The Annals of Statistics
1992, Vol. 20, No. 2, 1022–1039

LARGE SAMPLE STUDY OF EMPIRICAL DISTRIBUTIONS IN A RANDOM-MULTIPLICATIVE CENSORING MODEL

BY Y. VARDI[1] AND CUN-HUI ZHANG[2]

Rutgers University

Consider an incomplete data problem with the following specifications. There are three independent samples $(X_1, \dots, X_m)$, $(Z_1, \dots, Z_n)$ and $(U_1, \dots, U_n)$. The first two samples are drawn from a common lifetime distribution function G, while the third sample is drawn from the uniform distribution over the interval $(0, 1)$. In this paper we derive the large sample properties of $\hat{G}_{m,n}$, the nonparametric maximum likelihood estimate of G based on the observed data $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$, where $Y_i \equiv Z_i U_i$, $i = 1, \dots, n$. (The Z's and U's are unobservable.) In particular we show that if m and n approach infinity at a suitable rate, then $\sup_t |\hat{G}_{m,n}(t) - G(t)| \to 0$ (a.s.), $\sqrt{m+n}\,(\hat{G}_{m,n} - G)$ converges weakly to a Gaussian process and the estimate $\hat{G}_{m,n}$ is asymptotically efficient in a nonparametric sense.

1. Introduction and summary.

1.1. *The problem.* We consider the random-multiplicative censoring model introduced in Vardi (1989). There are three independent random samples: $(X_1, \dots, X_m)$, $(Z_1, \dots, Z_n)$ and $(U_1, \dots, U_n)$. The X and Z samples are both from a lifetime distribution G, while the U sample is from the uniform $(0, 1)$ distribution. Let

$$Y_i \equiv Z_i U_i, \qquad i = 1, \dots, n,$$

so that the Y's are continuous rv's with a common lifetime distribution F and density

$$f(y) = \int_{y<z} z^{-1}\, dG(z), \qquad 0 < y. \tag{1.1}$$

The observed data are the two random samples $(X_1, \dots, X_m)$ and $(Y_1, \dots, Y_n)$, while the Z's and the U's are unobservable. We can think of this as an *informative censoring* model where the probabilistic relationship between the censored observations Y's and the unobservable "parent" variables Z's is specified. The purpose of this paper is to study the large sample behavior of the nonparametric maximum likelihood estimate (NPMLE) for the unknown distribution G.

Received February 1989; revised March 1991.
[1]Research partially supported by NSF Grant DMS-88-02893.
[2]Research partially supported by NSF Grant DMS-89-16180 and ARO Grant DAAL03-91-G-0045.
AMS 1980 *subject classifications*. Primary 62G05, 62G20.
Key words and phrases. Censored data, informative censoring, nonparametric maximum likelihood estimation, weak convergence, efficiency, survival function.

Since X_i has distribution G and Y_j has density (1.1), the nonparametric likelihood function based on the observations $x_1, \dots, x_m, y_1, \dots, y_n$ is

$$L(G) = \left[\prod_{i=1}^{m} G(dx_i)\right] \prod_{j=1}^{n} \int_{y_j \le z} z^{-1} G(dz). \tag{1.2}$$

Vardi (1989) derived the NPMLE $\hat{G}_{m,n}$ which maximizes this likelihood. He also showed that the above model unifies several well studied statistical problems, which we briefly describe below. An elegant discussion of a related line-segment model was given by Laslett (1982).

Nonparametric estimation in renewal processes. Suppose we observe $A_k, R_k \wedge T_k, \delta_k = I\{R_k \le T_k\}, 1 \le k \le N$, where A_k and R_k are ages and residual lifetimes of N iid stationary renewal processes with an underlying distribution function F^0, and T_k are censoring times. Let x_i, $1 \le i \le m$, be those uncensored lifetimes $A_k + R_k$ with $\delta_k = 1$, and y_j, $1 \le j \le n$, be the censored lifetimes $A_k + T_k$ with $\delta_k = 0$. Then the likelihood of the data is

$$\left[\prod_{i=1}^{m} \frac{dF^0(x_i)}{\mu^0}\right] \prod_{j=1}^{n} \frac{\bar{F}^0(y_j -)}{\mu^0},$$

where $\bar{F}^0 = 1 - F^0$ and μ^0 is the mean of F^0. The problem of maximizing this likelihood is equivalent to that of maximizing (1.2) under the transformation $G(dx) = x\,dF^0(x)/\mu^0$ and $\bar{F}^0(y)/\mu^0 = \int_{y<z} z^{-1}G(dz)$.

Nonparametric deconvolution. In a deconvolution problem, we are interested in estimating the distribution of a random variable Z^0 based on a sample Y_j^0, $1 \le j \le n$, from the population $Y^0 = Z^0 + U^0$, where Z^0 and U^0 are independent. If U^0 is an exponential random variable with the density function e^{-u}, $u > 0$, and we assume no knowledge about the distribution of Z^0, then $\exp(-U^0)$ is uniform (0, 1) and the transformation $Y_i = \exp(-Y_i^0)$ gives us a sample from the density (1.1), where G is the distribution of $Z = \exp(-Z^0)$.

Estimating a decreasing density. This well known problem was originally studied by Grenander (1956). A nonnegative random variable has a decreasing density on $(0, \infty)$ if and only if (iff) it is a product of a nonnegative random variable and an independent uniform (0, 1) random variable.

1.2. *Organization of the paper and summary of results.* The NPMLE $\hat{G}_{m,n}$ is a consistent estimator in the sense that $\hat{G}_{m,n} \to_{\mathscr{D}} G$ (a.s.) as $(m + n) \to \infty$. This can be proved by the methods of Wang (1985) and Pfanzagl (1988). The uniform consistency of $\hat{G}_{m,n}$ is proved in Section 2. Specifically, we show that if m and n approach infinity such that $m^2/(m + n) \to \infty$, then $\sup_t|\hat{G}_{m,n}(t) - G(t)| \to 0$ in probability. Apparently, the condition $m^2/(m + n) \to \infty$ cannot be completely omitted due to the difference between convergence in distribution and uniform convergence when the limit G is not a continuous function. For example, if G puts unit mass at 1 and $m = 0$, then the NPMLE $\hat{G}_{0,n}$ is induced from the Grenander estimate, which puts no mass at 1, so that $\sup_t|\hat{G}_{0,n}(t) - G(t)| = 1$ (a.s.). In Section 3 we derive the weak convergence of $\hat{G}_{m,n}$. In particular we show that if $m, n \to \infty$ such that $m/(m+n) \to p > 0$, then $\sqrt{m+n}(\hat{G}_{m,n} - G)$ converges weakly to a Gaussian process. This limiting Gaussian process can be described as an inverse of a bounded linear operator (in function space) applied to a mixture of $B_X(G(t))$ and an integration of $B_Y(F(\cdot))$, where B_X and B_Y are two independent Brownian bridges.

