_s$ to be the maximum of the uncensored observations.

We have

$$\beta^{-1}(1 - V(t)) - (1 - \hat{V}_N(t)) = \beta^{-1}\int_0^{T_1} \cdots \int_0^{T_k}(1 - I[h_{k;r} \le t]) \prod_{(s,i)} dF_s(x_{si})$$
$$- \int_0^{T_1} \cdots \int_0^{T_k}(1 - I[h_{k;r} \le t]) \prod_{(s,i)} d\hat{F}_s(x_{si}) - 1 + \prod_{(s,i)} \hat{F}_s(\hat{T}_s)$$
$$= \int_0^{T_1} \cdots \int_0^{T_k}(1 - I[h_{k;r} \le t])\left[\prod_{(s,i)} \beta^{-d} dF_s(x_{si}) - \prod_{(s,i)} d\hat{F}_s(x_{si})\right]$$
$$-1 + \prod_{(s,i)} \hat{F}_s(\hat{T}_s),$$

where $d = [\sum_{s=1}^k r_s]^{-1}$. Thus, using the formula for the difference of products given in the proof of Theorem 3.1 we get

$$\beta^{-1}(1 - V(t)) - (1 - \hat{V}_N(t))$$
$$= \sum_{(s,i)} \int_0^{T_1} \cdots \int_0^{T_k}(1 - I[h_{k;r} \le t]) \prod{}^{*} d\hat{F}_{s_1}(x_{s_1 j})$$
$$\times d\left[\beta^{-d}F_s(x_{si}) - \hat{F}_s(x_{si})\right] \tag{A.8}$$
$$+ \text{ integrals involving } d\left[\beta^{-d}F_s(x_{si}) - \hat{F}_s(x_{si})\right] \text{ in at least second order}$$
$$+ \text{ terms of order } \left(1 - F_s(\hat{T}_s)\right).$$

But by Assumption 3.1 with $M_h = 1$, each of the terms in the sum on the right-hand side of (A.8) will be either of the form

$$\int_0^{T_1} \cdots \int_0^{T_k}[\beta^{-d}F_s(y) - \hat{F}_s(y)]\Pi^* d\hat{F}_{s_1}(x_{s_1 j})$$

or of the form

$$\int_0^{T_1} \cdots \int_0^{T_k}\{[\beta^{-d}F_s(T_s) - \hat{F}_s(T_s)] - [\beta^{-d}F_s(y) - \hat{F}_s(y)]\}\Pi^* d\hat{F}_{s_1}(x_{s_1 j}),$$

where y depends on $x_{s_1 j}$, $(s_1, j) \neq (s, i)$ and on t. But Theorem A.1 and Theorem 3.2.1 in Gill (1980) imply that both forms of integrals above are positive with high probability. Integrals involving $d[\beta^{-d}F_s(x_{si}) - \hat{F}_s(x_{si})]$ in at least second order may also be shown to be positive with a high probability. Finally there are negative terms of order $1 - \hat{F}_s(\hat{T}_s)$; these, however, converge to zero at least as fast as the positive terms and thus β can be chosen so the whole expression is positive with high probability.

REMARK A.1. The requirement that in Assumption 3.1 $M_h = 1$ is somewhat restrictive. It is clear from the proof of Theorem A.2 that without this requirement one would have to use linear bounds for the Kaplan–Meier estimator indexed by intervals. In the general case such results are not available (and indeed not true) even for the usual empirical distribution function. However, Theorem A.2 may be proved if instead of $M_h = 1$ one requires that there exists a $\delta > 0$ such that all the intervals in each of the partitions described in Assumption 3.1 are of length greater than δ almost surely $[P_s]$, $s = 1, \ldots, k$.

REFERENCES

AKRITAS, M. G. (1983). Linear bounds for the Kaplan–Meier estimator. Unpublished manuscript.

BICKEL, P. J. and LEHMANN, E. L. (1979). Descriptive statistics for nonparametric models IV. Spread. In *Contributions to Statistics* (J. Hájek Memorial Volume) (J. Jurečková, ed.).

CAMPBELL, G. and FÖLDES, A. (1980). Large-sample properties of nonparametric bivariate estimators with censored data. Technical Report 80-10, Dept. of Statistics, Purdue Univ.

EFRON, B. (1967). The two sample problem with censored data. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **4** 831–853. Univ. California Press.

GILL, R. D. (1980). *Censoring and Stochastic Integrals.* Mathematical Centre Tracts 124, Mathematisch Centrum, Amsterdam.

GILL, R. D. (1983). Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.* **11** 49–58.

HALL, W. J. (1982). Personal communication.

HARDY, G. H. (1952). *A Course of Pure Mathematics.* 10th ed. Cambridge University Press.

HODGES, J. L., JR. and LEHMANN, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34** 598–611.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.

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

NATANSON, I. P. (1961). *Theory of Functions of a Real Variable* **1**. Ungar, New York.

PADGETT, W. J. and WEI, L. J. (1982). Estimation of the ratio of scale parameters in the two sample problem with arbitrary right censorship. *Biometrika* **69** 252–256.

PRATT, J. W. (1960). On interchanging limits and integrals. *Ann. Math. Statist.* **31** 74–77.

SANDER, J. M. (1975). Asymptotic normality of linear combinations of functions of order statistics with censored data. Technical Report 8, Division of Biostatistics, Stanford Univ.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

SERFLING, R. J. (1984). Generalized L-, M- and R-statistics. *Ann. Statist.* **12** 76–86.

SERFLING, R. J. and THORNTON, D. H. (1982). An extension of Bahadur's representation of sample quantiles, with applications to versions of the Hodges–Lehmann location estimator. Technical Report 352, Johns Hopkins Univ.

SHORACK, G. R. (1972). Functions of order statistics. *Ann. Math. Statist.* **43** 412–427.

SILVERMAN, B. W. (1976). Limit theorems for dissociated random variables. *Adv. in Appl. Probab.* **8** 806–819.

SILVERMAN, B. W. (1983). Convergence of a class of empirical distribution functions of dependent random variables. *Ann. Probab.* **11** 745–751.

VERVAAT, W. (1972). Functional central limit theorems for processes with positive drift and their inverses. *Z. Wahrsch. verw. Gebiete* **23** 245–253.

WEI, L. J. and GAIL, M. H. (1983). Nonparametric estimation for a scale-change model with censored observations. *J. Amer. Statist. Assoc.* **78** 382–388.

DEPARTMENT OF STATISTICS
PENNSYLVANIA STATE UNIVERSITY
UNIVERSITY PARK, PENNSYLVANIA 16802