The Annals of Statistics
1982, Vol. 10, No. 3, 772–785

NONPARAMETRIC ESTIMATION IN RENEWAL PROCESSES

BY Y. VARDI[1]

Bell Laboratories

Data collected from many independent identically distributed renewal processes, each of which is observed for an arbitrary period of time, is usually affected by censoring coupled with length biased sampling. In this paper we derive an algorithm that produces the nonparametric maximum likelihood estimator (i.e., the analog of the single-sample empirical distribution function) of the common lifetime distribution, based on such data.

1. Introduction. We consider the problem of finding a maximum likelihood estimate (MLE) of the lifetime distribution, F, on the basis of data collected from independent identically distributed (iid) stationary renewal processes, such that in each process we sample a period of fixed length, whose location is independent of the process itself. Occasionally, we refer to these sampling periods as "windows". To avoid mathematical difficulties we assume that the processes are discrete (the applicability of our methods to the continuous case is remarked upon at the end of the paper) and so the problem is reduced to estimating the underlying lifetime probability function $f = (f(1), f(2), \cdots)$. In Section 2 we show that a nonparametric MLE (NPMLE) of f does not always exist, and we describe explicitly the types of data for which it does, and for which it does not, exist. We further give a simple algorithm that converges monotonically to a lifetime probability function $\hat{f}_M$, that maximizes the likelihood of the data in the set of all probability functions that are concentrated on $\{1, \cdots, M\}$. We call $\hat{f}_M$ an "M-restricted MLE", and we prove that whenever the data is such that a NPMLE does exist then for all large values of M, $\hat{f}_M$ is a NPMLE. We also show that when a NPMLE does not exist then the value of the likelihood function at $\hat{f}_M$ converges (as $M \to \infty$) to the supremum, over all lifetime probability functions, of the likelihood function, and the difference between the two quantities is $O(M^{-1})$. The optimal nature of the algorithm is proved in Section 3. In Section 4 we discuss some practical aspects concerning the use of the algorithm. In particular we point out that our assumptions about the sampling design are not restrictive and the algorithm yields the MLE for many other sampling schemes from renewal processes. For example, if we had K ordinary (as opposed to stationary) renewal processes, and each process is observed for a fixed period of time, starting at time zero, so that the likelihood of the data is of the type treated by Kaplan and Meier (1958), then our algorithm yields the Kaplan-Meier estimator. In Section 5 we describe applications from various fields of studies, such as absenteeism from work, reliability of components, medical studies, and more. Some of the examples we present were taken from works in which the need for our methodology arose but the problem was avoided.

2. The problem and the algorithm.

2.1 The data and the likelihood function. Consider a number, say K, of iid discrete time renewal processes, each of which began indefinitely far in the past so that the processes are stationary. For convenience we speak of the time period (counted in days) between successive "events" in each process as lifetimes of items and we assume that items

Received September 1981; revised January 1982.

[1] This research was partially done while the author was at the Technion, Israel Institute of Technology, Haifa, Israel.

AMS 1980 *subject classifications.* Primary 62G05; secondary 62E99, 62M99, 62P10.

Key words and phrases. Empirical distribution function, maximum likelihood, EM-algorithm, renewal theory, length biasing, survival analysis, censored data, Kaplan-Meier.

can fail only at midday, in which case they are replaced instantaneously. We denote the probability function of the items' lifetimes by $f = (f(1), f(2), \cdots)$, where $f(j) \geq 0$, $\sum_{j=1}^{\infty} f(j) = 1$, and we are interested in finding an MLE for f on the basis of the following data.

We assume that the ith process is observed for b_i consecutive days $a_i + 1, \cdots, a_i + b_i$, and no other information about the process is available. Here, a_i and b_i (≥ 1) are fixed integers, independent of the process itself. (Of course, because of the stationarity, we could assume $a_i = 0$.) We define

$N_i(j)$ = the total number of failures that have occurred during days $a_i + 1, \cdots, a_i + j$ in the ith process, $j = 1, 2, \cdots$.

$\xi_i(j)$ = the day on which the jth failure after day a_i occurred in the ith process, $j = 1, 2, \cdots$.

Since failures occur at midday, all lifetimes which are first (last) in their sampling periods are observed as censored on the left (right). Thus we can divide the observed lifetimes into the following four exclusive and exhaustive sets:

(i) $$\mathbf{X} = \{\xi_i(j) - \xi_i(j-1); j = 2, \cdots, N_i(b_i), N_i(b_i) \geq 2\},$$

the set of all lifetimes which originated and terminated within their sampling periods.

(ii) $$\mathbf{Y} = \{\xi_i(1) - a_i; N_i(b_i) \geq 1\},$$

the set of all (incomplete) lifetimes which are first, but not last, in their sampling periods. These lifetimes are censored on the left.

(iii) $$\mathbf{Z} = \{a_i + b_i + 1 - \xi_i(N_i(b_i)); N_i(b_i) \geq 1\},$$

the set of all (incomplete) lifetimes which are last, but not first, in their sampling periods. These lifetimes are censored on the right.