It is well known that estimation is less efficient and sometimes impossible if the data are dominated by censored observations, so that the relationship between the sample size of the censored data (n) and that of complete data (m) under which asymptotic theorems for $\hat{G}_{m,n}$ hold is very important in assessing the usefulness of the estimator in practice. For instance, the condition $m^2/(m+n) \to \infty$, besides the obvious possibility of $m \to \infty$ and n bounded (which is in agreement with our intuition of the need to keep the size of the Y sample small), allows the Y sample to be considerably larger in size than the X sample. This is true as long as the size of the X sample is also large. As an example take $m = N^{1/3}$ and $n = N^{1/2}$ for some large N, so that as $N \to \infty$ the proportion of the X's in the data approaches zero $[m/(m+n) \to 0]$, but $m^2/(m+n) \to \infty$ and the uniform consistency holds. The issue of the relative sizes of the two samples is interesting beyond the mathematical aspect, because unlike many censored data problems, here G can be consistently estimated solely on the basis of the Y's. Such an estimate, however, would be $n^{1/3}$-consistent for each t [Prakasa Rao (1969); Groeneboom (1985)] via its connection to the Grenander estimate [Vardi (1989)], as compared to the $m^{1/2}$ consistency of an estimate based solely on the X's. Thus, it is not clear what the asymptotic contribution of the Y sample to the $(m+n)^{1/2}$ consistency of $\hat{G}_{m,n}$ is. In Section 4 we present the asymptotic efficiency of $\hat{G}_{m,n}$ within a class of "regular" estimators, which shows that there is a genuine gain in asymptotic efficiency when the NPMLE $\hat{G}_{m,n}$ is used instead of the empirical cumulative distribution function (ECDF) based on the X sample alone. (This gain cannot be achieved by a linear combination of the NPMLE's based on individual samples.) Our approach to the asymptotic efficiency is simple but nonstandard, and the class of estimators we consider is slightly smaller than a usual one in a standard theorem.

Section 5 contains miscellaneous discussions. For ease of reading we postpone some more involved parts of our proofs to the Appendix.

2. Uniform consistency of $\hat{G}_{m,n}$. The key steps in the development of the asymptotic behavior of $\hat{G} \equiv \hat{G}_{m,n}$ in this section (as well as in Section 3) are (2.5) and Lemma 1. The basic approach is first to use the maximum likelihood score equation (2.3) to express

$$U_{m,n} \equiv \sqrt{m+n}\left(\hat{G}_{m,n} - G\right) \tag{2.1}$$

as an implicit linear function of the empirical processes

$$W_{X,m} \equiv \sqrt{m}\,(G_m - G) \quad \text{and} \quad W_{Y,n} \equiv \sqrt{n}\,(F_n - F), \tag{2.2}$$

with G_m and F_n being the ECDF's of $x_1, \dots, x_m$ and $y_1, \dots, y_n$, respectively.

This functional relation is given in (2.5). The second step is to show that this equation uniquely determines the behavior of $U_{m,n}$ in terms of the behavior of $W_{X,m}$ and $W_{Y,n}$. The final step in the analysis is then to use the weak convergence of $W_{X,m}$ and $W_{Y,n}$ to Brownian bridges in conjunction with (2.5) and Lemma 1 to establish the uniform consistency of $\hat{G}_{m,n}$ (in this section) and the weak convergence of $U_{m,n}$ (in Section 3). The details now follow.

Let $t_1 < \cdots < t_h$ be the distinct values of $x_1, \dots, x_m$ and $y_1, \dots, y_n$. The NPMLE must satisfy the score equation

$$d\hat{G}(t) = \frac{m}{m+n}\, dG_m(t) + \frac{n}{m+n}\int_{0<y\le t} \frac{dF_n(y)}{\int_{y\le z} z^{-1}\, d\hat{G}(z)} t^{-1}\, d\hat{G}(t),$$

subject to $\sum_{j=1}^{h} d\hat{G}(t_j) = 1$ and $dG(t_h) > 0$. This equation can be obtained by standard methods such as differentiating the logarithm of the likelihood function (1.2) with respect to G [cf. Gill (1989)] or taking the limit in the EM algorithm of Vardi [(1989), Equation (2.10)]. Integrating both sides of this score equation, we have

$$\begin{aligned}\hat{G}(t) &= \frac{m}{m+n} G_m(t) \\ &\quad + \frac{m}{m+n}\int_{0<x\le t}\left[\int_{0<y\le x} \frac{dF_n(y)}{\int_{y\le z} z^{-1}\, d\hat{G}(z)}\right] x^{-1}\, d\hat{G}(x).\end{aligned} \tag{2.3}$$

The asymptotics developed in this paper are entirely based on this equation and therefore hold for all its solutions.

Recall from (1.1) that $f(y) = F'(y) = \int_{y<z} z^{-1}\, dG(z)$ and define

$$\hat{f}(y) = \hat{f}_{m,n}(y) = \int_{y<z} z^{-1}\, d\hat{G}_{m,n}(z). \tag{2.4}$$

Note that $\hat{f}$ is a nonincreasing function and it is defined to be right-continuous to agree with the right continuity of CDF's in general, as they appear together in subsequent equations [e.g., (2.5)]. Denoting $m/(m+n)$ by $\hat{p}$, we get as a first order approximation to (2.3),

$$\begin{aligned}&\hat{G}(t) - G(t) \\ &\quad\approx \hat{p}(G_m - G)(t) + (1-\hat{p})\int_{0<x\le t}\int_{0<y\le x} \frac{d(F_n - F)(y)}{f(y)} x^{-1}\, dG(x) \\ &\qquad + (1-\hat{p})(\hat{G} - G)(t) \\ &\qquad + (1-\hat{p})\int_{0<x\le t}\int_{0<y\le x} \frac{\int_{x>y} z^{-1}\, d(\hat{G} - G)(z)}{f^2(y)}\, dF(y) x^{-1}\, dG(x),\end{aligned}$$

which is a nice linear equation in $\hat{G} - G$. A more careful treatment of the preceding derivation (in the Appendix) yields our key equation

$$\frac{m}{m+n}U_{m,n}(t) + \frac{n}{m+n}\hat{f}(t)\int_{0<y\le t} y\int_{y<z} z^{-2}U_{m,n}(z)\,dz\,d\frac{1}{\hat{f}(y)} = W_{m,n}(t), \tag{2.5}$$

where the integration is defined to be 0 for $t \ge t_h$ and the process $W_{m,n}$ is defined by

$$W_{m,n}(t) = \sqrt{\frac{m}{m+n}}\,W_{X,m}(t) + \sqrt{\frac{n}{m+n}}\,\hat{f}(t)\int_{0<y\le t} W_{Y,n}(y)\,d\frac{1}{\hat{f}(y)}, \tag{2.6}$$

with $W_{X,m}$ and $W_{Y,n}$ being the empirical processes in (2.2). Note that for a fixed t, $\hat{f}(t)/\hat{f}(y)$ is nondecreasing in y and bounded by 1 on $(0, t]$, so that the integrations in (2.5) and (2.6) are well defined. In addition, by (2.6) the process $W_{m,n}$ is an explicit linear function of empirical processes and bounded in absolute value by

$$|W_{m,n}(t)| \le \sqrt{\frac{m}{m+n}}\,|W_{X,m}(t)| + \sqrt{\frac{n}{m+n}}\sup_{0<y\le t}|W_{Y,n}(y)|. \tag{2.7}$$

In (2.5) the process $W_{m,n}$ is expressed as the image of a linear operator applied to $U_{m,n}$, say $R_{m,n}U_{m,n} = W_{m,n}$. By inverting this operator $R_{m,n}$, we obtain $U_{m,n} = R_{m,n}^{-1}W_{m,n}$, which suggests a limiting Gaussian process for $U_{m,n}$ of the form $U = R^{-1}W$. The invertibility of $R_{m,n}$ is essentially established in Lemma 1 and explicitly given as linear operators from and to Banach spaces of functions in Section 3, where we also prove the convergence of $W_{m,n}$ and justify the passage of limit from $R_{m,n}^{-1}$ to R^{-1}. For now, we state Lemma 1 and show in Theorem 1 the uniform consistency of $\hat{G}_{m,n}$.

