*The Annals of Statistics*
1982, Vol. 10, No. 2, 595–602

# ASYMPTOTIC OPTIMALITY OF THE PRODUCT LIMIT ESTIMATOR[1]

BY JON A. WELLNER

*University of Munich and University of Rochester*

The product limit estimator due to Kaplan and Meier (1958) is well-known to be the nonparametric maximum likelihood estimator of a distribution function based on censored data. It is shown here that the product limit estimator is an asymptotically optimal estimator in two senses: in the sense of a Hájek-Beran type representation theorem for regular estimators; and in an asymptotic minimax sense similar to the classical result for the uncensored case due to Dvoretzky, Kiefer, and Wolfowitz (1956). The proofs rely on the methods of Beran (1977) and Millar (1979).

**1. Introduction:** *the censored data problem.* Let $(Y_{1i}, Y_{2i})$, $i = 1, \cdots, n$, be independent identically distributed pairs of non-negative random variables with distribution function (df's) $G_1(t) = P(Y_{1i} \leq t)$ and $G_2(t) = P(Y_{2i} \leq t)$, $t \in [0, \infty)$. Assume that $Y_{1i}$ and $Y_{2i}$ are independent for all $i$, and that $G_1$ and $G_2$ are continuous. In the censored data problem with random censorship we observe the $n$ pairs of random variables $(X_i, \delta_i)$, $i = 1, \cdots, n$, where

$$X_i = \min\{Y_{1i}, Y_{2i}\}, \quad \delta_i = \begin{cases} 1 & \text{if} \quad X_i = Y_{1i} \\ 2 & \text{if} \quad X_i = Y_{2i}. \end{cases} \tag{1.1}$$

The problem is to estimate the distribution function $G_1$.

Let $(X_{(i)}, \delta_{(i)})$, $i = 1, \cdots, n$ denote the $(X, \delta)$ pairs ordered by the $X$'s; i.e. $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$. Then the well-known product-limit estimator $\hat{\mathbb{G}}_{1n}$ of $G_1$ is defined by $\hat{\mathbb{G}}_{1n} = 1 - \hat{\mathbb{S}}_{1n}$ where

$$\hat{\mathbb{S}}_{1n}(t) = \prod_{i: X_{(i)} \leq t} \left(\frac{n-i}{n-i+1}\right)^{2-\delta_{(i)}}, \quad 0 \leq t < \infty; \tag{1.2}$$

see e.g. Peterson (1977). The product-limit estimator $\hat{\mathbb{G}}_{1n}$ was derived by Kaplan and Meier (1958) as the "nonparametric maximum likelihood estimator" of $G_1$ in the above problem; see Johansen (1978) and Scholz (1980) for reexaminations of the maximum likelihood character of $\hat{\mathbb{G}}_1$. Because the product limit estimator is a nonparametric maximum likelihood estimator, it seems to be generally assumed that it will have the optimality properties characteristic of maximum likelihood estimates in more familiar finite-dimensional problems. But apparently no such optimality properties have been proved so far. The question remains: Is the product limit estimator a "good" estimator of $G_1$, at least asymptotically?

Our object here is to answer the above question affirmatively. Theorem 1 gives a representation for the asymptotic distribution of regular estimates of $G_1$ which asserts, roughly, that the limiting process for *any* sequence of regular estimates $\tilde{\mathbb{G}}_{n1}$ of $G_1$ must be at least as dispersed as the limit process corresponding to the product limit estimator $\hat{\mathbb{G}}_{n1}$. This theorem is analogous to the representation theorem established by Beran (1977) in the uncensored case. Theorem 2 gives asymptotic minimax properties of the product-limit estimator which cover a wide variety of loss functions. The proof of Theorem 2 is based on

Received February 1981; revised July 1981.

[1] Research supported by the Alexander von Humboldt Foundation while the author was visiting the University of Munich.

*AMS* 1980 *subject classifications.* Primary 62G05, 62G20; secondary 62E20.

*Key words and phrases.* Asymptotic minimax, censored data, convolution representation, distribution function, regular estimation.

work of Hájek (1972), LeCam (1972), and, especially, Millar (1979); see also Levit (1978). Theorem 2 is not quite as strong as the classical asymptotic minimax theorem for the sample distribution function due to Dvoretzky, Kiefer, and Wolfowitz (1956) (see also Millar, 1979) essentially because the sup norm of the limit process in the present censored case is *not* distribution free.