(iv) $$\mathbf{W} = \{b_i + 1; N_i(b_i) = 0\},$$

the set of all (incomplete) lifetimes which are both first and last in their sampling periods. These lifetimes are censored on both ends, and the only thing known about them is that their lengths exceed their sampling periods by at least 1.

Let $t_1 < t_2 < \cdots < t_h$ be the values taken by the observations in $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z} \cup \mathbf{W}$, in an increasing order, and let x_i, y_i, z_i, w_i be the multiplicities of observations from **X, Y, Z, W**, respectively, at t_i. The total numbers of observations in **X, Y, Z, W** are denoted by n_x, n_y, n_z, n_w, respectively. Note that

(2.1) $$n_x = \sum_{i=1}^{h} x_i, \qquad n_y = \sum_{i=1}^{h} y_i, \qquad n_z = \sum_{i=1}^{h} z_i, \qquad n_w = \sum_{i=1}^{h} w_i,$$

and although each of the n's could be zero, we must have $n_y + n_w = K$. In what follows we use the notation (for a general probability function f):

(2.2) $$S_f(j) = \sum_{k=j}^{\infty} f(k), \qquad \mu_f(j) = \sum_{k=j}^{\infty} kf(k), \qquad \mu_f = \mu_f(1),$$

and, as usual, the subscript f in various probability computations always indicates that the computation is done under the assumption that f is the true probability law.

To write the probability of the data, under the assumption that the true lifetime probability law is f, we recall (e.g., Feller, 1968, Chapter XIII) that observations in **Y** are realizations from the residual lifetime probability function $\mu_f^{-1}(S_f(1), S_f(2), \cdots)$, and observations in **W** are, therefore, censored observations from this distribution. Thus, the probability of the data is

(2.3) $$\begin{aligned} P_f(\text{data}) &= P_f(\{N_i(j); 1 \leq j \leq b_i\}, i = 1, \cdots, K) \\ &= \prod_{i=1}^{h} \{f(t_i)\}^{x_i} \{\mu_f^{-1} S_f(t_i)\}^{y_i} \{S_f(t_i)\}^{z_i} \{\mu_f^{-1} \textstyle\sum_{j=t_i}^{\infty} S_f(j)\}^{w_i}. \end{aligned}$$

There are three different data configurations that one should consider:

Configuration I: $n_w > 0$, $n_x + n_y + n_z = 0$. In this, rather trivial, case no failures have been observed and (2.3) becomes,

$$P_f(\text{data}) = \prod_{i=1}^{h} \left\{ \frac{\sum_{j=t_i}^{\infty} S_f(j)}{\sum_{j=1}^{t_i-1} S_f(j) + \sum_{j=t_i}^{\infty} S_f(j)} \right\}^{w_i}. \tag{2.4}$$

Since, however, for $x \geq 0$ and $a > 0$, fixed, $x/(a + x)$ is maximum at $x = \infty$, we see that (2.4) is maximized for any probability function f for which $\mu_f = \infty$; this is equivalent to $\sum_{j=M}^{\infty} S_f(j) = \infty$ for all M's. In the rest of this paper we assume that the data is never of this type; that is, we always observe at least one failure.

Configuration II: $n_w = 0$, $n_x + n_y + n_z > 0$. In this case at least one failure has been observed in each sampling period, and (2.3) becomes

$$P_f(\text{data}) = \mu_f^{-n_y} \prod_{i=1}^{h} \{f(t_i)\}^{x_i} \{S_f(t_i)\}^{y_i+z_i}. \tag{2.5}$$

Clearly, if f is such that $S_f(t_h + 1) > 0$, then by adding the mass $S_f(t_h + 1)$ to the mass at t_h we decrease μ_f, and either do not change or increase (according as $x_h = 0$ or $x_h > 0$, respectively) the product in (2.5), and so (2.5) increases. This shows that for the purpose of maximizing (2.5) we can restrict consideration to f's in Q_{t_h}, the set of all probability functions which are concentrated on $\{1, \cdots, t_h\}$.

In general we denote

$$Q_M = \{f(1), \cdots, f(M); \textstyle\sum_{j=1}^{M} f(j) = 1, f(j) \geq 0, j = 1, \cdots, M\}, \qquad Q = Q_\infty. \tag{2.6}$$

Configuration III: $n_w > 0$, $n_x + n_y + n_z > 0$. In this case failures have been observed in some of the sampling periods, but not in all of them. Using

$$\sum_{j=t_i}^{\infty} S_f(j) = \sum_{j=t_i}^{\infty} (j - t_i + 1) f(j) \tag{2.7}$$

we rewrite (2.3) as

$$\begin{aligned} P_f(\text{data}) = {} & \{\textstyle\sum_{j=1}^{t_h} j f(j) + u_f(t_h + 1)\}^{-(n_y+n_w)} \\ & \times \textstyle\prod_{i=1}^{h} (\{f(t_i)\}^{x_i} \{\sum_{j=t_i}^{t_h} f(j) + S_f(t_h + 1)\}^{y_i+z_i} \\ & \times \{\textstyle\sum_{j=t_i}^{t_h} j f(j) + \mu_f(t_h + 1) - (t_i - 1)[\sum_{j=t_i}^{t_h} f(j) + S_f(t_h + 1)]\}^{w_i}). \end{aligned} \tag{2.8}$$

An interesting feature of (2.8) is that there need not exist an $f \in Q$ for which (2.8) is maximum. Loosely speaking, the reason is that there are data sets for which it would be advantageous, in order to increase (2.8), to choose $S_f(t_h + 1) = 0$ even in cases where $\mu_f(t_h + 1) > 0$ at the optimum. Though there does not exist an $f \in Q$ for which this is possible, one can consider such an infeasible f as a limit of f's in Q. The following will clarify this point.