LEMMA 1. *Let $h(t)$ be a nonnegative, right-continuous nonincreasing function defined on $(0,\infty)$. Also let $u(t)$ be a function on $(0,\infty)$ which is right-continuous with left limits, satisfying $\int_{0<y\le t} y\int_{y<z} z^{-2}|u(z)|\,dz\,d[h(y)]^{-1} < \infty$ for $h(t) > 0$ (inner integral is Lebesgue's, outer integral is Stieltjes'). For $0 < p \le 1$, define*

$$w(t) = pu(t) + (1-p)h(t)\int_{0<y\le t} y\int_{y<z} z^{-2}u(z)\,dz\,d(h(y))^{-1}, \tag{2.8}$$

where the integration is defined to be 0 if either $t = 0$ or $h(t) = 0$. [Compare with (2.5).] Then the following hold:

(i) *If $w(\cdot) \equiv 0$, then $u(\cdot) \equiv 0$.*

(ii) *If $u(0+) = u(0) = 0$, then*

$$\sup_{0\le t<T}|u(t)| \le \frac{2}{p}\max\left(\sup_{0\le t<T}|w(t)|, \frac{1-p}{p}\sup_{t\ge T}|w(t)|\right) \tag{2.9}$$

and

$$\sup_{t \ge 0} |u(t)| \le c_p \sup_{t \ge 0} |w(t)| \tag{2.10}$$

with $c_p = (2/p)\max(1, (1-p)/p)$ *and* $T = \inf\{t\colon h(t) = 0\}$.

Lemma 1 is proved in the Appendix. The uniform consistency of the NPMLE follows immediately from (2.5) and Lemma 1 upon substituting in (2.8) $U_{m,n}$ for u, $W_{m,n}$ for w, $\hat{f}_{m,n}$ for h, $m/(m+n)$ for p and applying part (ii) of the lemma. This is stated and explained in the following theorem. [Lemma 1(i) is needed for the proof of Lemma 3, which comes later.]

THEOREM 1. *Let $\hat{G} = \hat{G}_{m,n}$ be the unique NPMLE for the lifetime distribution function G based on samples $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$.*

(i) *If $m/(m+n) \to p > 0$, then*

$$\sup_{t \ge 0} \left| \hat{G}_{m,n}(t) - G(t) \right| \to 0 \quad a.s. \tag{2.11}$$

(ii) *If $m^2/(m+n) \to \infty$, then $\sup_t |\hat{G}_{m,n}(t) - G(t)| \to 0$ in probability.*

PROOF. It follows from (2.5) and (2.10) that

$$\sup_{t \ge 0} \left| \hat{G}_{m,n}(t) - G(t) \right| \le \frac{2(m+n)}{m} \max\left(1, \frac{n}{m}\right) \sup_{t \ge 0} \left| \frac{W_{m,n}(t)}{\sqrt{m+n}} \right|,$$

which implies (2.11) by (2.7), (2.2) and the uniform consistency of the ECDF's.

For part (ii), we have

$$W_{m,n}(t) = \frac{m}{\sqrt{m+n}} \left(\hat{G}_{m,n}(t) - G(t) \right) = \frac{m}{\sqrt{m+n}} (1 - G(t)), \qquad t \ge t_h,$$

so that by (2.5) and (2.9),

$$\begin{aligned} &\sup_{0 \le t < t_h} \left| \hat{G}_{m,n}(t) - G(t) \right| \\ &\quad \le \frac{2\sqrt{m+n}}{m} \max\left\{ \sup_{0 \le t < t_h} |W_{m,n}(t)|, \frac{n}{\sqrt{m+n}} (1 - G(t_h)) \right\}. \end{aligned} \tag{2.12}$$

Since $t_h \ge x_m^* = \max(x_1, \dots, x_m)$,

$$1 - G(t_h) \le m^{-1} m (1 - G(x_m^*)) = m^{-1} O_p(1).$$

Hence, by (2.12), (2.7), (2.2) and the stochastic boundedness of the empirical processes,

$$\begin{aligned} &\sup_{t \ge 0} \left| \hat{G}_{m,n}(t) - G(t) \right| \\ &\quad \le \sup_{0 \le t < t_h} \left| \hat{G}_{m,n}(t) - G(t) \right| + 1 - G(t_h) \\ &\quad \le 2m^{-1} \sqrt{m+n} \max\left\{ O_p(1), n(m+n)^{-1/2} m^{-1} O_p(1) \right\} + o_p(1) \\ &\quad = o_p(1) \quad \text{as } m^2/(m+n) \to \infty. \end{aligned}$$

□

3. Weak convergence of $\sqrt{m+n}\,(\hat{G}_{m,n} - G)$ to a Gaussian process. Let $D_0[0,\infty]$ be the Banach space of all functions $u(\cdot)$ on $[0,\infty)$ that are right-continuous with left limits, converge to 0 at ∞ and vanish at 0:

$$D_0[0,\infty] = \{u\colon u(t+) = u(t),\, u(t-) \text{ exists}, \forall\, t;$$
$$u(t) \to 0 \text{ as } t \to \infty;\, u(0) = 0\},$$

with the sup norm $\|u\| = \sup_{0 \le t < \infty}|u(t)|$. This Banach space is an isomorphism of the subspace $\{u \in D[0,1],\, u(0) = 0,\, u(1) = u(1-) = 0\}$ of the usual $D[0,1]$ space via a one-to-one transformation $u(t) \to u(t/(1+t))$. We shall show that under the condition $m/(m+n) \to p > 0$, the processes $U_{m,n}$ in (2.1) converge weakly to a Gaussian process in the space $D_0[0,\infty]$. Let $\mathscr{B}$ be the σ algebra generated by all closed balls in $D_0[0,\infty]$. Here a stochastic process in $D_0[0,\infty]$ is a $\mathscr{F}/\mathscr{B}$ measurable mapping from a probability space $(\Omega, \mathscr{F}, P)$ to $D_0[0,\infty]$, and the weak convergence of $U_{m,n}$ to U in $D_0[0,\infty]$ means $E\varphi(U_{m,n}) \to E\varphi(U)$ for all bounded continuous real functions φ on $D_0[0,\infty]$ which are measurable with respect to $\mathscr{B}$ [cf. Pollard (1984), page 65]. As discussed in the paragraph preceding Lemma 1, the method of proof is to consider the left side of (2.5) as a linear operator applied to $U_{m,n}$ and to invert this linear operator.

For any nonnegative right-continuous nonincreasing function $h(\cdot)$ on $(0,\infty)$ define linear operators A_h and $\bar{A}_h$ from $D_0[0,\infty]$ to $D_0[0,\infty]$ by

$$(A_h u)(t) = h(t)\int_{0<y\le t} y\int_{y<z} z^{-2}u(z)\,dz\,d\frac{1}{h(y)}, \tag{3.1}$$

$$(\bar{A}_h u)(t) = h(t)\int_{0<y\le t} u(y)\,d\frac{1}{h(y)} \tag{3.2}$$

[cf. the integrations in (2.5) and (2.6)]. Here and in the sequel we use the convention that

$$h(t)\int_{0<y\le t} v(y)\,d\frac{1}{h(y)} = 0 \quad \text{if } t = 0 \text{ or } h(t) = 0, \forall\, v(\cdot).$$

Let f and $\hat{f}_{m,n}$ be the density of Y_1 in (1.1) and its right-continuous NPMLE in (2.4), respectively, and define via (3.1) and (3.2),

$$A_{m,n} \equiv A_{\hat{f}_{m,n}}, \qquad A \equiv A_f, \qquad \bar{A}_{m,n} \equiv \bar{A}_{\hat{f}_{m,n}}, \qquad \bar{A} \equiv \bar{A}_f, \tag{3.3}$$

$$R_{m,n} = \frac{m}{m+n}I + \frac{n}{m+n}A_{m,n}, \qquad R = pI + (1-p)A, \tag{3.4}$$

where I is the identity operator, $Iu = u$. Then (2.5) and (2.6) become

$$R_{m,n}U_{m,n} = W_{m,n} = \sqrt{\frac{m}{m+n}}\,W_{X,m} + \sqrt{\frac{n}{m+n}}\,\bar{A}_{m,n}W_{Y,n} \tag{3.5}$$

and (2.10) of Lemma 1(ii) can be restated as

$$|u| \le c_p\|(pI + (1-p)A_h)u\|, \qquad \forall\, u \in D_0[0,\infty],\, p > 0,\, h, \tag{3.6}$$

where $c_p = (2/p)\max(1, (1-p)/p)$. The limiting process of $U_{m,n}$ is obtained by inverting the operators $R_{m,n}$ in (3.5) and then taking the limit.