**2. The main results.** We first need some notation and easy facts concerning $L^2$ spaces: Let $\nu$ be a measure on $R = (-\infty, \infty)$ with respect to which $G_1$ and $G_2$ have densities $g_1$ and $g_2$ respectively (the measure $\nu$ induced by $G_1 + G_2$ always works). Then it is easy to see that the $(X, \delta)$ pairs observed have a density $f$ with respect to $\mu \equiv \nu \times$ (counting measure) on $S \equiv R \times \{1, 2\}$ given by

$$f(x, \delta) = [\{1 - G_2(x)\}g_1(x)]^{2-\delta}[\{1 - G_1(x)\}g_2(x)]^{\delta-1}, \quad (x, \delta) \in S. \tag{2.1}$$

It is also easy to recover $G_1$ (or $G_2$) from $f$: First by writing $f_i \equiv f(\cdot\,; i)$, $F_i(x) \equiv P(X \le x, \delta = 1) = \int_{-\infty}^{x} f_i\, d\nu$, $i = 1, 2$ and $F(x) = P(X \le x) = F_1(x) + F_2(x)$, we have $\bar{F} \equiv (1 - F) = (1 - G_1)(1 - G_2)$. Then note that

$$\exp\left(-\int_0^x \frac{1}{\bar{F}} f_1\, d\nu\right) = 1 - G_1(x), \tag{2.2}$$

by an easy computation, where the left side is just a function of $f$; e.g. see equation (2.2) of Peterson (1977).

Let $\|\cdot\|_\mu$ and $\langle\cdot,\cdot\rangle_\mu$, and $\|\cdot\|_\nu$ and $\langle\cdot,\cdot\rangle_\nu$, denote the usual norm and inner products on $L^2(S, \mu)$ and $L^2(R, \nu)$, respectively, and note that for arbitrary functions $f = (f_1, f_2)$ and $g = (g_1, g_2)$ in $L^2(S, \mu) = L^2(R, \nu) \times L^2(R, \nu)$

$$\langle f, g\rangle_\mu = \langle f_1, g_1\rangle_\nu + \langle f_2, g_2\rangle_\nu. \tag{2.3}$$

We will usually write $\langle\cdot,\cdot\rangle$ for both $\langle\cdot,\cdot\rangle_\mu$ and $\langle\cdot,\cdot\rangle_\nu$ (and similarly for the norms), and add the subscripts only when confusion might otherwise arise.

Let $\mathscr{F}(\mu)$ denote the set of all densities with respect to $\mu$ on $S$. Let $\mathscr{C}(f, \alpha)$ denote the set of all sequences of densities $\{f_m \in \mathscr{F}(\mu)\}$ such that

$$\lim_{m\to\infty} \| m^{1/2}(f_m^{1/2} - f^{1/2}) - \alpha \| = 0, \tag{2.4}$$

where $\alpha \in L^2(S, \mu)$. This implies that $\alpha$ is orthogonal to $f^{1/2}$ in $L^2(S, \mu)$, as is easily shown. Let $\mathscr{C}(f)$ denote the union of all sets $\{\mathscr{C}(f, \alpha) : \alpha \in L^2(S, \mu),\ \alpha \perp f^{1/2}\}$.

Set $T_i \equiv \inf\{t : G_i(t) = 1\}$, equal to $\infty$ if the set is empty, for $i = 1, 2$. Note that we cannot hope to estimate $G_1$ to the right of $T_0 \equiv \min(T_1, T_2)$ when we observe only the $(X, \delta)$ pairs, since $P(X_i \le T_0$ for all $i = 1, \cdots, n) = 1$. (In many practical situations $T_0 = T_2 < T_1 \le \infty$.) Thus we will consider estimation of $G_1$ on $[0, T]$ where $T < T_0$; this will also simplify our proofs.

Let $\{f_m\} \in \mathscr{C}(f)$ and let $\{G_{1m}\}$ be the corresponding sequence of $G_1$'s obtained via (2.2): i.e.

$$1 - G_{1m}(x) \equiv \exp\left(-\int_0^x \frac{1}{\bar{F}_m} f_{1m}\, d\nu\right). \tag{2.5}$$

Consider the corresponding sequence of experiments where, in the $n$th experiment, we observe $n$ independent pairs $(X_{ni}, \delta_{ni})$, $i = 1, \cdots, n$ with joint density $\prod_{i=1}^n f_n(x_{ni}, \delta_{ni})$ on $S^n$. Let $\{\tilde{\mathscr{G}}_{1n}\}$ be any sequence of $C[0, T]$-valued estimators where $\tilde{\mathscr{G}}_{1n}$ is a function of $\{(X_{ni}, \delta_{ni}), i = 1, \cdots, n\}$.