Three Examples. The examples which follow show that when the data is of Configuration III, any of the following situations is possible: (i) there does not exist an MLE (Example 1); (ii) there exists a unique MLE (Example 2); (iii) there exist infinitely many MLE's (Example 3).

EXAMPLE 1. (Based on 3 windows). $\mathbf{Y} = \{2\}$, $\mathbf{X} = \{2\}$, $\mathbf{Z} = \{1\}$, $\mathbf{W} = \{2, 2\}$, or alternatively, $(t_1, t_2) = (1, 2)$, $(y_1, y_2) = (0, 1)$, $(x_1, x_2) = (0, 1)$, $(z_1, z_2) = (1, 0)$, $(w_1, w_2) = (0, 2)$. Substituting the data in (2.8) we get, after some simple algebra,

$$P_f(\text{data}) = \frac{\{1 - f(1)\} f(2) (\mu_f - 1)^2}{\mu_f^3}. \tag{2.9}$$

In trying to maximize (2.9) we would like to choose $f(1) = 0$, $f(2) = 1$, and $\mu_f = 3$ (the maximizer of $\mu^{-3}(\mu - 1)^2$ over $\mu \geq 1$). Since this is impossible, we have for each $f \in Q$

$$P_f(\text{data}) < \max_{1 \leq \mu} \frac{(\mu - 1)^2}{\mu^3} = \frac{4}{27} = \mathscr{L}^*.$$

Here, and in the sequel, we denote

$$\mathscr{L}^* = \sup_{f \in Q} P_f(\text{data}). \tag{2.10}$$

Nevertheless, we can get arbitrarily close to $\mathscr{L}^*$ by a sequence of probability functions f_M from Q. Define for $M > 2$

$$f_M(j) = \begin{cases} 1 - \dfrac{1}{M-2} & j = 2 \\ \dfrac{1}{M-2} & j = M \\ 0 & \text{otherwise.} \end{cases} \tag{2.11}$$

Then $\mu_{f_M} = 3$ for all $M > 2$, and

$$P_{f_M}(\text{data}) = \frac{M-3}{M-2}\mathscr{L}^*. \tag{2.12}$$

We note that the infeasible law, which satisfies $f(2) = 1$ and $\mu_f = 3$, can be considered as a limit of the sequence f_M.

EXAMPLE 2. $\mathbf{Y} = \{2\}$, $\mathbf{X} = \{2\}$, $\mathbf{Z} = \{1\}$, $\mathbf{W} = \{2\}$. (Based on two of the three windows of the previous example.) For this data we have

$$P_f(\text{data}) = \frac{\{1 - f(1)\} f(2)(\mu_f - 1)}{\mu_f^2}, \tag{2.13}$$

and the unique probability function that maximizes this quantity is $f(2) = 1$ (and, of course, $\mu_f = 2$).

EXAMPLE 3. $\mathbf{Y} = \{2\}$, $\mathbf{X} = \{1\}$, $\mathbf{Z} = \{1\}$, $\mathbf{W} = \{2, 2\}$. (Similar to Example 1, except that now $(x_1, x_2) = (1, 0)$.) Here

$$P_f(\text{data}) = \frac{\{1 - f(1)\} f(1)(\mu_f - 1)^2}{\mu_f^3}, \tag{2.14}$$

and this quantity is maximized for $f(1) = \frac{1}{2}$ and $\mu_f = 3$. For example, any of the probability functions

$$f_M(j) = \begin{cases} \dfrac{1}{2} & j = 1 \\ \dfrac{M-5}{2(M-2)} & j = 2 \\ \dfrac{3}{2(M-2)} & j = M \\ 0 & \text{otherwise} \end{cases} \tag{2.15}$$

for $M = 5, 6, \cdots$ will maximize (2.14).

2.2 *The M-restricted* MLE. We summarize the situation which is exhibited by the above examples:

LEMMA 2.1. *There exist data sets, with $n_w > 0$, for which P_f (data) $< \mathscr{L}^*$ for all f's in Q.*

Let $\hat{f}_M$ be a solution (one always exists) to the problem of maximizing P_f(data) over all f's in Q_M, and denote

$$\mathscr{L}_M^* = P_{\hat{f}_M}(\text{data}) = \max_{f \in Q_M} P_f(\text{data}).$$

That is, $\hat{f}_M$ is an MLE of the lifetime probability function restricted to the set Q_M("M-restricted MLE," hereafter), and $\mathscr{L}_M^*$ is the corresponding value of the likelihood function.