THEOREM 2. *Let the processes* $U_{m,n}$ *be given by* (2.1). *Suppose that* $m/(m+n) \to p > 0$. *Then*

(3.7) $U_{m,n}$ *converges weakly to a Gaussian process* $U = R^{-1}W$ *in* $D_0[0,\infty]$,

where R^{-1}, *a bounded linear operator frcm* $D_0[0,\infty]$ *to* $D_0[0,\infty]$, *is the inverse of* R *in* (3.4), *and for some independent Brownian bridge processes* B_X *and* B_Y,

$$(3.8)\quad W(t) = \sqrt{p}\,B_X(G(t)) + \sqrt{1-p}\,f(t)\int_{0<y\le t} B_Y(F(y))\,d\frac{1}{f(y)}.$$

[*Note*: *The right side of* (3.8) *is the limiting version of the right side of* (3.5).]

PROOF. The proof has two parts.

(i) Finding the limit of $W_{m,n}$: By the asymptotic theory of ECDF's there exist Brownian bridge processes $B_{X,m}$, $B_{Y,n}$, $m \ge 1$, $n \ge 1$, such that $B_{X,m}$ is independent of $B_{Y,n}$ and

$$(3.9)\quad \|W_{X,m} - B_{X,m}\circ G\| + \|W_{Y,n} - B_{Y,n}\circ F\| \to 0, \quad \text{in probability,}$$

where $(u\circ v)(t) = u(v(t))$ for any functions u and v. Let $C_0[0,b]$ be the Banach space of all continuous functions on $[0,b]$ which vanish at 0 and b with the sup norm. Since $B_{X,m}$, $B_{Y,n}$, $m \ge 1$, $n \ge 1$, are tight, for any $\varepsilon > 0$ there exists a compact set K in $C_0[0,1]$ such that

$$P\{B_{X,m} \in K, B_{Y,n} \in K\} \ge 1 - \varepsilon, \qquad \forall\, m, n,$$

so that by the continuity of $F(\cdot)$ there exists a compact set K_1 in $C_0[0,\infty]$ satisfying $P\{B_{Y,n}\circ F \in K_1\} \ge 1-\varepsilon$ for every n.

LEMMA 2. *Let the operators* $\bar{A}_{m,n}$, $\bar{A}$, $A_{m,n}$, A, $R_{m,n}$ *and* R *be defined by* (3.3) *and* (3.4). *Then* $\|\bar{A}_{m,n}\| \le 1$, $\|A_{m,n}\| \le 1$, $\|R_{m,n}\| \le 1$,

$$\left\|\bar{A}_{m,n}u - \bar{A}u\right\| \to 0 \quad a.s., \forall\, u \in C_0[0,\infty]$$

and

$$\|A_{m,n}u - Au\| \to 0 \quad a.s., \qquad \|R_{m,n}u - Ru\| \to 0 \quad a.s., \qquad \forall\, u \in D_0[0,\infty].$$

Lemma 2 is proved in the Appendix. Since K_1 is compact, it follows that $\sup_{w\in K_1}\|\bar{A}_{m,n}w - \bar{A}w\| \to 0$ a.s., which implies by (3.5), (3.9) and the boundedness of $\|\bar{A}_{m,n}\|$ that

$$(3.10)\qquad \|W_{m,n} - V_{m,n}\| \to 0 \quad \text{in probability,}$$

where $V_{m,n} = \sqrt{p}\,B_{X,m}\circ G + \sqrt{1-p}\,\bar{A}(B_{Y,n}\circ F)$. Since $\bar{A}K_1$ is compact, for every $\varepsilon > 0$ there exists a compact set K_2 in $D_0[0,\infty]$ such that

$$(3.11)\qquad P\{V_{m,n} \in K_2\} \ge 1-\varepsilon, \qquad \forall\, m, n.$$

(ii) Taking the limit of the inverse of the operator:

LEMMA 3. *Suppose* $0 < p \le 1$ *and* $m \ge 1$. *Then the linear operators* $R_{m,n}$ *and* R *are one-to-one linear mappings from* $D_0[0,\infty]$ *onto* $D_0[0,\infty]$ *with* $\|R_{m,n}^{-1}\| \le 2(m+n)^2/m^2$ *and* $\|R^{-1}\| \le 2/p^2$.

Lemma 3 is also proved in the Appendix. Let $K_3 = \{R^{-1}v\colon v \in K_2\}$ with K_2 in (3.11). Then K_3 is compact in $D_0[0,\infty]$ by Lemma 3 and $\sup_{u \in K_3}\|R_{m,n}u - Ru\| \to 0$, a.s., by Lemma 2. Hence, on the event $V_{m,n} \in K_2$, by (3.10) and Lemma 3, as $m/(m+n) \to p > 0$,

$$\begin{aligned}
&\|U_{m,n} - R^{-1}V_{m,n}\| \\
&\quad = \|R_{m,n}^{-1}W_{m,n} - R^{-1}V_{m,n}\| \\
&\quad \le 2(m+n)^2 m^{-2}\|W_{m,n} - V_{m,n}\| + \|R_{m,n}^{-1}V_{m,n} - R^{-1}V_{m,n}\| \\
&\quad = o_p(1) + \|R_{m,n}^{-1}(R_{m,n} - R)R^{-1}V_{m,n}\| \\
&\quad \le o_p(1) + 2(m+n)^2 m^{-2} \sup_{u \in K_3}\|R_{m,n}u - Ru\| \\
&\quad = o_p(1).
\end{aligned}$$

Since this inequality holds for an event whose probability of occurring is arbitrarily close to 1 [$\varepsilon > 0$ is arbitrary in (3.11)] and the processes $V_{m,n}$ and W in (3.8) have the same distribution, the assertion of the theorem follows. □

4. Asymptotic efficiency. In this section we shall demonstrate the asymptotic efficiency of the NPMLE $\hat{G}$. The original Hajék–Le Cam convolution theorem has been generalized to various nonparametric cases by essentially calculating a kernel operator at a fixed underlying probability measure P_G; see for example Beran (1977), Wellner (1982), Begun, Hall, Huang and Wellner (1983) and Millar (1985), among others. Here we shall take a slightly different approach: first proving a finite-dimensional convolution theorem for the case where the underlying distribution G is discrete with finite support, and then taking a smooth transition from the finite-dimensional case to the infinite-dimensional case. This produces a simple proof of the superiority of the NPMLE over all regular estimators whose finite-dimensional limiting distributions are continuous in G.

