The Annals of Statistics
1991, Vol. 19, No. 3, 1457–1470

ALMOST SURE ASYMPTOTIC REPRESENTATION FOR A CLASS OF FUNCTIONALS OF THE KAPLAN–MEIER ESTIMATOR

BY IRÈNE GIJBELS AND NOËL VERAVERBEKE

Limburgs Universitair Centrum

This paper deals with censored data estimation of a general class of von Mises-type functionals of the survival time distribution F. Conditions are given under which an almost sure asymptotic representation holds for the estimator, obtained by applying the same functional to $\hat{F}_n$, the product-limit estimator of Kaplan and Meier.

1. Introduction. Let $X_1, \dots, X_n$ be independent and identically distributed (i.i.d.) nonnegative random variables with common continuous distribution function F, called the survival time distribution. Our model is that of right random censoring, that is, associated with each X_i there is an independent nonnegative censoring time Y_i and $Y_1, \dots, Y_n$ are assumed to be i.i.d. random variables with continuous distribution function G. The observations in this model are the pairs (T_i, δ_i), where $T_i = \min(X_i, Y_i)$ and $\delta_i = I(X_i \le Y_i)$, $i = 1, \dots, n$. Clearly, the T_i are i.i.d. with continuous distribution function $H = 1 - (1 - F)(1 - G)$.

For estimation of F, based on (T_i, δ_i), $i = 1, \dots, n$, Kaplan and Meier (1958) suggested the so-called product-limit estimator $\hat{F}_n$ defined by

$$\hat{F}_n(t) = \begin{cases} 1 - \prod_{T_{(i)} \le t} \left(\dfrac{n-i}{n-i+1} \right)^{\delta_{(i)}}, & \text{if } t < T_{(n)}, \\ 1, & \text{if } t \ge T_{(n)} \text{ and } \delta_{(n)} = 1, \\ \text{undefined}, & \text{if } t \ge T_{(n)} \text{ and } \delta_{(n)} = 0, \end{cases}$$

where $T_{(1)} \le T_{(2)} \le \cdots \le T_{(n)}$ are the order statistics of the T_i and $\delta_{(1)}, \dots, \delta_{(n)}$ are the corresponding δ_i.

For any distribution function L, we use as a notation: $T_L = \inf\{t: L(t) = 1\}$. Note that $T_H = \min(T_F, T_G)$.

Many characteristics of the survival time distribution F may be expressed as

$$(1.1) \qquad \int \cdots \int h_t(x_1, \dots, x_m) \prod_{i=1}^{m} dF(x_i),$$

where $h_t(x_1, \dots, x_m)$ is some m-variate function, symmetric in its $m \ge 1$

Received May 1989; revised June 1990.
AMS 1980 *subject classifications*. Primary 62G05; secondary 60F15.
Key words and phrases. Censored data, almost sure representation, V-statistics, asymptotic normality, law of iterated logarithm, quantiles.

variables and possibly depending on some real parameter t. Some examples are: (i) $m = 1$, $h_t(x) = x$ gives the mean of F; (ii) $m = 2$, $h_t(x_1, x_2) = (x_1 - x_2)^2/2$ gives the variance of F; (iii) $h_t(x_1, \dots, x_m) = I(\tilde{h}(x_1, \dots, x_m) \le t)$, $t \in \mathbb{R}$, gives the distribution function at t of the function $\tilde{h}(X_1, \dots, X_m)$.

If no censoring is present, then a natural way to obtain estimators is to replace F in (1.1) by the empirical distribution function. This then leads to the well-developed theory of V-statistics and U-statistics, see, for example, Serfling (1980).

Applying the same idea in the censored data case leads to substitution of the Kaplan–Meier estimator $\hat{F}_n$ in (1.1). The resulting estimator has a more complicated structure because the simple sum structure of the empirical distribution function is lost by using the Kaplan–Meier estimator. Our approach to this problem is to use an asymptotic representation result for the Kaplan–Meier estimator in which $\hat{F}_n(t)$ is decomposed into a leading term which is an average of i.i.d. random variables and a lower order term $r_n(t)$ (see Lemma 2). But, without further assumptions, this asymptotic representation only holds for $t < T_H$ and the order of a.s. convergence of the remainder term $r_n(t)$ is uniform for $t \in [0, T]$, where $T < T_H$. This leads us to considering the functional

$$\theta_t(F) = \int_0^T \cdots \int_0^T h_t(x_1, \dots, x_m) \prod_{i=1}^m dF(x_i),$$

where the upper bound T in the integrals satisfies $T < T_H$ and which is a restricted version of (1.1). The corresponding estimator is then given by $\theta_t(\hat{F}_n)$, where $\hat{F}_n$ is the Kaplan–Meier estimator. Thus

$$\theta_t(\hat{F}_n) = \int_0^T \cdots \int_0^T h_t(x_1, \dots, x_m) \prod_{i=1}^m d\hat{F}_n(x_i).$$