LEMMA 2.2. (i) *If there exists an $f \in Q$ such that P_f (data) $= \mathscr{L}^*$ then, for all sufficiently large M, $\mathscr{L}_M^* = \mathscr{L}^*$.* (ii) *If P_f (data) $< \mathscr{L}^*$ for every $f \in Q$, then the sequence $\mathscr{L}_M^*$ converges monotonically to $\mathscr{L}^*$. Furthermore,*

$$\frac{\mathscr{L}_M^*}{\mathscr{L}^*} = 1 - O(M^{-1}) \quad \text{as} \quad M \to \infty. \tag{2.16}$$

(Cf. Example 1.) The proof is similar to the proof of Lemma 2.4 and hence will be omitted.

From a practical standpoint Examples 1 through 3, Lemmas 2.1 and 2.2, and (2.16), suggest that we should replace the original problem of finding an unrestricted MLE with the problem of finding an M-restricted MLE (for a fixed large M). We now reduce this last problem to a maximization problem with $h + 1$ variables, instead of M ($\gg h + 1$) variables.

LEMMA 2.3. *An M-restricted* MLE *is a solution to the problem: Maximize*

$$\mathscr{L}_M(p|\text{data}) = \left(\sum_{i=1}^{h+1} t_i p_i\right)^{-(n_y+n_w)} \prod_{i=1}^{h} \left[p_i^{x_i} \left(\sum_{j=i}^{h+1} p_j\right)^{y_i+z_i} \left\{\sum_{j=i}^{h+1} (t_j - t_i + 1) p_j\right\}^{w_i}\right], \tag{2.17}$$

subject to

$$\sum_{i=1}^{h+1} p_i = 1, \qquad p_i \geq 0, \qquad i = 1, \cdots, h+1. \tag{2.18}$$

Here $p_i \equiv p(t_i)$, $i = 1, \cdots, h + 1$, and the subscript M indicates that

$$t_{h+1} \equiv M. \tag{2.19}$$

Furthermore, if $f \in Q_M$ and $f(j) > 0$ for $j \notin \{t_1, \cdots, t_h\}$, $j < t_h$, then f cannot be an M-restricted MLE.

The proof is similar to the proof of Lemma 2.4 and hence will be omitted.

LEMMA 2.4. *Let $f \in Q$. Then for all sufficiently large M each Q_M contains a probability function p_M which satisfies*

$$\sum_{i=1}^{h+1} p_M(t_i) = 1 \ (t_{h+1} \equiv M) \quad \text{and} \quad P_f(\text{data}) \leq P_{p_M}(\text{data}).$$

Furthermore, if $f(k) > 0$ for some $k \notin \{t_1, \cdots, t_h\}$, $k < t_h$, then the inequality above is strict.

The proof of the Lemma is given in the Appendix.

2.3 *The* RT *Algorithm.* We describe now an algorithm for solving (2.17). The optimum properties of the algorithm are given in Theorem 1 below.

Step A. Start with an initial estimate $\{p_k^{\text{old}}\}$ satisfying

$$p_k^{\text{old}} > 0, \ k = 1, \cdots, h+1, \ \sum_{k=1}^{h+1} p_k^{\text{old}} = 1.$$

Step B. Evaluate, for $k = 1, \cdots, h + 1$,

$$r_k = x_k + p_k^{\text{old}} \sum_{i=1}^{k} \left\{ \frac{y_i + z_i}{\sum_{j=i}^{h+1} p_j^{\text{old}}} + \frac{(t_k - t_i + 1)w_i}{\sum_{j=i}^{h+1} (t_j - t_i + 1)p_j^{\text{old}}} \right\} \tag{2.20}$$

and solve the equation

$$\sum_{k=1}^{h+1} \frac{r_k t_k}{(n_x + n_z)\mu + (n_y + n_w)t_k} = 1 \tag{2.21}$$

for μ. This can be done by successively bisecting the interval $[t_1, t_{h+1}]$. Denote the solution by μ^{new}.

Step C. Set

$$p_k^{\text{new}} = \frac{r_k \mu^{\text{new}}}{(n_x + n_z)\mu^{\text{new}} + (n_y + n_w)t_k}, \qquad k = 1, \cdots, h+1. \tag{2.22}$$

Step D. If p^{new} is sufficiently close to p^{old}, so that the required accuracy has been achieved, then stop. Otherwise, return to Step B with p^{new} replacing p^{old}.

THEOREM 1. *The* RT *algorithm converges monotonically to a fixed point* p^* *which satisfies the Kuhn-Tucker conditions*

$$p_k \frac{\partial \ell(p)}{\partial p_k}\bigg|_{p^*} = 0, \qquad k = 1, \cdots, h+1 \tag{2.23}$$

$$\frac{\partial \ell(p)}{\partial p_k}\bigg|_{p^*} \le 0 \quad \text{if} \quad p_k^* = 0, \qquad k = 1, \cdots, h+1, \tag{2.24}$$

where

$$\ell(p) = \log \mathcal{L}_M(p \mid \text{data}) - (n_x + n_z)\left(\sum_{i=1}^{h+1} p_i - 1\right), \tag{2.25}$$

is the Lagrangian of $\log \mathcal{L}_M$ ($\mathcal{L}_M$ *is given in* (2.17)), *and* $(n_x + n_z)$ *is the Lagrange multiplier.*