For any stochastic process $H(\cdot)$ in $D_0[0,\infty]$ we shall denote by $\mathscr{L}(H; G)$ the distribution of $H(\cdot)$ in $D_0[0,\infty]$ under the probability P_G and by $\mathscr{L}(H; G, s_1, \dots, s_k)$ the k-dimensional joint distribution of $H(s_1), \dots, H(s_k)$ under P_G. Assume that $m/(m+n) \to p > 0$. Let μ be a measure on $(-\infty, \infty)$ with respect to which the distribution G has a density g. Let $\mathscr{C}(G)$ be the set of all sequences of distribution functions $\{G_{m,n}\}$ such that the density $g_{m,n} = dG_{m,n}/d\mu$ exists and

$$\sqrt{m+n}\left(g_{m,n}^{1/2} - g^{1/2}\right) \text{ converges in } L^2(\mu). \tag{4.1}$$

As in Beran (1977), we shall say that a sequence of estimators $\tilde{G} = \tilde{G}_{m,n}$ based on X and Y samples of sizes m and n is regular if

$$\mathscr{L}\left(\sqrt{m+n}\left(\tilde{G}_{m,n} - G\right); G_{m,n}\right) \to_{\mathscr{D}} \mathscr{L}(\tilde{U}; G) \tag{4.2}$$

for all sequences $\{G_{m,n}\} \in \mathscr{C}(G)$, where $\tilde{U}$ is a stochastic process in $D_0[0,\infty]$ such that the distribution $\mathscr{L}(\tilde{U}; G)$ in (4.2) depends only upon G and not upon the choice of the sequence $\{G_{m,n}\} \in \mathscr{C}(G)$ that determines the sampling scheme. Since the functional $\varphi_t(h) = \arctan(h(t))$ is bounded and continuous on $D_0[0,\infty]$ and measurable with respect to the σ algebra $\mathscr{B}$ generated by all closed balls, (4.2) implies

$$\mathscr{L}\left(\sqrt{m+n}\left(\tilde{G}_{m,n} - G_{m,n}\right); G_{m,n}, s_1, \dots, s_k\right) \to_{\mathscr{D}} \mathscr{L}\left(\tilde{U}; G, s_1, \dots, s_k\right) \tag{4.3}$$

for all $s_1, \dots, s_k$.

THEOREM 3. *Suppose that $m/(m+n) \to p > 0$. Let $\tilde{G}$ be a sequence of regular estimators with a limiting distribution $\mathscr{L}(\tilde{U}; G)$ whose finite-dimensional distributions $\mathscr{L}(\tilde{U}; G, s_1, \dots, s_k)$ are continuous in G under the* sup-*norm topology for G. Then there exists a stochastic process $H(\cdot)$ in $D_0[0,\infty]$ such that*

$$\mathscr{L}(\tilde{U}; G) = \mathscr{L}(H; G) * \mathscr{L}(U; G),$$

where U is as in Theorem 2 and $*$ denotes the convolution.

A sequence of estimators $\tilde{G}$ satisfies the conditions of Theorem 3 if

$$\|G_{m,n} - G\| \to 0 \Rightarrow \mathscr{L}\left(\sqrt{m+n}\left(\tilde{G} - G_{m,n}\right); G_{m,n}\right) \to_{\mathscr{D}} \mathscr{L}(\tilde{U}; G),$$

since (4.1) implies $\|G_{m,n} - G\| \to 0$. Our argument in Sections 2 and 3 can be slightly modified to show that the NPMLE $\hat{G}$ possesses this property.

PROOF. There are two steps.

(i) Finite-dimensional case: Suppose G is a discrete distribution function with a finite support $\{s_1, \dots, s_k\}$. Let $\hat{G}_{\mathrm{ML}}$ be the maximum likelihood estimate (MLE) of G based on the observations and the extra knowledge of the support $\{s_1, \dots, s_k\}$. Then $\hat{G}_{\mathrm{ML}}$ is a solution of (2.3), provided that $t_h = s_k^*$, where $s_k^* = \max(s_1, \dots, s_k)$. Since our derivation of the limiting processes in Sections 2 and 3 holds for all solutions of (2.3) (e.g., $\hat{G}_{\mathrm{ML}} I\{t_h = s_k^*\} + \hat{G} I\{t_h \neq s_k^*\}$) and $P\{t_h = s_k^*\} \to 1$,

$$\mathscr{L}\left(\sqrt{m+n}\left(\hat{G}_{\mathrm{ML}} - G\right); G\right) \to_{\mathscr{D}} \mathscr{L}(U; G). \tag{4.4}$$

Therefore, by (4.3) and the Hajék–Le Cam convolution theorem for the MLE in the finite-dimensional parametric case, there exists a k-dimensional joint distribution $\mu_{s_1,\dots,s_k}$ such that

$$\mathscr{L}\left(\tilde{U}; G, s_1, \dots, s_k\right) = \mu_{s_1,\dots,s_k} * \mathscr{L}(U; G, s_1, \dots, s_k). \tag{4.5}$$

Note that the support of G is assumed to be $\{s_1, \dots, s_k\}$ here.

(ii) Transition to the infinite-dimensional case: For an arbitrary fixed G, let s_k, $k \geq 1$, be a sequence of distinct positive numbers dense in $(0, \infty)$, and $G_{(k)}$, $k \geq 1$, be distribution functions such that the support of $G_{(k)}$ is $\{s_1, \dots, s_k\}$ and $\|G_{(k)} - G\| \to 0$. Then by (4.5)

$$\mathscr{L}\left(\tilde{U}; G_{(k)}, s_1, \dots, s_j\right) = \mu_{jk} * \mathscr{L}(U; G_{(k)}, s_1, \dots, s_j), \qquad j \leq k,$$

where μ_{jk} is the marginal distribution of $\mu_{s_1, \dots, s_k}$ for the first j coordinates. Letting $k \to \infty$ by the continuity of $\mathscr{L}(U; G, s_1, \dots, s_j)$ and $\mathscr{L}(\tilde{U}; G, s_1, \dots, s_j)$ in G, we can find a j-dimensional distribution μ_j such that

$$\mathscr{L}\left(\tilde{U}; G, s_1, \dots, s_j\right) = \mu_j * \mathscr{L}(U; G, s_1, \dots, s_j). \tag{4.6}$$

Let $s_{(1)} < \cdots < s_{(j)}$ be the ordered values of $s_1, \dots, s_j$, and $s_{(0)} = 0$. Define stochastic processes

$$U_{(j)}(s) = U(s_{(i)}), \qquad \tilde{U}_{(j)}(s) = \tilde{U}(s_{(i)}), \qquad s_{(i)} \leq s < s_{(i+1)}, \qquad 0 \leq i < j,$$

and $U_{(j)}(s) = \tilde{U}_j(s) = 0$ for $s \geq s_{(j)}$. Then by (4.6) there exist stochastic processes $H_{(j)}$ independent of the process U such that

$$\mathscr{L}\left(\tilde{U}_{(j)}; G\right) = \mathscr{L}(H_{(j)} + U_{(j)}; G) = \mathscr{L}(H_{(j)}; G) * \mathscr{L}(U_{(j)}; G). \tag{4.7}$$

Since $U_{(j)} \to U$ a.s. and $\tilde{U}_{(j)} \to \tilde{U}$ a.s. in the Skorohod topology, the processes $U_{(j)}, \tilde{U}_{(j)}$, $j \geq 1$, are tight in the Skorohod topology, which implies the tightness of $H_{(j)}$, $j \geq 1$, so that by (4.7) there exists a process H independent of U such that

$$\mathscr{L}(\tilde{U}; G) = \mathscr{L}(H + U; G) = \mathscr{L}(H; G) * \mathscr{L}(U; G)$$

[cf. Billingsley (1968), pages 37, 121 and 123]. □

5. Remarks.

5.1. *The inversion of the operator R.* Although the inverse operator R^{-1} in Theorem 2 is not explicitly given, it can be calculated by multistage infinite series expansions. For simplicity, let us assume $f(t) > 0$ for all t. Then by Lemmas 2 and 1,

$$\|A\| \leq 1, \qquad \left\|(\lambda + A)^{-1}\right\| \leq 2/\lambda, \qquad \lambda = p/(1-p) > 0,$$

so that the following expansions converge in the norm $\|\cdot\|$ of operators

$$(\lambda + A)^{-1} = \lambda^{-1}(I + A/\lambda)^{-1} = \lambda^{-1} \sum_{k=0}^{\infty} (-1)^k (A/\lambda)^k, \qquad \lambda > 1, \tag{5.1}$$

and for $0 < \lambda_j/2 < \lambda_{j+1} \leq \lambda_j$,

$$\begin{aligned} (\lambda_{j+1} + A)^{-1} &= (\lambda_{j+1} - \lambda_j + \lambda_j + A)^{-1} \\ &= (\lambda_j + A)^{-1}\left[I + (\lambda_{j+1} - \lambda_j)(\lambda_j + A)^{-1}\right]^{-1} \\ &= (\lambda_j + A)^{-1} \sum_{k=0}^{\infty} \left[(\lambda_j - \lambda_{j+1})(\lambda_j + A)^{-1}\right]^k. \end{aligned} \tag{5.2}$$