We say that an estimating sequence, or estimator, $\{\tilde{\mathscr{G}}_{1n}\}$ of $G_1$ is *regular at f* if the distributions $\mathscr{L}\{n^{1/2}(\tilde{\mathscr{G}}_{1n} - G_{1n})\}$ on $C[0, T]$ converge weakly to the same distributions $\mathscr{D} = \mathscr{D}_f$, depending only on $f$, for all sequences $\{f_m\} \in \mathscr{C}(f)$. Of course $\mathscr{D}$ may also depend on the estimator $\tilde{\mathscr{G}}_{1n}$. Let $\tilde{Z}$ denote a process with law $\mathscr{D}$ on $C[0, T]$.

Now let $Z = \{Z(t)\}_{0 \le t \le T}$ be a mean-zero Gaussian process on $[0, T]$ with covariance function

$$E\{Z(s)Z(t)\} = \bar{G}_1(s)\bar{G}_1(t)C(s \wedge t) \tag{2.6}$$

where

$$C(t) \equiv \int_0^t \frac{1}{\bar{\bar{F}}^2} f_1 \, d\nu = \int_0^t \frac{1}{\bar{G}_1^2 \bar{G}_2} \, dG_1. \tag{2.7}$$

Let $\mathscr{D}_Z = \mathscr{L}(Z)$ denote the distribution of $Z$ on $C[0, T]$.

The following theorem extends the result of Beran (1977) to the case of randomly censored data.

THEOREM 1. *For any regular estimator $\tilde{\mathbb{G}}_{1n}$ of $G_1$ in the random-censorship model (based only on observation of $\{(X_i, \delta_i), i = 1, \cdots, n\}$), the limiting law $\mathscr{D}$ on $C[0, T]$, $T < T_0$, may be represented as $\mathscr{D}_Z * \mathscr{D}_W$ where $\mathscr{D}_Z$ is the distribution of the mean zero Gaussian process $Z$ with covariance function given by (2.6) and $\mathscr{D}_W$ is the distribution of some independent process $W$. Equivalently,*

$$\tilde{Z} = Z + W \tag{2.8}$$

*in distribution where $Z$ and $W$ are independent.*

To see that the product-limit estimator $\hat{\mathbb{G}}_{1n}$ given in (1.2) is asymptotically optimal recall that, by Theorem 5 of Breslow and Crowley (1974), the process

$$\hat{Z}_n \equiv n^{1/2}(\hat{\mathbb{G}}_{1n} - G_1) \tag{2.9}$$

converges weakly to the process $Z$. In other words, $\hat{\mathbb{G}}_{1n}$ is a sequence of estimators for which $W \equiv 0$ in (2.8), and hence the product limit estimator is optimal in the sense of Theorem 1. Note that the obvious "lower linear interpolation" of $\hat{S}_{1n}$, $\hat{S}_{1n}^c$ yields a continuous estimator $\hat{\mathbb{G}}_{1n}^c$ with corresponding process $\hat{Z}_n^c$ satisfying $\|\hat{Z}_n - \hat{Z}_n^c\|_0^T = o_p(1)$. Hence $\hat{Z}_n^c$ also converges weakly to $Z$, since $\hat{Z}_n$ converges weakly to a continuous $Z$, $n^{1/2} \times$ (maximum jump of $\hat{\mathbb{G}}_{1n}$ on $[0, T]$) $= o_p(1)$.

Now we turn to our asymptotic minimax result for the product limit estimator $\hat{\mathbb{G}}_{1n}$.

As in Millar (1979), let $\ell : C[0, T] \to R^+$ be subconvex; e.g. $\ell(x) = \sup_t |x(t)| \equiv \|x\|$, $\ell(x) = \int |x(t)|^2 \, dt$, and $\ell(x) = 1\{x : \|x\| > c\}$ are all subconvex. Let $\mathscr{F}$ be the collection of all continuous distributions on $S$ of the form (2.1); i.e.

$$\mathscr{F} = \{F = (F_1, F_2) : F_1 = \int^{\cdot} (1 - G_2) \, dG_1,\ F_2 = \int^{\cdot} (1 - G_1) \, dG_2 \text{ for some continuous df's } G_1, G_2 \text{ on } R^+\}.$$

Although we are now using "$F$" in two ways, this should cause no confusion. As in Millar (1979), let $b$ denote the procedures (i.e. Markov kernels: for each $s \in S^n$, $b(s, \cdot)$ is a probability on $(C[0, T], \mathscr{C})$ where $\mathscr{C}$ = Borel subsets of $C[0, T]$ with the supremum norm; and for each $A \in \mathscr{C}$, $b(\cdot, A)$ is measurable), and let $F^n$ denote the $n$-fold product measure corresponding to $F$. See pages 234–235 of Millar (1979) for more details.