The truncation of the functional at a fixed time T opens the way to obtain results for $\theta_t(\hat{F}_n) - \theta_t(F)$ from the known ones about $\hat{F}_n(t) - F(t)$ for $t \in [0, T]$. This route requires some careful steps so as to end up with weak conditions on the kernel function h_t. From the practical point of view, the truncation is not so desirable. With for example $m = 1$ and $h_t(x) = x$, this means that, instead of the mean survival time, we obtain the quantity $\int_0^T x\, dF(x)$. This functional is related in a simple way to $\int_0^T (1 - F(x))\, dx$, the mean survival time over $[0, T]$. This difficulty is well known, see, for example, Sander (1975), Reid (1981), Akritas (1986). Efforts to avoid this unpleasant restriction have been done in, for example, Susarla and Van Ryzin (1980), Gill (1983). It should be noted however that very often the truncated functional θ_t is not too different from the one in (1.1). For instance in the important case where $h_t(x_1, \dots, x_m) = I(\tilde{h}(x_1, \dots, x_m) \le t)$, we have that $\theta_t(F) = P(\tilde{h}(X_1, \dots, X_m) \le t,\ X_1 \le T, \dots, X_m \le T)$ and for many functions $\tilde{h}$ this equals $P(\tilde{h}(X_1, \dots, X_m) \le t)$ for t in some large interval. Some examples are: (i) if $m = 1$ and $\tilde{h}(x) = x$, then $\theta_t(F) = F(t)$ for $t \le T$; (ii) if $m = 2$ and $\tilde{h}(x_1, x_2) = (x_1 + x_2)/2$, then $\theta_t(F) = P((X_1 + X_2)/2 \le t)$ for $t \le T/2$. Hence, the restriction we have to impose on the functional (for mathematical reasons) still allows to estimate the quantity in (1.1), but for a somewhat restricted range for the parameter t. This also means that the results in the last section allow to study pth quantiles of X, but also of, for example, $(X_1 + X_2)/2, \dots$ in a restricted range for p (for instance only small and moderate p).

The main result of this paper is to prove that, under general conditions on the kernel h_t, $\theta_t(\hat{F}_n) - \theta_t(F)$ can be written as an ordinary V-statistic, based on the i.i.d. observations (T_i, δ_i), plus a remainder term which is almost surely of the order $O(n^{-1} \log n)$, uniformly in t. It follows from such a representation that many of the standard limit theorems for ordinary V-statistics continue to hold for $\theta_t(\hat{F}_n)$, after suitable normalization.

The paper is organized as follows. After preliminaries in Section 2, the main result of the paper, the almost sure asymptotic representation for $\theta_t(\hat{F}_n)$, is presented as Theorem 1 in Section 3. Asymptotic normality and almost sure behaviour of $\theta_t(\hat{F}_n)$ are discussed in Section 4. Finally, in Section 5, we concentrate on the special choice $h_t(x_1, \dots, x_m) = I(\tilde{h}(x_1, \dots, x_m) \le t)$, which leads to additional results for the corresponding quantile estimator. The almost sure representations obtained in this paper go beyond the results in two related papers: Reid (1981) and Akritas (1986). Reid (1981) considered related estimators via the influence curve approach. In the special case of indicator kernels, Akritas (1986) obtained weak convergence for the resulting process and the corresponding quantile process.

2. Preliminaries. In this section we present some preliminaries which will be required in the proof of the main result.

We first introduce some notation. For a function g of one real variable, the total variation on an interval $[a, b]$ will be denoted by $TV_{[a,b]}g$. If g is a function of p real variables, then $TV_{[a,b]}g(\cdot, y_2, \dots, y_p)$ stands for the total variation of the univariate function $g(\cdot, y_2, \dots, y_p)$ with $y_2, \dots, y_p$ fixed. We will write $g(y-, y_2, \dots, y_p)$ and $g(y+, y_2, \dots, y_p)$ for left- and right-hand limits of this function at y. Integrals of the form $\int_a^b$ are meant as $\int_{[a,b]}$ throughout and integration over other types of intervals will be mentioned explicitly. For $k = 1, \dots, m$, we will use the abbreviation $\mathbf{x}_{(k)} = (x_k, x_{k+1}, \dots, x_m)$ and writing $\mathbf{x}_{(k)} \in [a, b]$ will mean that $x_j \in [a, b]$ for $j = k, \dots, m$.

An important tool in the proofs of our results will be an integration by parts formula for integrals of the form $\int_0^T h_t(x_1, \mathbf{x}_{(2)})\, dK(x_1)$, where K is some function of bounded variation. Having in mind the applications in which h_t is of the indicator form, we would like to perform this partial integration under the weakest possible conditions on $h_t(\cdot, \mathbf{x}_{(2)})$. One of the conditions will be that for all t and for all $\mathbf{x}_{(2)} \in [0, T]$, $h_t(\cdot, \mathbf{x}_{(2)})$ is continuous, except at some finite number of points, say $J(t)$, which is independent of $\mathbf{x}_{(2)}$. We denote these ordered (from smallest to largest) discontinuity points by $k^{(j)}(\mathbf{x}_{(2)}, t)$, $j = 1, \dots, J(t)$. Further, we define for $j = 1, \dots, J(t)$,

$$\tilde{k}^{(j)} = \tilde{k}^{(j)}(\mathbf{x}_{(2)}, t) = \min\Big(T, \max\Big(0, k^{(j)}(\mathbf{x}_{(2)}, t)\Big)\Big)$$

and also $\tilde{k}^{(0)} = 0$ and $\tilde{k}^{(J(t)+1)} = T$.