REMARK. Though it is generally considered satisfactory, in the theory of nonlinear programming, to prove that an algorithm converges to a point which satisfies the Kuhn-Tucker conditions, we have to remember that these conditions are sufficient for optimality only if the problem can be suitably transformed into a convex programming problem. We show this for the special case $n_w = 0$ (i.e., Configuration II) and $x_i > 0$ $i = 1, \cdots, h$. In this case maximizing (2.17) subject to (2.18) is the same as

$$\text{maximize } \left(\prod_{i=1}^{h} p_i^{x_i} q_i^{y_i + z_i}\right) u^{-n_y} \tag{2.26}$$

subject to

$$\sum_{j=1}^{h} p_j = 1, \quad \sum_{j=i}^{h} p_j = q_i, \quad i = 1, \cdots, h, \quad \sum_{j=1}^{h} t_j p_j = u, \quad p_j > 0, \quad j = 1, \cdots, h. \tag{2.27}$$

Now, the maximum of (2.26) restricted to (2.27) is attained at the same points as the maximum of (2.26) restricted to

$$\begin{gathered} \sum_{j=1}^{h} p_j \le 1, \qquad q_1 \le \sum_{j=1}^{h} p_j, \qquad p_i + q_{i+1} \le q_i, \qquad i = 1, \cdots, h, \qquad (q_{h+1} \equiv 0), \\ \sum_{j=1}^{h} t_j p_j \le u, \qquad p_j > 0, \qquad j = 1, \cdots, h, \end{gathered} \tag{2.28}$$

and the maximum of (2.26) restricted (2.28) is attained at the same points as the maximum of (2.26) restricted to

$$\begin{gathered} \sum_{j=1}^{h} p_j \le 1, \qquad q_1 \le 1, \qquad p_i + q_{i+1} \le q_i, \qquad i = 1, \cdots, h, \qquad (q_{h+1} \equiv 0), \\ \sum_{j=1}^{h} t_j p_j \le u, \qquad p_j > 0, \qquad j = 1, \cdots, h. \end{gathered} \tag{2.29}$$

Substituting $p_j = e^{-\alpha_j}$, $q_j = e^{-\beta_j}$, $j = 1, \cdots, h$ and $u = e^{\gamma}$ in (2.26) and (2.29) we get a problem of minimizing a convex function over a convex region which has a unique solution. This unique solution is the unique point which satisfies the Kuhn-Tucker conditions for (2.26) and (2.29). One can then verify that a Kuhn-Tucker point for the latter problem is mapped into a Kuhn-Tucker point of the original problem (2.17, 2.18) and vice versa.

3. Proof of Theorem 1. To prove Theorem 1, and to understand the rationale behind the algorithm, we need to review a few facts.

In this section U always stands for a random variable (rv) with a probability law given by

$$P(U = t_i) = p(t_i) = p_i \geq 0, \qquad i = 1, \cdots, h+1, \tag{3.1}$$

$\sum_{i=1}^{h+1} p_i = 1$, and V always stands for a rv, independent of U, with the length biased probability law:

$$P(V = t_i) = \frac{t_i p_i}{\sum_{j=1}^{h+1} t_j p_j}, \qquad i = 1, \cdots, h+1. \tag{3.2}$$

Suppose now that $u_1, \cdots, u_m$ is a random sample from (3.1) and $v_1, \cdots, v_n$ is a random sample from (3.2), and let ξ_i and η_i denote the multiplicities of the u's and the v's at t_i, respectively. Then we have the following.

FACT 1 (Vardi, 1982). The unique vector $\hat{p} = (\hat{p}_1, \cdots, \hat{p}_{h+1})$ that maximizes the likelihood function

$$\mathcal{L}_1(p \mid u_1, \cdots, u_m, v_1, \cdots, v_n) = \prod_{i=1}^{h+1} p_i^{\xi_i} \left[\frac{t_i p_i}{\sum_{j=1}^{h+1} t_j p_j}\right]^{\eta_i} \tag{3.3}$$

is given by

$$\hat{p}_k = \frac{(\xi_k + \eta_k)\hat{\mu}}{m\hat{\mu} + nt_k}, \qquad k = 1, \cdots, h+1 \tag{3.4}$$

where $\hat{\mu}$ is the unique solution of

$$\sum_{k=1}^{h+1} \frac{(\xi_k + \eta_k)t_k}{m\mu + nt_k} = 1. \tag{3.5}$$

Furthermore, $\hat{\mu}$ of (3.5) satisfies $t_1 \leq \hat{\mu} = \sum_{i=1}^{h+1} t_i \hat{p}_i \leq t_{h+1}$, and since the left hand side of (3.5) is monotone in μ, we can use the interval bisection method to approximate $\hat{\mu}$, numerically, with accuracy of at least $(t_{h+1} - t_1)2^{-N}$ in N evaluations (and comparisons) of the left hand side of (3.5).