THEOREM 2. *For fixed $0 < T < \infty$ and $0 < \delta < 1$, let $\ell : C[0, T] \to R^+$ be subconvex, and let $\mathscr{F}^+ \equiv \mathscr{F}^+(T, \delta) \equiv \{F = (F_1, F_2) \in \mathscr{F} : 1 - F(T) \ge \delta > 0\}$. Then*

$$\liminf_{n \to \infty} \inf_b \sup_{F \in \mathscr{F}^+} \iint \ell\{n^{1/2}(y - G_1)\} b(s, dy) F^n(ds) \ge \sup_{F \in \mathscr{F}^+} E\ell(Z), \tag{2.10}$$

*where $Z$ is the mean-zero Gaussian process on $[0, T]$ with covariance given by (2.6).*

*Furthermore, if*

$$\lim_{n\to\infty} \int \ell\{n^{1/2}(\hat{\mathcal{G}}^c_{1n} - G_1)\} F^n(ds) = E\ell(Z) \text{ uniformly in } F \in \mathscr{F}^+, \tag{2.11}$$

*then* $\hat{\mathcal{G}}^c_{1n}$ *is asymptotically minimax in* $\mathscr{F}^+$; *i.e.*

$$\lim_{n\to\infty} \frac{\sup_{F\in\mathscr{F}^+} \int \ell\{n^{1/2}(\hat{\mathcal{G}}^c_{1n} - G_1)\} F^n(ds)}{\inf_b \sup_{F\in\mathscr{F}^+} \int\int \ell\{n^{1/2}(y - G_1)\} b(s, dy) F^n(ds)} = 1. \tag{2.12}$$

REMARK 1. If $\ell(x) = 1\{x: \|x\|_0^T > c\}$, $\ell(x) = \|x\|_0^T$, or some nice function of $\|x\|_0^T$, (2.11) can be verified by using the exponential bound given in Theorem 2, page 82, of Földes and Rejtő (1981) together with the strong approximation result contained in Corollary 6.1, page 104, of Burke, Csörgő, and Horváth (1981). In particular, taking $b_n \equiv \{1 - F(T)\}^{-1} \le \delta^{-1} < \infty$ for all $F \in \mathscr{F}^+$ in (4.4) on page 94, and thus also on pages 101 and 104, of Burke et al. yields $r(n) = O(n^{-1/3}(\log n)^{3/2})$ uniformly in $F \in \mathscr{F}^+$. Thus their Corollary 6.1 on page 104 implies that

$$\sup_{F\in\mathscr{F}^+} P\{\|\hat{Z}_n - Z_n\|_0^T > r(n)\} < \text{constant}\cdot n^{-(1+\delta)} \tag{a}$$

where $\hat{Z}_n =_d n^{1/2}(\hat{\mathcal{G}}_{1n} - G_1)$ and $Z_n =_d Z$ for all $n \ge 1$. After a truncation argument, with the truncation error controlled uniformly in $F \in \mathscr{F}^+$ by use of the exponential bound of Földes and Rejtő (1981), (a) implies (2.11) for loss functions $\ell$ which are nice functions of $\|x\|_0^T$.

REMARK 2. A more satisfactory asymptotic minimax theorem for the product limit estimator would allow loss functions $\ell$ defined on $C[0, T_0]$, $T_0 = \min(T_1, T_2) \le \infty$. Results of this type may be possible using the fact that if $C(T_0) = \infty$, then $\|Zw_F\|_0^{T_0} =_d \|B^0\|_0^1$ is distribution free where $w_F = (1 + C)^{-1}(1 - G_1)^{-1}$ is a weight function, $C$ is the function defined in (2.7), and $B^0$ denotes a Brownian bridge process on [0, 1]; see Hall and Wellner (1980).