If $\lambda = p/(1-p) > 1$, then $R^{-1} = (p + (1-p)A)^{-1} = (1-p)^{-1}(\lambda + A)^{-1}$ is given by (5.1); otherwise, there exist constants $\lambda_0 > \lambda_1 > \cdots > \lambda_k = p/(1-p)$ with $\lambda_0 > 1$ and $\lambda_j/2 < \lambda_{j+1} \le \lambda_j$, so that we can calculate $(\lambda_0 + A)^{-1}$ by (5.1) and then apply (5.2) successively for $j = 0, \dots, k-1$ to obtain $R^{-1} = (1-p)^{-1}(\lambda_k + A)^{-1}$. In addition, by (3.1) and (3.3),

$$(Au)(t) = \int_0^\infty A(t,x)u(x)\,dx, \qquad A(t,x) = f(t)x^{-2}\int_{0<y\le t\wedge x} y\,d\frac{1}{f(y)},$$

so that the above expansions imply that the operator R^{-1} has the form

$$(R^{-1}u)(t) = p^{-1}u(t) + \int_0^\infty K(t,x)u(x)\,dx \tag{5.3}$$

for a kernel $K(t,x)$ satisfying

$$p^2K(t,x) + (1-p)A(t,x) + p(1-p)\int_0^\infty K(t,z)A(z,x)\,dz = 0,$$

$$\int_0^\infty K(t,z)A(z,x)\,dz = \int_0^\infty A(t,z)K(z,x)\,dz.$$

For example, if $\lambda = p/(1-p) > 1$, then by (5.1),

$$K(t,x) = \frac{1}{1-p}\sum_{k=1}^{\infty}(-1)^k\lambda^{-(k+1)}A^k(t,x),$$

where $A^{k+1}(t,x) = \int_0^\infty A^k(t,y)A(y,x)\,dy$ and $A^1(t,x) = A(t,x)$. The details of the analysis leading to the above conclusions involve infinite series expansions of resolvent operators in Banach spaces [Dunford and Schwartz (1958), page 566].

5.2. *The covariance function.* The covariance function of the Gaussian process $\{U(t),\ 0 \le t\}$ is defined by

$$\psi(s,t) = \psi_G(s,t) = E_G U(s)U(t). \tag{5.4}$$

Since the operator R^{-1} can be calculated by expansions and has the form (5.3) and the covariance function $\varphi(s,t) = EW(s)W(t)$ can be explicitly written using (3.8), the covariance function of $U = R^{-1}W$ in (3.7) can be calculated in principle.

It follows from (3.8) that

$$\begin{aligned}\varphi(s,t) &= p[G(s\wedge t) - G(s)G(t)]\\ &\quad + (1-p)\int_0^s\int_0^t [F(x\wedge y) - F(x)F(y)]\,d_y\frac{f(t)}{f(y)}\,d_x\frac{f(s)}{f(x)}.\end{aligned} \tag{5.5}$$

Integrating by parts, we have

$$\begin{aligned}f(t)\int_{0<y\le t} F(x\wedge y)\,d\frac{1}{f(y)} &= F(x\wedge t) - f(t)\int_{0<y\le t\wedge x}\frac{1}{f(y)}\,dF(y)\\ &= F(x\wedge t) - f(t)(x\wedge t),\end{aligned}$$

so that by (2.2) of Vardi (1989)

$$f(t)\int_{0<y\le t} F(y)\,d\frac{1}{f(y)} = F(t) - tf(t) = G(t)$$

and for $s \le t$,

$$\int_0^s\int_0^t F(x\wedge y)\,d_y\frac{f(t)}{f(y)}\,d_x\frac{f(s)}{f(x)} = \int_{0<x\le s}[F(x) - f(t)x]\,d_x\frac{f(s)}{f(x)}$$
$$= G(s) - f(t)f(s)\int_{0<x\le s} x d\frac{1}{f(x)},$$

which imply by (5.5) that

$$\varphi(s,t) = [G(s\wedge t) - G(s)G(t)] - (1-p)f(s)f(t)\int_{0<x\le s\wedge t} xd[f(x)]^{-1}.$$

Finally, by (5.3) the covariance function of the limiting process U is

$$\psi(s,t) = \frac{1}{p^2}\varphi(s,t) + \frac{1}{p}\int_0^\infty K(s,x)\varphi(x,t)\,dx$$
$$+ \frac{1}{p}\int_0^\infty \varphi(s,x)K(t,x)\,dx + \int_0^\infty\int_0^\infty K(s,x)K(t,y)\varphi(x,y)\,dx\,dy.$$

We can also consistently estimate the covariance function $\psi_G(s,t)$ through its continuity in G. If G has a finite discrete support, then by (4.4) the limiting distribution of the NPMLE is the same as that of MLE with the extra knowledge of the support set, so that $\psi_G(s,t)$ in (5.4) can be calculated by inverting the Fisher information matrix for the finite-dimensional parametric case. In particular, by the discreteness of $\hat{G}_{m,n}$, this provides us with a method of calculating $\psi_{\hat{G}_{m,n}}(s,t)$ which can be used to estimate $\psi_G(s,t)$. Since the covariance function is continuous in G and $\|\hat{G}_{m,n} - G\| \to 0$, we have the consistency $\psi_{\hat{G}_{m,n}}(s,t) \to \psi_G(s,t)$ a.s.

APPENDIX

PROOF OF (2.5). Since

$$\frac{1}{\sqrt{n}}\hat{f}(t)\int_{0<y\le t} W_{Y,n}(y)\,d\hat{f}^{-1}(y)$$
$$= -\int_{0<y\le t}(F_n(y) - F(y))d\left(1 - \frac{\hat{f}(t)}{\hat{f}(y)}\right)$$
$$= \int_{0<y\le t}\left(1 - \frac{\hat{f}(t)}{\hat{f}(y-)}\right)d(F_n(y) - F(y)) \quad \text{(integrating by parts)}$$
$$= \int_{0<y\le t}\int_{y\le x\le t} x^{-1}\,d\hat{G}(x)\hat{f}^{-1}(y-)d(F_n(y) - F(y)) \quad \text{[by (2.4)]}$$
$$= \int_{0<x\le t}\int_{0<y\le x}\frac{d(F_n(y) - F(y))}{\int_{y\le z} z^{-1}\,d\hat{G}(z)}x^{-1}\,d\hat{G}(x) \quad \text{(change order)},$$

it follows that for $t < t_h$,

$$\left(\frac{m}{m+n}U_{m,n}(t) - W_{m,n}(t)\right)\Big/\sqrt{m+n}$$

$$= \frac{m}{m+n}\left(\hat{G}(t) - G_m(t)\right) - \frac{\sqrt{n}}{m+n}\hat{f}(t)\int_{0<y\le t} W_{Y,n}(t)\, d\hat{f}^{-1}(y)$$

[by (2.1) and (2.6)]

$$= \frac{m}{m+n}\left(\hat{G}(t) - G_m(t)\right)$$

$$- \frac{n}{m+n}\int_{0<x\le t}\int_{0<y\le x}\frac{d(F_n(y) - F(y))}{\int_{y\le z} z^{-1}\, d\hat{G}(z)} x^{-1}\, d\hat{G}(x)$$

$$= \frac{n}{m+n}\int_{0<x\le t}\left[\int_{0<y\le x}\frac{dF(y)}{\int_{y\le z} z^{-1}\, d\hat{G}(z)} - x\right] x^{-1}\, d\hat{G}(x) \quad \text{[by (2.3)]}$$

$$= \frac{-n}{(m+n)^{3/2}}\int_{0<x\le t}\int_{0<y<x}\frac{\int_{y<z} z^{-1}\, dU_{m,n}(z)}{\int_{y<z} z^{-1}\, d\hat{G}(z)}\, dy\, x^{-1}\, d\hat{G}(x)$$

[by (2.1) and (1.1)]

$$= \frac{-n}{(m+n)^{3/2}}\int_{0<y\le t}\int_{y<x\le t} x^{-1}\, d\hat{G}(x)\frac{\int_{y<z} z^{-1}\, dU_{m,n}(z)}{\hat{f}(y)}\, dy$$

(change order).