Let R be a uniform (0, 1) rv, independent of V, and let $[x]$ denote the smallest integer that is *larger* than or equal to x. Then we have

FACT 2.

$$P([RV] = j \mid V = t_k) = \frac{1}{t_k}, \qquad j = 1, \cdots, t_k, \qquad k = 1, \cdots, h+1. \tag{3.6}$$

FACT 3.

$$P([RV] = j) = \frac{S_p(j)}{\mu_p}, \qquad j = 1, \cdots. \tag{3.7}$$

Note that $S_p(j) = \sum_{i \geq j} p(i) = \sum_{t_i \geq j} p_i$, so that

$$P([RV] = t_j) = \frac{\sum_{i=j}^{h+1} p_i}{\sum_{i=1}^{h+1} t_i p_i}, \qquad j = 1, \cdots, h+1. \tag{3.8}$$

FACT 4.

$$P(V = t_k \mid [RV] = j) = \frac{p(t_k)}{S_p(j)} I[j \le t_k] = \frac{p_k}{\sum_{t_i \ge j} p_i} I[j \le t_k],$$

so that

$$P(V = t_k \mid [RV] = t_j) = \frac{p_k}{\sum_{i=j}^{h+1} p_i} I[j \le k]. \tag{3.9}$$

Here and in the sequel, $I[\ ]$ denotes the indicator function.

FACT 5.

$$P(V = t_k \mid [RV] \ge j) = \frac{(t_k - j + 1)p_k}{\sum_{t_i \ge j} (t_i - j + 1)p_i} I[j \le t_k]$$

so that

$$P(V = t_k \mid [RV] \ge t_j) = \frac{(t_k - t_j + 1)p_k}{\sum_{i=j}^{h+1} (t_i - t_j + 1)p_i} I[j \le k]. \tag{3.10}$$

Using these five facts, we shall tailor an artificial problem to our data, and show that the likelihood function of the artificial problem coincides with the likelihood function of our original problem. We then show that our algorithm coincides with the EM algorithm (Dempster et al., 1977) for the artificial problem.

The artificial problem. Consider a situation where we have two independent samples: $u = (u_1, \cdots, u_{n_x+n_z})$ from (3.1) and $v = (v_1, \cdots, v_{n_y+n_w})$ from (3.2). Nevertheless, we do not observe the *complete data* (u, v), but rather the *incomplete data* A_x, A_y, A_z, A_w given by

$$\begin{aligned} A_x &= \{u_1, \cdots, u_{n_x}\}, \qquad A_z = \{u_{n_x+1} \wedge \bar{z}_1, \cdots, u_{n_x+n_z} \wedge \bar{z}_{n_z}\}, \\ A_y &= \{[R_1 v_1], \cdots, [R_{n_y} v_{n_y}]\}, \\ A_w &= \{[R_{n_y+1} v_{n_y+1}] \wedge \bar{w}_1, \cdots, [R_{n_y+n_w} v_{n_y+n_w}] \wedge \bar{w}_{n_w}\}, \end{aligned} \tag{3.11}$$

where $a \wedge b = \min(a, b)$. Here $R_1, \cdots, R_{n_y+n_w}$ are iid uniform (0, 1) rv's $\{\bar{z}_1, \cdots, \bar{z}_{n_z}\} = \mathbf{Z}$ and $\{\bar{w}_1, \cdots, \bar{w}_{n_w}\} = \mathbf{W}$. Suppose further that it so happened that

$$\begin{aligned} &u_i = \bar{x}_i, \quad i = 1, \cdots, n_x, \qquad u_i \ge \bar{z}_{i-n_x}, \quad i = n_x + 1, \cdots, n_x + n_z, \\ &[R_i v_i] = \bar{y}_i, \quad i = 1, \cdots, n_y, \qquad [R_i v_i] \ge \bar{w}_{i-n_y}, \quad i = n_y + 1, \cdots, n_y + n_w, \end{aligned} \tag{3.12}$$

where $\{\bar{x}_1, \cdots, \bar{x}_{n_x}\} = \mathbf{X}$, and $\{\bar{y}_1, \cdots, \bar{y}_{n_y}\} = \mathbf{Y}$.

The problem is to find the MLE of the p_i's in (3.1) on the basis of the incomplete data A_x, A_y, A_z, A_w. Because of (3.11) and (3.12) we have

$$A_x = \mathbf{X}, \qquad A_y = \mathbf{Y}, \qquad A_z = \mathbf{Z}, \qquad A_w = \mathbf{W}, \tag{3.13}$$

and so, using (3.8), we get the following:

LEMMA 3.1. *The likelihood function of the artificial problem,* $\mathcal{L}_A(p \mid A_x, A_y, A_z, A_w)$, *coincides with* $\mathcal{L}_M(p \mid \text{data})$ *of* (2.17).