**3. Proof of the theorems.** We begin with several lemmas, whose proofs we will defer until Section 4. Lemma 1 describes the behaviour of the likelihood ratios

$$L_n \equiv 2 \log \prod_{i=1}^n \{f_n^{1/2}(X_i, \delta_i)/f^{1/2}(X_i, \delta_i)\} \tag{3.1}$$

for $\{f_n\} \in \mathscr{C}(f, \alpha)$; this result has been used repeatedly by Beran, e.g. Beran (1977), and can be deduced easily from LeCam's second lemma. Lemmas 2 and 3 give the limiting behaviour of $n^{1/2}(G_{1n} - G_1)$ for $\{f_n\} \in \mathscr{C}(f, \alpha)$; the characteristic functional of the process $Z$ is computed in Lemma 4; the results of some straightforward $L^2$ computations are summarized in Lemma 5.

LEMMA 1. *If* $\{f_n\} \in \mathscr{C}(f, \alpha)$, $\alpha \in L^2(S, \mu)$, *then for every* $\varepsilon > 0$, $\lim_{n\to\infty} P_f\{|L_n - 2n^{-1/2} \sum_{i=1}^n \alpha(X_i, \delta_i) f^{-1/2}(X_i, \delta_i) + 2\|\alpha\|^2| > \varepsilon\} = 0$.

For each $t \in [0, T]$ define $\gamma_t: S \to R$ by

$$\gamma_t(s, \delta) = \begin{cases} \gamma_t(s, 1) = 1_{[0,t]}(s)\left\{\dfrac{1}{\bar{F}(s)} + C(t) - C(s)\right\} f^{1/2}(s, 1) \\ \\ \gamma_t(s, 2) = 1_{[0,t]}(s)\{C(t) - C(s)\} f^{1/2}(s, 2) \end{cases} \tag{3.2}$$

where $C$ is given by (2.7).

LEMMA 2. *If* $\{f_n\} \in \mathscr{C}(f, \alpha)$ *and* $\{G_{1n}\}$ *is defined by* (2.5), *then* $\sup_{0\le t\le T} | n^{1/2}\{G_{1n}(t) - G_1(t)\} - 2\bar{G}_1(t)\langle\alpha, \gamma_t\rangle | \to 0$ *as* $n \to \infty$; *recall that* $T < T_0 = \min(T_1, T_2)$.

Let $v$ be a function of bounded variation on $[0, T]$, set

$$V(s) \equiv \int_s^T \bar{G}_1 \, dv, \quad U(s) \equiv \int_s^T \bar{G}_1 C \, dv, \tag{3.3}$$

and define $\eta : S \to R$ by

$$\eta(s, \delta) = \begin{cases} \eta(s, 1) = [\{V(s)/\bar{F}(s)\} + U(s) - C(s)V(s)]f^{1/2}(s, 1)1_{[0,T]}(s) \\ \eta(s, 2) = \{U(s) - C(s)V(s)\}f^{1/2}(s, 2)1_{[0,T]}(s). \end{cases} \tag{3.4}$$

LEMMA 3. *If* $\{f_n\} \in \mathscr{C}(f, \alpha)$, $\{G_{1n}\}$ *is given by* (2.5), *and* $v$ *is a function of bounded variation on* $[0, T]$, *then*

$$\lim_{n\to\infty} \int_0^T n^{1/2}(G_{1n} - G_1) \, dv = 2 \int_0^T \bar{G}_1(t)\langle\alpha, \gamma_t\rangle \, dv(t) = 2\langle\alpha, \eta\rangle.$$

LEMMA 4. *The characteristic functional of the Gaussian process* $Z$ *on* $[0, T]$ *with covariance function* (2.6) *is*

$$E \exp\left(i \int_0^T Z \, dv\right) = \exp\left(-\frac{1}{2}\sigma^2\right) \tag{3.5}$$

*where*

$$\sigma^2 = \sigma^2(v) \equiv \int_0^T V^2 \, dC = \int_0^T \left(\int_\cdot^T \bar{G}_1 \, dv\right)^2 \frac{1}{\bar{G}_1^2 \bar{G}_2} \, dG_1. \tag{3.6}$$

LEMMA 5. *Let* $\eta$ *be the function defined in* (3.4); *then*

(i) $$\|\eta\|^2 = \sigma^2 + b^2 \quad \text{and} \quad \langle\eta, f^{1/2}\rangle = b,$$

*where* $\sigma^2 = \sigma^2(v)$ *is defined in* (3.6) *and* $b \equiv U(0)$. *Hence* $\eta_0 \equiv \eta - \langle\eta, f^{1/2}\rangle f^{1/2} = \eta - bf^{1/2}$ *and* $\alpha_* \equiv \eta_0/\sigma$ *satisfy* $\eta_0 \perp f^{1/2}$, $\alpha_* \perp f^{1/2}$ *(trivially) and*