Since

$$\frac{d}{dy}\left(y\int_{y<z} z^{-2}U_{m,n}(z)\, dz\right) = -\int_{y<z} U_{m,n}(z)\, dz^{-1} - y^{-1}U_{m,n}(y)$$

$$= \int_{y<z} z^{-1}\, dU_{m,n}(z) \quad \text{(integrating by parts)},$$

the above equation can be written as

$$\frac{m}{m+n}U_{m,n}(t) - W_{m,n}(t)$$

$$= \frac{-n}{m+n}\int_{0<y\le t}\left(1 - \frac{\hat{f}(t)}{\hat{f}(y)}\right) d\left(y\int_{y<z}\frac{U_{m,n}(z)}{z^2}\, dz\right) \quad \text{[by (2.4)]}$$

$$= \frac{-n}{m+n}\int_{0<y\le t} y\int_{y<z}\frac{U_{m,n}(z)}{z^2}\, dz\, d_y\frac{\hat{f}(t)}{\hat{f}(y)} \quad \text{(integrating by parts)}.$$

Hence, we have (2.5). □

Proof of Lemma 1. We shall only prove part (ii) since the proof of part (i) is similar and easier. Since $|u(t)| = |w(t)/p| \le c_p|w(t)|$ for $t \ge T$, (2.10) follows immediately from (2.9). Now, let $t_0 \in (0, T)$ be a point with $u(t_0) > 0$ and M be a constant such that

$$\max\Big(\sup_{0 \le t < T} |w(t)|, \sup_{t \ge T} |w(t)|(1-p)/p \Big) < M.$$

Our goal here is to show that $pu(t_0) \le 2M$. We shall treat the following two cases separately.

CASE a. For some $0 \le t_1 \le t_0$, $u(t) \ge 0$ for all $t \ge t_1$ and $u(t_1 -) \le 0$ if $t_1 > 0$. Let

$$v(t) = \frac{pu(t) - w(t)}{h(t)} = -(1-p)\int_{0<y\le t} y \int_{y<z} z^{-2} u(z)\, dz\, d\frac{1}{h(y)}.$$

Then $v(t_0) \le v(t_1 -)$, since the right-hand side of the above equation is decreasing from $t_1 -$ to t_0. It follows that

$$pu(t_0) \le w(t_0) + h(t_0)v(t_1 -) \le w(t_0) + |w(t_1 -)| \le 2M,$$

which proves the result for Case a.

CASE b. For some $0 \le t_1 \le t_0 < t_2 < \infty$,

(A1) $\quad u(t_1 -) \le 0$ if $t_1 > 0$, $\quad u(t) \ge 0 \quad \forall\, t_1 \le t < t_2$, $\quad u(t_2) \le 0$.

We shall assume $pu(t_0) > 2M$ and establish a contradiction. For $t < T$ set

$$v(t) = \frac{pu(t) - w(t) - M}{h(t)}$$

$$= -Mh^{-1}(t) - (1-p)\int_{0<y\le t} y \int_{y<z} z^{-2}u(z)\, dz\, d\frac{1}{h(y)}.$$

Then $v(t_1 -) \le 0$ if $t_1 > 0$, $v(0+) \le 0$ and $v(t_0) > 0$, so that we can find a positive constant $y_1 \in [t_1, t_0]$ such that $dv(y_1) > 0$. Since $d[1/h(y)] \ge 0$ and

$$dv(y) = -\left[M + (1-p)y\int_{y<z} z^{-2}u(z)\, dz\right] d\frac{1}{h(y)},$$

we have

(A2) $\quad M + (1-p)y_1\int_{y_1<z} z^{-2}u(z)\, dz < 0 \quad$ for some $t_1 \le y_1 \le t_0$.

If $t_2 < T$, then $v(t_2) \le 0$ and the above argument also leads to

(A3) $\quad M + (1-p)y_2\int_{y_2<z} z^{-2}u(z)\, dz > 0 \quad$ for some $t_0 \le y_2 \le t_2$.

For the case $t_2 \ge T$, we have

$$M + (1-p)T\int_{T<z} z^{-2}u(z)\, dz \ge M - \frac{1-p}{p}\sup_{z \ge T}|w(z)| > 0,$$

so that (A3) remains valid for $y_2 = T$. Now, putting (A2) and (A3) together, we have $t_1 \le y_1 \le t_0 < y_2 \le t_2$ and by algebra

$$\frac{1-p}{M}\int_{y_1<z<y_2} z^{-2}u(z)\,dz < -\frac{1}{y_1} + \frac{1}{y_2} = -\frac{y_2-y_1}{y_1y_2} \le 0,$$

which is a contradiction to (A1). Hence, the proof is complete. □

PROOF OF LEMMA 2. The bounds for $\|\bar{A}_{m,n}\|$, $\|A_{m,n}\|$ and $\|R_{m,n}\|$ follow from the monotonicity of $\hat{f}(y)$; see (3.1)–(3.4) and (2.7). Let $\bar{u}(z) = y\int_{y<z} z^{-2}u(z)\,dz$. Then

$$\lim_{y\to 0}\bar{u}(y) = u(0+) = 0 \quad \text{and} \quad \lim_{y\to\infty}\bar{u}(y) = \lim_{z\to\infty}u(z) = 0,$$

so that $\bar{u} \in C_0[0,\infty]$. Since $(A_{m,n} - A)u = (\bar{A}_{m,n} - \bar{A})\bar{u}$ and

$$(R_{m,n} - R)u = \left(\frac{m}{m+n} - p\right)(u - A_{m,n}u) + (1-p)\big(\bar{A}_{m,n} - \bar{A}\big)\bar{u},$$

it suffices for us to show that

$$\big\|\bar{A}_{m,n}u - \bar{A}u\big\| = \big\|\big(\bar{A}_{m,n} - \bar{A}\big)u\big\| \to 0 \quad \text{a.s., } \forall\, u \in C_0[0,\infty].$$

Let $T = \inf\{t\colon f(t) = 0\}$. By (2.4) and (1.1),

$$\hat{f}_{m,n}(y) - f(y) = \int_{y<z} z^{-2}\big(\hat{G}_{m,n}(z) - G(z)\big)\,dz - y^{-1}\big(\hat{G}_{m,n}(y) - G(y)\big),$$

so that by Theorem 1 for any $\varepsilon > 0$ and $f(T_\varepsilon -) > 0$,

$$\varepsilon_{m,n} = \sup_{t\ge\varepsilon}\ \sup_{\varepsilon<y<\min(t,T_\varepsilon)}\left|\frac{\hat{f}_{m,n}(t)}{\hat{f}_{m,n}(y)} - \frac{f(t)}{f(y)}\right| \to 0 \quad \text{a.s.}$$