Being a standard incomplete data problem, it is natural to try the EM algorithm on the artificial problem. The following is a single iteration of the EM algorithm applied to the artificial problem:

Let $p^{\text{old}} = (p_1^{\text{old}}, \cdots, p_{h+1}^{\text{old}})$ be the current estimate of p. Then we first compute (E-step)

$$E\left\{\log \prod_{i=1}^{h+1} p_i^{\xi_i}\left(\frac{t_i p_i}{\sum_{j=1}^{h+1} t_j p_j}\right)^{\eta_i} \middle| \mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}, p^{\text{old}}\right\}$$

$$= \sum_{i=1}^{h+1} \tilde{\xi}_i \log p_i + \sum_{i=1}^{h+1} \tilde{\eta}_i \log \frac{t_i p_i}{\sum_{j=1}^{h+1} t_j p_j}, \tag{3.14}$$

where ξ_i and η_i are, as before, the multiplicities of the u's and the v's at t_i, respectively, and $\tilde{\xi}_i$ and $\tilde{\eta}_i$ are their conditional expectations given the data and p^{old}. These conditional expectations are derived using (3.9) and (3.10):

$$\tilde{\xi}_k = x_k + \sum_{i=1}^{k} z_i \frac{p_k^{\text{old}}}{\sum_{j=i}^{h+1} p_j^{\text{old}}},$$

$$\tilde{\eta}_k = \sum_{i=1}^{k} y_i \frac{p_k^{\text{old}}}{\sum_{j=i}^{h+1} p_j^{\text{old}}} + \sum_{i=1}^{k} w_i \frac{(t_k - t_i + 1)p_k^{\text{old}}}{\sum_{j=i}^{h+1} (t_j - t_i + 1)p_j^{\text{old}}}, \tag{3.15}$$

$k = 1, \cdots, h+1$. Completing this step, we proceed to find (M-step) the p that maximizes (3.14), subject to $\sum_{i=1}^{h+1} p_i = 1$ and $p_i \geq 0$, $i = 1, \cdots, h+1$. This is done in two steps, using Fact 1. First we solve

$$\sum_{i=1}^{h+1} \frac{(\tilde{\xi}_i + \tilde{\eta}_i)t_i}{(n_x + n_z)\mu + (n_y + n_w)t_i} = 1 \tag{3.16}$$

for μ (note that $\sum_{i=1}^{h+1} \tilde{\xi}_i = n_x + n_z$ and $\sum_{i=1}^{h+1} \tilde{\eta}_i = n_y + n_w$) and then, after denoting the solution by μ^{new}, we set

$$p_k^{\text{new}} = \frac{(\tilde{\xi}_k + \tilde{\eta}_k)\mu^{\text{new}}}{(n_x + n_z)\mu^{\text{new}} + (n_y + n_w)t_k}, \qquad k = 1, \cdots, h+1. \tag{3.17}$$

Since r_k of (2.20) coincides with $\tilde{\xi}_k + \tilde{\eta}_k$ of (3.15), we see that the RT algorithm coincides with the EM algorithm based on the artificial problem.

It is important to note, however, that the recipe given in Dempster et al. (1977) for the EM algorithm, in general, does not induce any particular algorithm in our original problem, because the probability specification of the original problem does not fall in the category of what they call a "standard incomplete data problem."

The advantage of deriving the RT algorithm as an EM algorithm (and, indeed, the only place where we use this fact, mathematically) is the following.

LEMMA 3.2. (Monotonicity of the algorithm.) *If in the* RT *algorithm* $p^{\text{new}} \neq p^{\text{old}}$, *then* $\mathcal{L}_M(p^{\text{new}}|\text{data}) > \mathcal{L}_M(p^{\text{old}}|\text{data})$.

PROOF. Because of Fact 1, if $p^{\text{new}} \neq p^{\text{old}}$, then (3.14) with $p_k = p_k^{\text{new}}$ is strictly bigger than (3.14) with $p_k = p_k^{\text{old}}$. The proof follows now from Theorem 1 of Dempster et al. (1977).

PROOF OF THEOREM 1. It is easily verified that if p^* is a fixed point of the algorithm, then it satisfies (2.23). Because of Lemma 3.2, the continuity of the mapping from p^{old} to p^{new} in the algorithm, and Convergence Theorem A of Zangwill (1969, page 91), we conclude that indeed the algorithm *does* converge to a fixed point, and so the fixed point satisfies (2.23). Suppose then, by negation, that the point of convergence, p^*, does not satisfy (2.24). That is, there is a k such that $p_k^* = 0$ but

$$\left.\frac{\partial \ell(p)}{\partial p_k}\right|_{p^*} > 0 \tag{3.18}$$

and let $\{p^{(n)}\}$ denote the sequence of p's (produced by the algorithm) that converges to p^*.

Since $p_k^* = 0$, we must have $x_k = 0$, because otherwise (2.22) combined with $r_k^{(n)} \geq x$ would imply $p_k^* > 0$. Therefore by combining (2.20) with (2.22) we can write

$$r_k^{(n+1)} = r_k^{(n)} \frac{\sum_{i=1}^{k} \left\{ \frac{y_i + z_i}{\sum_{j=i}^{h+1} p_j^{(n+1)}} + \frac{(t_k - t_i + 1)w_i}{\sum_{j=i}^{h+1} (t_j - t_i + 1)p_j^{(n+1)}} \right\}}{(n_x + n_z) + (n_y + n_w)t_k/\mu^{(n+1)}} \equiv r_k^{(n)} q_k^{(n+1)}. \tag{3.19}$$

Now, because of (3.18) and the continuity of $\partial \ell(p)/\partial p_k$, there exists $\delta > 0$ such that for all $n \geq n_0$

$$\left. \frac{\partial \ell(p)}{\partial p_k} \right|_{p^{(n)}} > 1 + \delta. \tag{3.20}$$