(ii) $$\|\eta_0\|^2 = \sigma^2, \quad \|\alpha_*\|^2 = 1, \quad \langle\alpha_*, \eta\rangle = \sigma.$$

PROOF OF THEOREM 1. Let $\{\hat{G}_{1n}\}$ be a regular estimator of $G_1$. The characteristic functional of $n^{1/2}(\hat{G}_{1n} - G_{1n})$ under $f_n$ is given by

$$\begin{aligned} E_{f_n}\exp\left\{i \int_0^T n^{1/2}(\hat{G}_{1n} - G_{1n}) \, dv\right\} &= E_f \exp\left\{i \int_0^T n^{1/2}(\hat{G}_{1n} - G_{1n}) \, dv + L_n\right\} \\ &= E_f \exp\left\{i \int_0^T n^{1/2}(\hat{G}_{1n} - G_1) \, dv - 2i\langle\alpha, \eta\rangle + L_n\right\} + o(1), \end{aligned} \tag{3.7}$$

by Lemma 3, the latter being true for all $\alpha \in L^2(S, \mu)$, $\alpha \perp f^{1/2}$, and all functions $v$ of bounded variation on $[0, T]$. By regularity, (3.7) converges to $E \exp(i \int_0^T \tilde{Z} \, dv)$. The rest of the proof proceeds exactly as in Beran (1977), except that we choose $\alpha = h\alpha_*$ where $\alpha_*$ given in Lemma 5 satisfies $\alpha_* \perp f^{1/2}$, $\|\alpha_*\| = 1$, and $\langle\alpha_*, \eta\rangle = \sigma$. The final result is that

$$\varphi(v, 0) = \varphi(v, -\sigma)\exp(-\tfrac{1}{2}\sigma^2)$$

where $\varphi(v, 0)$ is the characteristic functional of $\tilde{Z}$, $\exp(-\tfrac{1}{2}\sigma^2)$ is the characteristic functional of $Z$ by Lemma 4, and $\varphi(v, -\sigma)$ is the characteristic functional of $W$. □

PROOF OF THEOREM 2. Most of the proof follows directly from Lemmas 1 through 5 together with the methods and results of Millar (1979), so we only give a sketch here. To make appropriate identifications with Millar's Section 3, for fixed $F \in \mathscr{F}^+$ with density $f$, take $H = \{\alpha \in L^2(S, \mu) : \alpha \perp f^{1/2}\}$, $B = C[0, T]$, $B^* = BV[0, T]$ = functions of bounded variation on $[0, T]$, and define $\tau : L^2(S, \mu) \to C[0, T]$ by $\tau\alpha(t) = \bar{G}_1(t)\langle \alpha, \gamma_t \rangle$ where $\gamma_t$ is given in (3.2). Also define $\tau^* : BV[0, T] \to H$ by $\tau^* v = \eta_0$ where $\eta_0$ is given by (3.4) and Lemma 5. Then $\tau^*$ is the adjoint of $\tau$: by Lemmas 3 and 5 $\langle \alpha, \tau^* v \rangle = \langle \alpha, \eta_0 \rangle = \langle \alpha, \eta \rangle = \int_0^T \bar{G}_1 \langle \alpha, \gamma_. \rangle \, dv = \int_0^T (\tau\alpha) \, dv$ for $\alpha \in H$. Thus, by way of Lemmas 4 and 5 $\| \tau^* v \|^2 = \| \eta_0 \|^2 = \sigma^2(v)$ and hence Millar's $P_0$ is the law of $Z$ on $C[0, T]$, while $P_h$ is the law of $Z + \tau h = Z + 2\tau\alpha$, letting $h = 2\alpha$.

Once these identifications have been made, the proof of (2.10) proceeds much as the proof of Millar's Proposition 5.1, using Millar's Propositions 2.1 and 3.1 and Remark 1; we omit the details. Then (2.12) follows directly from (2.10) and (2.11). □