LEMMA 1. *If* (i) *K is of bounded variation on* $[0, T]$ *(ii) for all t and for all* $\mathbf{x}_{(2)} \in [0, T]$: $h_t(\cdot, \mathbf{x}_{(2)})$ *is of bounded variation on* $[0, T]$ *and is continuous except at a finite number* $J(t)$ *of points, then*

$$\begin{aligned}
&\int_{[0,T]} h_t(x_1, \mathbf{x}_{(2)})\, dK(x_1) \\
&\quad = [K(0+) - K(0-)]h_t(0, \mathbf{x}_{(2)}) + [K(T+) - K(T-)]h_t(T, \mathbf{x}_{(2)}) \\
&\qquad + \sum_{j=1}^{J(t)} I(0 < \tilde{k}^{(j)} < T) h_t\big(\tilde{k}^{(j)}, \mathbf{x}_{(2)}\big)\big[K(\tilde{k}^{(j)}+) - K(\tilde{k}^{(j)}-)\big] \\
&\qquad + \sum_{j=0}^{J(t)} I(\tilde{k}^{(j)} < \tilde{k}^{(j+1)})\Big[h_t\big(\tilde{k}^{(j+1)}-, \mathbf{x}_{(2)}\big)K(\tilde{k}^{(j+1)}-) \\
&\qquad\qquad - h_t\big(\tilde{k}^{(j)}+, \mathbf{x}_{(2)}\big)K(\tilde{k}^{(j)}+)\Big] \\
&\qquad - \sum_{j=0}^{J(t)} \int_{(\tilde{k}^{(j)}, \tilde{k}^{(j+1)})} K(x_1)\, d_{x_1} h_t(x_1, \mathbf{x}_{(2)}).
\end{aligned} \tag{2.1}$$

PROOF. The proof is immediate by applying the following integration by parts formula for Lebesgue–Stieltjes integrals: If f and g are of bounded variation on $[a, b]$ $(a < b)$ and if f is continuous on (a, b), then

$$\int_{(a,b)} f(x)\, dg(x) = f(b-)g(b-) - f(a+)g(a+) - \int_{(a,b)} g(x)\, df(x)$$

[see also Hewitt and Stromberg (1975), page 419]. □

Another important tool for establishing our results is an asymptotic representation for the product-limit estimator $\hat{F}_n(t)$, due to Lo and Singh (1986). In the next lemma we formulate this result but for the remainder term we state the order which was recently obtained by Major and Rejtő (1988).

LEMMA 2. *If F is continuous, then for* $t < T_H$,

$$\hat{F}_n(t) = F(t) + \frac{1}{n}\sum_{i=1}^{n} \psi_i(t) + r_n(t), \tag{2.2}$$

where the $\psi_i(t)$ *are i.i.d. zero mean random variables which are bounded uniformly in* $0 \le t \le T$, *with* $T < T_H$ *and where*

$$\sup_{t \in [0,T]} |r_n(t)| = O(n^{-1}\log n) \quad a.s. \tag{2.3}$$

Moreover,

$$\psi_i(t) = (1 - F(t))\left[\int_0^t \frac{I(T_i \le y) - H(y)}{(1 - H(y))^2}\, dH^u(y) + \frac{I(T_i \le t, \delta_i = 1) - H^u(t)}{1 - H(t)} - \int_0^t \frac{I(T_i \le y, \delta_i = 1) - H^u(y)}{(1 - H(y))^2}\, dH(y)\right]$$

and

$$\mathrm{Cov}(\psi_i(t), \psi_i(s)) = (1 - F(t))(1 - F(s))\gamma(\min(t, s)),$$

where H^u is the subdistribution function of the uncensored observations given by $H^u(t) = P(T_1 \le t, \delta_1 = 1)$ and

$$\gamma(t) = \int_0^t \frac{dH^u(y)}{(1 - H(y))^2}.$$

We conclude this preliminary section with some technical facts about the functions ψ_i and r_n appearing in Lemma 2.

First, for each $\omega \in \Omega$, the functions $\psi_i(\cdot, \omega)$, $i = 1, \dots, n$, are right-continuous and of bounded variation on $[0, T]$. Indeed, one easily calculates, using basic properties concerning total variation of sums and products [see, e.g., Apostol (1982), page 130], that for each $i = 1, \dots, n$ and $\omega \in \Omega$:

$$TV_{[0,T]}\psi_i(\cdot, \omega) \le C_1, \tag{2.4}$$

where

$$C_1 = 2\left[\int_0^T \frac{1 + H(y)}{(1 - H(y))^2}\, dH^u(y) + \frac{1 + H(T) + H^u(T)}{1 - H(T)} + \int_0^T \frac{1 + H^u(y)}{(1 - H(y))^2}\, dH(y)\right].$$

Secondly, for the function r_n defined by (2.2), we also have that $r_n(\cdot, \omega)$ is right-continuous and of bounded variation on $[0, T]$. Indeed, from (2.2), $TV_{[0,T]}r_n(\cdot, \omega) \le TV_{[0,T]}\hat{F}_n(\cdot, \omega) + TV_{[0,T]}F + n^{-1}\sum_{i=1}^n TV_{[0,T]}\psi_i(\cdot, \omega) \le 2 + C_1$. Hence, for each $\omega \in \Omega$,

$$TV_{[0,T]}r_n(\cdot, \omega) \le C_2 = 2 + C_1. \tag{2.5}$$

3. Almost sure asymptotic representation. In this section we show that, up to a remainder term, $\theta_t(\hat{F}_n) - \theta_t(F)$ can be represented as a V-statistic with kernel of degree m, based on the bivariate i.i.d. observations (T_i, δ_i), $i = 1, \dots, n$. The almost sure bound of the remainder term is established under minimal conditions on h.