For any differentiable function $u(\cdot)$ in $C_0[0,\infty]$ satisfying

$$\text{(A4)} \qquad \int|u'(y)|\,dy < \infty, \quad u(t) = 0 \ \text{ for } t < \varepsilon, \quad u'(t) = 0 \ \text{ for } T_\varepsilon < t < T,$$

we have

$$\begin{aligned}
&\big\|\big(\bar{A}_{m,n} - \bar{A}\big)u\big\| \\
&\quad = \sup_{\varepsilon<t<T}\left|\hat{f}_{m,n}(t)\int_{\varepsilon<y\le t} u(y)\,d\frac{1}{\hat{f}_{m,n}(y)} - f(t)\int_{\varepsilon<y\le t} u(y)\,d\frac{1}{f(y)}\right| \\
&\quad = \sup_{\varepsilon<t<T}\left|\int_{\varepsilon<y<\min(t,T_\varepsilon)} u'(y)\left[\frac{\hat{f}_{m,n}(t)}{\hat{f}_{m,n}(y)} - \frac{f(t)}{f(y)}\right]dy\right| \\
&\quad \le \varepsilon_{m,n}\int|u'(y)|\,dy \to 0.
\end{aligned}$$

The set of functions satisfying (A4) for all $\varepsilon > 0$ and $T_\varepsilon < T$ is dense in the set of all differentiable functions in $C_0[0,\infty]$ and therefore dense in $C_0[0,\infty]$. Since $\|(\bar{A}_{m,n} - \bar{A})\| \le 2$, the proof is complete. □

PROOF OF LEMMA 3. Let $R = pI + (1-p)A$. Then $\|R\| \le 1$ and by (3.6),

$$\|u\| \le c_p \|Ru\| \le 2p^{-2}\|Ru\|, \qquad \forall\, u \in D_0[0,\infty].$$

It follows that R is a one-to-one mapping from $D_0[0,\infty]$ onto its range $RD_0[0,\infty]$, the range $RD_0[0,\infty]$ is closed and R^{-1} is a linear operator from $RD_0[0,\infty]$ to $D_0[0,\infty]$ with norm $\|R^{-1}\| \le c_p$. It suffices to prove that $RD_0[0,\infty] = D_0[0,\infty]$. By the Hahn–Banach theorem we only need to show that $\mu = 0$ is the only bounded linear functional μ on $D_0[0,\infty]$ satisfying $\mu(u) = 0$, $\forall\, u \in RD_0[0,\infty]$.

Let μ be a bounded linear functional such that $\mu(u) = 0$, $\forall\, u \in RD_0[0,\infty]$. Then

$$\text{(A5)} \qquad p\mu(u) = -(1-p)\mu(Au), \qquad \forall\, u \in D_0[0,\infty].$$

Since μ is also a bounded linear functional on $C_0[0,\infty]$, there exists a measure μ_0 with support $(0,\infty)$ such that $\mu(u) = \int u(y)\mu_0(dy)$, $\forall\, u \in C_0[0,\infty]$. For any given $u \in D_0[0,\infty]$ there exists a sequence $u_k \in C_0[0,\infty]$ such that $|u_k(t)| \le |u(t)|$, $\forall\, t$, and $u_k \to u$ a.e. in both Lebesgue measure and the measure μ_0, so that

$$\sup_{0 \le y \le \infty} \left| y\int_{y<z} z^{-2}u(z)\,dz - y\int_{y<z} z^{-2}u_k(z)\,dz \right| \to 0 \quad \text{as } k \to \infty,$$

which implies that $\|Au - Au_k\| \to 0$ by the fact that $\|\bar{A}\| \le 1$. It follows that $\mu(Au) = \lim \mu(Au_k)$ and $\mu(u) = \lim \int u_k(y)\mu_0(dy)$ as $k \to \infty$. Therefore, $\mu(u) = \int u(y)\mu_0(dy)$, $\forall\, u \in D_0[0,\infty]$. [Note that this is not true in general for bounded linear functionals on $D_0[0,\infty]$. For example, $\mu(u) = u(b) - u(b-)$.] By (A5) and Fubini's theorem,

$$\begin{aligned} &-p\int u(y)\mu_0(dy) \\ &\quad = (1-p)\int f(t)\int_{0<y\le t} y\int_{y<z} z^{-2}u(z)\,dz\,d[f(y)]^{-1}\mu_0(dt) \\ &\quad = (1-p)\int u(z)\left\{z^{-2}\int_{0<y\le z} y\int_{y\le t} f(t)\mu_0(dt)\,d[f(y)]^{-1}\right\}dz, \end{aligned}$$

which implies that there exists a Lebesgue integrable function $v_0(t)$ such that

$$\mu(u) = \int u(t)\mu_0(dt) = \int u(t)v_0(t)\,dt$$

and

$$\text{(A6)} \quad pv_0(z) + (1-p)z^{-2}\int_{0<y\le z} y\int_{y<t} f(t)v_0(t)\,dt\,d[f(y)]^{-1} = 0, \qquad \forall\, z.$$

Set $u_0(t) = v_0(t)t^2 f(t)$. Then,

$$\int_{0<y\le t} y \int_{y<z} z^{-2} |u_0(z)|\, dz\, d\frac{1}{f(y)}$$
$$\le \int_{0<y\le t} y \int_{y<z} |v_0(z)| f(z)\, dz\, d\frac{1}{f(y)}$$
$$= \int_{0<z<\infty} f(z)|v_0(z)| \int_{0<y\le z\wedge t} y d\frac{1}{f(y)}\, dz \le t\int_0^\infty |v_0(z)|\, dz < \infty,$$

and, by (A6),

$$pu_0(t) + (1-p)f(t)\int_{0<y\le t} y\int_{y<z} z^{-2}u_0(z)\, dz\, df^{-1}(y) = 0, \qquad \forall\, t,$$

so that by Lemma 1(i), $u_0 \equiv 0$ and $v_0 \equiv 0$. Hence $\mu = 0$ and the proof is complete. □

REFERENCES

BEGUN, J. M., HALL, W. J., HUANG, W.-M. and WELLNER, J. A. (1983). Information and asymptotic efficiency in parametric–nonparametric models. *Ann. Statist.* **11** 432–452.

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

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

DUNFORD, N. and SCHWARTZ, J. (1958). *Linear Operators. Part I: General Theory*. Wiley, New York.

GILL, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method, Part I. *Scand. J. Statist.* **16** 97–128.

GRENANDER, U. (1956). On the theory of mortality measurement, Part II. *Scand. Actuar. J.* **39** 125–153.

GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. M. Le Cam and R. A. Olshen, eds.) **2** 539–555. Wadsworth, Belmont, Calif.

LASLETT, G. M. (1982). The survival curve under monotone density constraints with applications to two-dimensional line segment processes. *Biometrika* **69** 153–160.

MILLAR, P. W. (1985). Non-parametric applications of an infinite dimensional convolution theorem. *Z. Wahrsch. Verw. Gebiete* **68** 545–556.

PFANZAGL, J. (1988). Consistency of maximum likelihood estimators for certain nonparametric families, in particular: Mixtures. *J. Statist. Plann. Inference* **19** 137–158.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36.

VARDI, Y. (1989). Multiplicative censoring, renewal process, deconvolution, and decreasing density: Nonparametric estimation. *Biometrika* **76** 751–761.

WANG, J.-L. (1985). Strong consistency of approximate maximum likelihood estimators with applications to nonparametrics. *Ann. Statist.* **13** 932–946.

WELLNER, J. A. (1982). Asymptotic optimality of the product limit estimator. *Ann. Statist.* **10** 595–602.

DEPARTMENT OF STATISTICS
HILL CENTER FOR THE MATHEMATICAL SCIENCES
BUSCH CAMPUS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903