## 4. Proofs of the lemmas.

PROOF OF LEMMA 2. Let $\| \cdot \|_0^T \equiv \sup_{0 \le t \le T} | \cdot |$ denote the supremum norm on $C[0, T]$. First note that

$$\sup_{0 \le t \le T} | n^{1/2}\{F_n(t) - F(t)\} - 2\langle \alpha, f^{1/2} 1_{[0,t]} \rangle | \to 0 \tag{4.1}$$

where

$$\langle \alpha, f^{1/2} 1_{[0,t]} \rangle = \int_0^t \{\alpha(s, 1) f^{1/2}(s, 1) + \alpha(s, 2) f^{1/2}(s, 2)\} \, d\nu(s); \tag{4.2}$$

this is similar to (2.9) of Beran (1977). Thus $\| F_n - F \|_0^T \to 0$, and $\| \bar{F}/\bar{F}_n \|_0^T \to 1$ since $T < T_0$ implies $\bar{F}(T) > 0$. Hence we have

$$\begin{aligned}
n^{1/2}(\Lambda_n(t) - \Lambda(t)) &\equiv n^{1/2}\left\{ \int_0^t \frac{f_n(s, 1)}{\bar{F}_n(s)} \, d\nu(s) - \int_0^t \frac{f(s, 1)}{\bar{F}(s)} \, d\nu(s) \right\} \\
&= \int_0^t n^{1/2}\{ f_n^{1/2}(s, 1) - f^{1/2}(s, 1)\}\{ f_n^{1/2}(s, 1) + f^{1/2}(s, 1)\} \frac{1}{\bar{F}(s)} \, d\nu(s) \\
&\quad + \int_0^t n^{1/2}\{F_n(s) - F(s)\} \frac{f(s, 1)}{\bar{F}(s)\bar{F}_n(s)} \, d\nu(s) \\
&\to 2 \int_0^t \alpha(s, 1) f^{1/2}(s, 1) \frac{1}{\bar{F}(s)} \, d\nu(s) + 2 \int_0^t \langle \alpha, f^{1/2} 1_{[0,s]} \rangle \frac{f(s, 1)}{\bar{F}(s)^2} \, d\nu(s) \\
&= 2\langle \alpha, \gamma_t \rangle,
\end{aligned}$$

where $\gamma_t$ is defined by (3.2), by using (4.2) and then Fubini's theorem on the second term. It is easily shown, using $\bar{F}(T) > 0$, that this convergence is uniform in $0 \le t \le T$; i.e.

$$\| n^{1/2}(\Lambda_n - \Lambda) - 2\langle \alpha, \gamma_. \rangle \|_0^T \to 0 \quad \text{as } n \to \infty.$$

Then it follows, by a mean value theorem argument, that with $\| \Lambda_n^* - \Lambda \|_0^T \le \| \Lambda_n - \Lambda \|_0^T \to 0$,

$$\begin{aligned}
n^{1/2}(\bar{G}_{n1} - \bar{G}_1) \equiv n^{1/2}\{\exp(-\Lambda_n) - \exp(-\Lambda)\} &= -\exp(-\Lambda_n^*) n^{1/2}(\Lambda_n - \Lambda) \\
&\to -\exp(-\Lambda) 2\langle \alpha, \gamma_. \rangle = -2\bar{G}_1 \langle \alpha, \gamma_. \rangle
\end{aligned}$$

uniformly on $[0, T]$. □

PROOF OF LEMMA 3. The convergence part of Lemma 3 follows immediately from

Lemma 2 together with the fact that $v$ is of bounded variation. The second equality results from straightforward computation using Fubini's theorem: $\langle \alpha, \gamma_t \rangle = \langle \alpha_1, \gamma_{1t} \rangle + \langle \alpha_2, \gamma_{2t} \rangle$ where

$$\langle \alpha_1, \gamma_{1t} \rangle = \int_0^t \alpha_1 f_1^{1/2}\left(\frac{1}{\bar{F}} - C\right) d\nu + C(t) \int_0^t \alpha_1 f_1^{1/2} \, d\nu,$$

$$\langle \alpha_2, \gamma_{2t} \rangle = -\int_0^t \alpha_2 f_2^{1/2} C \, d\nu + C(t) \int_0^t \alpha_2 f_2^{1/2} \, d\nu.$$

Hence, using Fubini's theorem with $V$ and $U$ given in (3.3),

$$\int_0^T \bar{G}_1(t) \langle \alpha, \gamma_t \rangle \, dv(t) = \int_0^T \left\{ V\left(\frac{1}{\bar{F}} - C\right) + U \right\} \alpha_1 f_1^{1/2} \, d\nu + \int_0^T (U - CV) \alpha_2 f_2^{1/2} \, d\nu$$

$$= \langle \alpha_1, \eta_1 \rangle + \langle \alpha_2, \eta_2 \rangle = \langle \alpha, \eta \rangle. \qquad \square$$