The following conditions will be considered.

(A1) $$\sup_t \sup_{\mathbf{x}_{(1)} \in [0,T]} \left| h_t(\mathbf{x}_{(1)}) \right| = M_0 < \infty.$$

(A2) $$\sup_t \sup_{\mathbf{x}_{(2)} \in [0,T]} \left(TV_{[0,T]} h_t(\cdot, \mathbf{x}_{(2)}) \right) = M_1 < \infty.$$

(A3) For all t and all $\mathbf{x}_{(2)} \in [0,T]$: $h_t(\cdot, \mathbf{x}_{(2)})$ is continuous except at some finite number of points, say $J(t)$, which is independent of $\mathbf{x}_{(2)}$. Moreover, $\sup_t J(t) = M_2 < \infty$.

It can be noticed from the proof that the previous conditions are slightly too strong but are put in this form for ease of presentation. Slight extensions could be obtained by replacing boundedness conditions by certain integrability conditions.

We consider some examples of kernels which satisfy our conditions.

EXAMPLE 1 (Mean). Take $h_t(x_1, \dots, x_m) = (1/m)(x_1 + \cdots + x_m)$. Then (A1) and (A3) are trivially fulfilled. Also (A2) holds, since $TV_{[0,T]} h_t(\cdot, \mathbf{x}_{(2)}) = T/m$.

EXAMPLE 2 (Variance). Take $m = 2$ and $h_t(x_1, x_2) = (x_1 - x_2)^2/2$. Again (A1) and (A3) are trivial and for (A2), note that $TV_{[0,T]} h_t(\cdot, x_2) = x_2^2 - x_2 T + T^2/2$.

EXAMPLE 3 (Gini's mean difference). Take $m = 2$ and $h_t(x_1, x_2) = |x_1 - x_2|$. Here $TV_{[0,T]} h_t(\cdot, x_2) = T$.

EXAMPLE 4 (Indicator function). Here

$$h_t(x_1, \dots, x_m) = I\left(\tilde{h}(x_1, \dots, x_m) \le t\right), \tag{3.1}$$

where $\tilde{h}(x_1, \dots, x_m)$ is some function, symmetric in its arguments and satisfying the following condition (B) [which is similar to Assumption 3.1 in Akritas (1986)]:

(B) There is a partition of the real line such that the function $\tilde{h}(\cdot, \mathbf{x}_{(2)})$ is continuous and monotonic within each interval of the partition. Moreover, the number of intervals in this partition, say $M_{\tilde{h}}$, is finite and independent of $\mathbf{x}_{(2)}$.

It is easily seen that under the assumption (B) on $\tilde{h}(x_1, \dots, x_m)$, the kernel $h_t(x_1, \dots, x_m)$ in (3.1) will satisfy (A1)–(A3), with M_1 and M_2 bounded by $2M_{\tilde{h}}$. Some examples of functions $\tilde{h}$ are (i) $\tilde{h}(x_1, \dots, x_m) = (x_1 + \cdots + x_m)/m$. Here $M_{\tilde{h}} = 1$, $J(t) = 1$; (ii) $\tilde{h}(x_1, x_2) = (x_1 - x_2)^2/2$. Here $M_{\tilde{h}} = 2$, $J(t) = 0$ if $t < 0$, 1 if $t = 0$, 2 if $t > 0$; (iii) $\tilde{h}(x_1, x_2) = |x_1 - x_2|$. Here $M_{\tilde{h}}$ and $J(t)$ are as in (ii).

Our main result is now formulated in the next theorem. It provides an almost sure asymptotic representation for $\theta_t(\hat{F}_n) - \theta_t(F)$ as a V-statistic plus a remainder term. (It should be noticed that we will use sum and product signs with the usual conventions concerning empty sums and empty products).

THEOREM 1. *Assume conditions* (A1)–(A3). *Then*

$$\theta_t(\hat{F}_n) = \theta_t(F) + V_n(t) + R_n(t),$$

where $V_n(t)$ *is a V-statistic*

$$V_n(t) = n^{-m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} k_t\big((T_{i_1}, \delta_{i_1}), \dots, (T_{i_m}, \delta_{i_m})\big)$$

with kernel k_t *defined by*

$$\begin{aligned} &k_t((T_1, \delta_1), \dots, (T_m, \delta_m)) \\ &\quad = \sum_{k=1}^{m} \sum_{1 \le j_1 < \cdots < j_k \le m} \int_0^T \cdots \int_0^T h_t(x_1, \dots, x_m) \\ &\qquad \times \prod_{i=1}^{k} d\psi_{j_i}(x_i) \prod_{j=k+1}^{m} dF(x_j) \end{aligned} \tag{3.2}$$

and where

$$\sup_t |R_n(t)| = O(n^{-1} \log n) \quad a.s.$$

PROOF. Using Lemma 2 we can write

$$\begin{aligned} \theta_t(\hat{F}_n) &= \int_0^T \cdots \int_0^T h_t(x_1, \dots, x_m) \\ &\quad \times \prod_{i=1}^{m} \left(dF(x_i) + \frac{1}{n} \sum_{j=1}^{n} d\psi_j(x_i) + dr_n(x_i) \right). \end{aligned} \tag{3.3}$$

The product in this integral will be expanded using the following general formula:

$$\begin{aligned} &\prod_{i=1}^{m} (a_i + b_i + c_i) \\ &\quad = \prod_{i=1}^{m} a_i + \sum_{k=1}^{m} \sum_{1 \le i_1 < \cdots < i_k \le m} \left[\prod_{\substack{j=1 \\ j \notin \{i_1, \dots, i_k\}}}^{m} a_j \right] b_{i_1} \cdots b_{i_k} \\ &\qquad + \sum_{k=1}^{m} \sum_{l=0}^{m-k} \sum_{1 \le i_1 < \cdots < i_k \le m} \sum_{\substack{1 \le j_1 < \cdots < j_l \le m \\ j_1, \dots, j_l \notin \{i_1, \dots, i_k\}}} \left[\prod_{\substack{q=1 \\ q \notin \{i_1, \dots, i_k, j_1, \dots, j_l\}}}^{m} a_q \right] \\ &\qquad \times b_{j_1} \cdots b_{j_l} c_{i_1} \cdots c_{i_k}. \end{aligned}$$

Applying this formula to (3.3) gives

$$\theta_t(\hat{F}_n) - \theta_t(F)$$

$$= \sum_{k=1}^{m} n^{-k} \sum_{j_1=1}^{n} \cdots \sum_{j_k=1}^{n} \sum_{1 \le i_1 < \cdots < i_k \le m} \int_0^T \cdots \int_0^T h_t(x_1, \dots, x_m)$$

$$\times \left[\prod_{\substack{j=1 \\ j \notin \{i_1, \dots, i_k\}}}^{m} dF(x_j) \right] d\psi_{j_1}(x_{i_1}) \cdots d\psi_{j_k}(x_{i_k})$$

$$(3.4) \quad + \sum_{k=1}^{m} \sum_{l=0}^{m-k} n^{-l} \sum_{q_1=1}^{n} \cdots \sum_{q_l=1}^{n} \sum_{1 \le i_1 < \cdots i_k \le m} \sum_{\substack{1 \le j_1 < \cdots < j_l \le m \\ j_1, \dots, j_l \notin \{i_1, \dots, i_k\}}}$$

$$\int_0^T \cdots \int_0^T h_t(x_1, \dots, x_m) \left[\prod_{\substack{q=1 \\ q \notin \{i_1, \dots, i_k, j_1, \dots, j_l\}}}^{m} dF(x_q) \right]$$

$$\times d\psi_{q_1}(x_{j_1}) \cdots d\psi_{q_l}(x_{j_l})\, dr_n(x_{i_1}) \cdots dr_n(x_{i_k}).$$

Using the symmetry of h_t, the first term on the right-hand side of (3.4) simplifies to

$$\sum_{k=1}^{m} \binom{m}{k} n^{-k} \sum_{j_1=1}^{n} \cdots \sum_{j_k=1}^{n} \int_0^T \cdots \int_0^T h_t(x_1, \dots, x_m)$$

$$\times d\psi_{j_1}(x_1) \cdots d\psi_{j_k}(x_k) \prod_{j=k+1}^{m} dF(x_j)$$

and using the ideas of Hoeffding's decomposition theorem for U-statistics [see, e.g., Serfling (1980), Lemma A, page 178–179] this is also equal to $V_n(t)$ as defined in the statement of the theorem. The symmetry of h_t also simplifies the second term in (3.4) and leads to the decomposition $\theta_t(\hat{F}_n) = \theta_t(F) + V_n(t) + R_n(t)$, where

$$R_n(t) = \sum_{k=1}^{m} \sum_{l=0}^{m-k} \binom{m}{k} \binom{m-k}{l} n^{-l} \sum_{q_1=1}^{n} \cdots \sum_{q_l=1}^{n}$$

$$(3.5) \quad \int_0^T \cdots \int_0^T h_t(x_1, \dots, x_m)\, dr_n(x_1) \cdots dr_n(x_k)$$

$$\times d\psi_{q_1}(x_{k+l}) \cdots d\psi_{q_l}(x_{k+l}) \prod_{q=k+l+1}^{m} dF(x_q).$$

To show that $\sup_t |R_n(t)| = O(n^{-1} \log n)$ a.s., we consider the integral in (3.5) and proceed as follows:

$$\begin{aligned}
&\left|\int_0^T \cdots \int_0^T h_t(x_1,\dots,x_m)\,dr_n(x_1)\cdots\, dr_n(x_k)\right.\\
&\qquad\left.\times d\psi_{q_1}(x_{k+1})\cdots\, d\psi_{q_l}(x_{k+l})\prod_{q=k+l+1}^{m} dF(x_q)\right|\\
(3.6)\qquad &\le \left[TV_{[0,T]}r_n\right]^{k-1}\left[\prod_{j=1}^{l} TV_{[0,T]}\psi_{q_j}\right]\left[TV_{[0,T]}F\right]^{m-k-l}\\
&\qquad\times \sup_t \sup_{\mathbf{x}_{(2)}\in[0,T]}\left|\int_0^T (x_1,\mathbf{x}_{(2)})\,dr_n(x_1)\right|\\
&\le C_2^{k-1}C_1^l \sup_t \sup_{\mathbf{x}_{(2)}\in[0,T]}\left|\int_0^T h_t(x_1,\mathbf{x}_{(2)})\,dr_n(x_1)\right|,
\end{aligned}$$