PROOF OF LEMMA 4. Since $Z$ is mean-zero Gaussian, its characteristic functional on $[0, T]$ is given by $\exp\{-\frac{1}{2}\sigma^2(v)\}$ where, for functions of bounded variation $v$ on $[0, T]$,

$$\sigma^2(v) = E\left(\int_0^T Z \, dv\right)^2 = \int_0^T \int_0^T \bar{G}_1(s) \bar{G}_1(t) C(s \wedge t) \, dv(s) \, dv(t)$$

$$= \int_0^T \int_0^T \bar{G}_1(s) \bar{G}_1(t) \int_0^T 1_{[0, s \wedge t]}(r) \, dC(r) \, dv(s) \, dv(t) = \int_0^T V(r)^2 \, dC(r)$$

by repeated use of Fubini's theorem. $\square$

PROOF OF LEMMA 5. Now

$$\| \eta_1 \|^2 = \int_0^T (V/\bar{F})^2 f_1 \, d\nu + 2 \int_0^T (V/\bar{F})(U - VC) f_1 \, d\nu + \int_0^T (U - VC)^2 f_1 \, d\nu$$

and

$$\| \eta_2 \|^2 = \int_0^T (U - VC)^2 f_2 \, d\nu.$$

This yields, noting that $\int_0^T (V/\bar{F})^2 f_1 \, d\nu = \int_0^T V^2 \, dC \equiv \sigma^2(v)$,

$$\| \eta \|^2 = \sigma^2(v) + 2 \int_0^T (V/\bar{F})(U - VC) f_1 \, d\nu + \int_0^T (U - VC)^2 \, dF.$$

But, upon noting that $d(U - VC) = dU - C\,dV - V\,dC = -\bar{G}_1 C \, dv + \bar{G}_1 C \, dv - V\,dC = -V\,dC$, an integration by parts yields

$$\int_0^T (U - VC)^2 \, dF = -2 \int_0^T (V/\bar{F})(U - VC) f_1 \, d\nu + U(0)^2.$$

Thus $\| \eta \|^2 = \sigma^2 + U(0)^2 \equiv \sigma^2 + b^2$. Moreover, by way of a similar integration by parts, $\langle \eta, f^{1/2} \rangle = U(0) \equiv b$. Part (ii) of the Lemma follows easily from (i). $\square$

## REFERENCES


BERAN, RUDOLF (1977). Estimating a distribution function. *Ann. Statist.* **5** 400–404.

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

BURKE, MURRAY D., CSÖRGŐ, SÁNDOR, and HORVÁTH, LAJOS (1981). Strong approximations of some biometric estimates under random censorship. *Z. Wahrsch. verw. Gebiete* **56** 87–112.

DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

FÖLDES, A. and REJTŐ, L. (1981). A LIL type result for the product limit estimator. *Z. Wahrsch. verw. Gebiete* **56** 75–86.

HÁJEK, JAROSLAV (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. verw. Gebiete* **14** 323–339.

HÁJEK, JAROSLAV (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth. Berkeley Symp. Math. Statist. and Probability.* Univ. California Press **1** 175–194.

HALL, W. J. and WELLNER, JON A. (1980). Confidence bands for a survival curve from censored data. *Biometrika* **67** 133–143.

JOHANSEN, SOREN (1978). The product limit estimator as maximum likelihood estimator. *Scand. J. Statist.* **5** 195–199.

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

LECAM, L. (1972). Limits of Experiments. *Proc. Sixth. Berkeley Symp. Math. Statist. and Probability.* Univ. California Press **1** 245–261.

LEVIT, B. YA. (1978). Infinite dimensional informational-lower bounds. *Theory Probability Appl.* **23** 371–377.

MILLAR, P. W. (1979). Asymptotic minimax theorems for the sample distribution function. *Z. Wahrsch. verw. Gebiete* **48** 233–252.

PETERSON, ARTHUR V. (1977). Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions. *J. Amer. Statist. Assoc.* **72** 854–858.

SCHOLZ, F. W. (1980). Towards a unified definition of maximum likelihood. *Canad. J. Statist.* **8** 193–203.

MATHEMATISCHES INSTITUT
DER LUDWIG-MAXIMILIANS-UNIVERSITÄT
THERESIENSTRAßE 39
D-8000 MÜNCHEN 2
WEST GERMANY

DEPARTMENT OF STATISTICS
UNIVERSITY OF ROCHESTER
ROCHESTER, NY 14627
U.S.A.