where C_1 and C_2 are the constants in (2.4) and (2.5). By Lemma 1 and conditions (A1)–(A3), we have for all t and all $\mathbf{x}_{(2)} \in [0,T]$:

$$(3.7)\qquad \left|\int_0^T h_t(x_1,\mathbf{x}_{(2)})\,dr_n(x_1)\right| \le K \sup_{y\in[0,T]} |r_n(y)|,$$

where K is a constant only depending on M_0, M_1, M_2. Hence, from (3.5), (3.6) and (3.7),

$$\begin{aligned}
\sup_t |R_n(t)| &\le K \sup_{y\in[0,T]} |r_n(y)| \sum_{k=1}^{m}\sum_{l=0}^{m-k}\binom{m}{k}\binom{m-k}{l} C_2^{k-1}C_1^l\\
&= O(n^{-1}\log n) \quad \text{a.s.}
\end{aligned}$$

by (2.3). This proves the theorem. □

COROLLARY 1. *Assume conditions* (A1)–(A3). *Then*

$$\theta_t(\hat{F}_n) = \theta_t(F) + U_n(t) + \tilde{R}_n(t),$$

where $U_n(t)$ *is a* U*-statistic*

$$U_n(t) = \binom{n}{m}^{-1} \sum_{1\le i_1<\cdots<i_m\le n} k_t\big((T_{i_1},\delta_{i_1}),\dots,(T_{i_m},\delta_{i_m})\big)$$

and where

$$\sup_t |\tilde{R}_n(t)| = O(n^{-1}\log n) \quad a.s.$$

PROOF. This follows immediately from the well-known relation between U-statistics and V-statistics [Serfling (1980), page 206] and the fact that for all

t and all $1 \le i_1, \dots, i_m \le n$,

$$\left|k_t\big((T_{i_1}, \delta_{i_1}), \dots, (T_{i_m}, \delta_{i_m})\big)\right| \le \sum_{k=1}^{m} \binom{m}{k} M_0 C_1^k. \tag{3.8}$$ □

4. Further asymptotic results. The representation in Theorem 1 opens a way to carry over asymptotic properties of the V-statistic $V_n(t)$ [or, by Corollary 1, also of the U-statistic $U_n(t)$] to the quantity of interest $\theta_t(\hat{F}_n) - \theta_t(F)$.

First note that $E[k_t((T_1, \delta_1), \dots, (T_m, \delta_m))] = 0$. This is easily seen from (3.2) by using Lemma 1 with $K = \psi_{j_1}$ and the fact that $E[\psi_{j_1}(x-)] = E[\psi_{j_1}(x) - 1/(1 - G(x))I(T_{j_1} = x, \delta_{j_1} = 1)] = E[\psi_{j_1}(x)] = 0$ (since F and hence H^u is continuous).

An important quantity in the limiting theory of U-statistics is the conditional expectation of the kernel k_t, denoted by

$$k_1((t_1, d_1); t) = E\big[k_t((T_1, \delta_1), \dots, (T_m, \delta_m)) | (T_1, \delta_1) = (t_1, d_1)\big].$$

Again using Lemma 1, it can be calculated that

$$k_1((T_1, \delta_1); t) = \frac{1}{m} \int_0^T g(x|t)\, d\psi_1(x),$$

where

$$g(x|t) = m \int_0^T \cdots \int_0^T h_t(x, x_2, \dots, x_m) \prod_{j=2}^{m} dF(x_j). \tag{4.1}$$

The processes $\psi_1(x)$ and $B^0(\gamma(x)/(1 + \gamma(x)))(1 - F(x))(1 + \gamma(x))$, with B^0 the Brownian bridge process on [0, 1], have the same covariance structure [see, for example, Shorack and Wellner (1986), page 308]. Hence, for calculating the covariance of $\int_0^T g(x|t)\, d\psi_1(x)$ one can proceed as in Akritas [(1986), page 625], keeping in mind our present definition of $g(x|t)$. Therefore,

$$\begin{aligned}
&\operatorname{Cov}\big(k_1((T_1, \delta_1); v), k_1((T_1, \delta_1); t)\big) \\
&\quad = \frac{1}{m^2} \Bigg\{ \int_0^T g(x|t) g(x|v) (1 - F(x))^2\, d\gamma(x) \\
&\qquad + \int_0^T g(x|t) \left[\int_0^x g(y|v)\, d((1 - F(y))\gamma(y)) \right] \\
&\qquad\qquad \times d((1 - F(x))(1 + \gamma(x))) \\
&\qquad + \int_0^T g(x|v) \left[\int_0^x g(y|t)\, d((1 - F(y))\gamma(y)) \right] \\
&\qquad\qquad \times d((1 - F(x))(1 + \gamma(x))) \\
&\qquad - \int_0^T g(x|v)\, d((1 - F(x))\gamma(x)) \int_0^T g(x|t)\, d((1 - F(x))\gamma(x)) \Bigg\}
\end{aligned} \tag{4.2}$$

and

$$\begin{aligned}\xi_1(t) &= \operatorname{Var}(k_1((T_1,\delta_1);t))\\ &= \frac{1}{m^2}\left\{\int_0^T g^2(x|t)(1-F(x))^2\,d\gamma(x)\right.\\ &\qquad\left.-2\int_0^T g(x|t)\left[\int_0^x g(y|t)\,d((1-F(y))\gamma(y))\right]dF(x)\right\}.\end{aligned} \tag{4.3}$$

We can now state the following theorem on strong consistency, law of iterated logarithm and asymptotic normality of $\theta_t(\hat{F}_n)$, by invoking the same results for U-statistics [see, e.g., Serfling (1980)] and using the representations in Corollary 1.

THEOREM 2. *Assume conditions* (A1)–(A3). *Then, for all* t,

$$\theta_t(\hat{F}_n) \to \theta_t(F) \quad a.s.$$

and if $\xi_1(t) > 0$,

$$\limsup_{n\to\infty} \frac{n^{1/2}\big(\theta_t(\hat{F}_n) - \theta_t(F)\big)}{\big(2m^2\xi_1(t)\log\log n\big)^{1/2}} = 1 \quad a.s.$$

and

$$n^{1/2}\big(\theta_t(\hat{F}_n) - \theta_t(F)\big) \to_d N\big(0; m^2\xi_1(t)\big).$$

EXAMPLE (Variance). Take $m = 2$ and $h_t(x_1, x_2) = (x_1 - x_2)^2/2$. A straightforward calculation shows that for this particular example, expression (4.2) leads to

$$\begin{aligned}4\xi_1(t) &= g^2(T)(1-F(T))^2\gamma(T)\\ &\quad - 2g(T)(1-F(T))\int_0^T(1-F(y))\gamma(y)g'(y)\,dy\\ &\quad + \int_0^T \frac{1}{(1-H(y))^2}\left[\int_y^T (1-F(x))g'(x)\,dx\right]^2 dH^u(y),\end{aligned} \tag{4.4}$$

with, from (4.1),

$$g(x|t) = x^2F(T) - 2x\alpha_1(T) + \alpha_2(T)$$

and $\alpha_k(T) = \int_0^T x^k\,dF(x)$, $k = 1, 2$.

Assume that $E(X^4) < \infty$ and that $\int_0^\infty dH^u(y)/(1-H(y))^2 < \infty$. If we take the limit for $T \to \infty$, then only the last term in the sum on the right-hand side of (4.4) will give a contribution different from zero. Hence, for $T \to \infty$, $4\xi_1(t)$ tends to

$$\int_0^\infty \frac{1}{(1-H(y))^2}\left[\int_y^\infty (1-F(x))(2x - 2\alpha_1)\,dx\right]^2 dH^u(y),$$

with $\alpha_1 = E(X)$, an expression which can also be derived from formula (3.5) in Reid (1981). We finally mention that in the uncensored case with $T \to \infty$, it can be calculated that (4.4) reduces to the familiar expression for the asymptotic variance of the variance estimator, namely $E[(X - E(X))^4] - \operatorname{Var}^2(X)$.

5. Asymptotic representation for quantiles. In this section we focus on Example 4 in Section 3, that is, the case where h_t has the indicator form

$$h_t(x_1, \dots, x_m) = I(\tilde{h}(x_1, \dots, x_m) \le t),$$

where the function $\tilde{h}$ satisfies condition (B). Instead of $\theta_t(F)$ and $\theta_t(\hat{F}_n)$, we will use a notation which is more convenient for this case, namely

$$H_F(t) = \int_0^T \cdots \int_0^T I(\tilde{h}(x_1, \dots, x_m) \le t) \prod_{i=1}^m dF(x_i),$$

$$H_{\hat{F}_n}(t) = \int_0^T \cdots \int_0^T I(\tilde{h}(x_1, \dots, x_m) \le t) \prod_{i=1}^m d\hat{F}_n(x_i).$$

Note that $H_F(t)$ and $H_{\hat{F}_n}(t)$ are increasing functions of t and that $H_F(\infty) = (F(T))^m$ and $H_{\hat{F}_n}(\infty) = (\hat{F}_n(T))^m$.

For $0 < p < H_F(\infty)$, the quantity

$$Q_p = H_F^{-1}(p) = \inf\{t \colon H_F(t) \ge p\}$$

is well defined. Similarly, for $0 < p < H_{\hat{F}_n}(\infty)$, we put

$$Q_{pn} = H_{\hat{F}_n}^{-1}(p) = \inf\{t \colon H_{\hat{F}_n}(t) \ge p\}.$$

The next theorem gives a Bahadur-type representation for Q_{pn} with remainder term of order $o(n^{-1/2})$ in probability. For quantiles of the classical empirical distribution function, the analogous result has been obtained by Ghosh (1971). Also note that, instead of dealing with Q_{pn} with p fixed, our theorem allows quantiles of the form $Q_{p_n n}$, with $\{p_n\}$ a sequence of numbers tending to some fixed p.

THEOREM 3. *Let $0 < p < H_F(\infty)$. Suppose that H_F is differentiable at Q_p, with $H_F'(Q_p) = h_F(Q_p) > 0$. Let $\{p_n\}$ be a sequence of numbers ($0 < p_n < H_{\hat{F}_n}(\infty)$) such that $p_n - p = O(n^{-1/2})$. Then, as $n \to \infty$,*

$$Q_{p_n n} = Q_p + \frac{p_n - H_{\hat{F}_n}(Q_p)}{h_F(Q_p)} + o_p(n^{-1/2}).$$

PROOF. We only give a sketch of the proof since it follows more or less the same lines as that of Theorem 1 in Ghosh (1971). If we denote $V_n = n^{1/2}[Q_{p_n n} - Q_p + (p - p_n)/h_F(Q_p)]$ and $W_n = n^{1/2}(p - H_{\hat{F}_n}(Q_p))/h_F(Q_p)$, then we prove that $V_n - W_n \to_P 0$ by verifying the conditions of Lemma 1 in Ghosh (1971).

The first condition requires the boundedness in probability of W_n, but this is clear from the representation in Corollary 1 and the boundedness in probability of $n^{1/2}\tilde{R}_n(Q_p)$ and $n^{1/2}U_n(Q_p)$. The latter can be seen as follows. For any $\delta > 0$,

$$P\big(n^{1/2}|U_n(Q_p)| > \delta\big) \le n\delta^{-2}\,\mathrm{Var}\big(U_n(Q_p)\big)$$

and $\mathrm{Var}(U_n(Q_p))$ can be approximated by $m^2 n^{-1}\mathrm{Var}(k_1((T_1,\delta_1);Q_p))$ [see property (3.8) and Lemma A(iii) in Serfling (1980), page 183].

The second condition of Lemma 1 in Ghosh (1971) reduces (omitting details) to verification of the fact that

$$\big[H_{\hat{F}_n}(t_n^*) - H_F(t_n^*)\big] - \big[H_{\hat{F}_n}(Q_p) - H_F(Q_p)\big] = o_P(n^{-1/2}),$$

where $t_n^* = Q_p + O(n^{-1/2})$. Because of the result in Corollary 1, this reduces to showing that $U_n(t_n^*) - U_n(Q_p) = o_P(n^{-1/2})$ or to proving that

$$n\,\mathrm{Var}\big[U_n(t_n^*) - U_n(Q_p)\big] \to 0.$$

Using Lemma A(iii) on page 183 in Serfling (1980) and property (3.8), the variance of the U-statistic $U_n(t_n^*) - U_n(Q_p)$ can be approximated by

$$\frac{m^2}{n}\mathrm{Var}\big[k_1((T_1,\delta_1);t_n^*) - k_1((T_1,\delta_1);Q_p)\big].$$

But

$$\begin{aligned}\mathrm{Var}&\big[k_1((T_1,\delta_1);t_n^*) - k_1((T_1,\delta_1);Q_p)\big]\\ &= \mathrm{Var}\big[k_1((T_1,\delta_1);t_n^*)\big] - 2\,\mathrm{Cov}\big(k_1((T_1,\delta_1);t_n^*), k_1((T_1,\delta_1);Q_p)\big)\\ &\quad + \mathrm{Var}\big[k_1((T_1,\delta_1);Q_p)\big]\end{aligned}$$

tends to zero as $n \to \infty$. This can be shown from expressions (4.2) and (4.3) by applying the dominated convergence theorem for the multiple integrals [see, e.g., Loève (1977), pages 126–127]. One has to take into account condition (B) on the kernel $\tilde{h}$ and the continuity of F and hence of γ. □

The previous representation obviously leads to further results. As an example we state the asymptotic normality of the quantiles $Q_{p_n n}$. For the case of fixed p, this is in compliance with Corollary 3.2 of Akritas (1986).

COROLLARY 2. *Under the conditions of Theorem* 3 *and the additional assumption that* $n^{1/2}(p_n - p) \to c$, *we have*

$$n^{1/2}(Q_{p_n n} - Q_p) \to_d N\left(\frac{c}{h_F(Q_p)}; \frac{m^2\xi_1(Q_p)}{h_F^2(Q_p)}\right).$$

REFERENCES

AKRITAS, M. G. (1986). Empirical processes associated with V-statistics and a class of estimators under random censoring. *Ann. Statist.* **14** 619–637.

APOSTOL, T. M. (1982). *Mathematical Analysis*. Addison-Wesley, Reading, Mass.

GHOSH, J. K. (1971). A new proof of the Bahadur representation of quantiles and an application. *Ann. Math. Statist.* **42** 1957–1961.

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

HEWITT, E. and STROMBERG, K. (1975). *Real and Abstract Analysis*. Springer, New York.

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

LO, S.-H. and SINGH, K. (1986). The product-limit estimator and the bootstrap: Some asymptotic representations. *Probab. Theory Related Fields* **71** 455–465.

LOÈVE, M. (1977). *Probability Theory I*. Springer, New York.

MAJOR, P. and REJTŐ, L. (1988). Strong embedding of the estimator of the distribution function under random censorship. *Ann. Statist.* **16** 1113–1132.

REID, N. (1981). Influence functions for censored data. *Ann. Statist.* **9** 78–92.

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

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

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

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

LIMBURGS UNIVERSITAIR CENTRUM
UNIVERSITAIRE CAMPUS
B-3590 DIEPENBEEK
BELGIUM