Since, however,

$$(q_k^{(n)} - 1)\left\{ (n_x + n_z) + (n_y + n_w) \frac{t_k}{\mu^{(n)}} \right\} = \left. \frac{\partial \ell(p)}{\partial p_k} \right|_{p^{(n)}} \tag{3.21}$$

it follows from (3.20) and (3.21) that there exists an $\varepsilon > 0$ such that for all $n \geq n_0$

$$q_k^{(n)} > 1 + \varepsilon. \tag{3.22}$$

Substituting in (3.19), we see that for all $n \geq n_0$

$$r_k^{(n)} > r_k^{(n_0)}(1 + \varepsilon)^{n - n_0} \to \infty, \qquad \text{as} \quad n \to \infty, \tag{3.23}$$

and this is a contradiction. Note that since we start the algorithm with $p_j^{(0)} > 0$, $j = 1, \cdots, h + 1$, we have (i) $r_j^{(n)} > 0$, $j = 1, \cdots, h + 1$ for all n's, and (ii) $\sum_{j=i}^{h+1} r_j^{(n)} = n_x + n_y + n_z + n_w < \infty$, for all n's.

This completes the proof of the Theorem.

4. Some practical considerations.

4.1 *Applicability of the algorithm to other sampling designs.* So far we have assumed that all the processes are stationary and in each process the sampling period is a window of fixed length whose location is independent of the process itself. Though this sampling design seems quite restrictive, the truth is that it gives rise to a likelihood function which is general enough to include the likelihood function of most other practical designs.

For example, suppose that the sampling design is such that the sampling windows always start at a failure time. E.g., we put k items on trial, we replace each item upon failure, and we stop the experiment after a fixed period of time. In this case $y_i = w_i = 0$, $i = 1, \cdots, h$, and the likelihood function that should be maximized is

$$\mathscr{L}(p|\text{data}) = \prod_{i=1}^{h} p_i^{x_i} (\textstyle\sum_{j=i}^{h} p_j)^{z_i}. \tag{4.1}$$

This is a Kaplan-Meier (1958) likelihood function, and indeed if all the y_i's and w_i's are zero, our algorithm becomes KM(**X**; **Z**), the Kaplan-Meier estimator based on **X** as the complete data and **Z** as the censored data. See Efron (1967) for the form of the KM estimator which coincides with the RT algorithm.

As another example, suppose all the windows end at failure times. E.g., the sampling design is to sample each process until we observe a certain number of failures. In this case $z_i = w_i = 0$, $i = 1, \cdots, h$, and our algorithm finds the p that maximizes

$$\mathscr{L}(p|\text{data}) = (\textstyle\sum_{i=1}^{h} t_i p_i)^{-n_y} \prod_{i=1}^{h} p_i^{x_i} (\textstyle\sum_{j=i}^{h} p_j)^{y_i}. \tag{4.2}$$

In the special case where each process is sampled until we observe a single failure, we are in the problem of estimating the lifetime distribution on the basis of a sample from the residual-lifetime distribution. Two estimation methods, neither of which is a maximum likelihood type method, for this problem are proposed in Cox (1969, Section 5.3).

It is of course possible to pool data from different experiments and create the joint likelihood. A product of (4.1) and (4.2) would be such an example. The RT algorithm can then be used to find the overall nonparametric MLE on the basis of the pooled data.

4.2 *A note on the estimability of the mean lifetime.* An interesting property of our sampling design, which is very much in contrast with other sampling designs that give rise to incomplete data problems, such as Kaplan and Meier's (1958), and Turnbull's (1974, 1976), is that even when all the windows' lengths are bounded (say by b) so that it is impossible to observe a lifetime which is longer than b, we can still estimate the mean lifetime consistently! (The asymptotic properties of the RT estimator will be included in a future paper. Note, however, that $\sum_{i=1}^{K} b_i / \sum_{i=1}^{K} N_i(b_i)$ is a consistent estimate of the mean lifetime.)

Consider, for example, the following design problem. We have K iid *ordinary* renewal processes, each of which started at time 0, and suppose that from each process we are allowed to sample only a single window of (predetermined) length b. We are free to determine the location of the sampling windows. Where should we locate them? Without being very formal about the purpose of the experiment that, with the introduction of a loss function, can be formulated as a decision theory problem, we can assume that in most applications we shall be interested in estimating the underlying distribution function and also the mean lifetime. Clearly, if we locate the windows at the origin (i.e., each process is observed during $[0, b]$, so that we get a Kaplan-Meier liklihood function) then the mean lifetime is not estimable. The same situation holds for most other choices of locations, if they depend on the sample paths of the processes themselves. If, however, we decide to sample at $[a, a + b]$ for some arbitrary $a \gg 0$, then for all practical purposes we can assume that the processes are stationary from a on, and the RT algorithm can be used to estimate the distribution function and the mean lifetime.

4.3 *On the choice of* $M \equiv t_{h+1}$. If there are no w's ($w_i = 0, i = 1, \cdots, h$) then it follows from the discussion of Configuration II (Section 2.1) that we can (and should!) carry out the algorithm without the additional point t_{h+1}, or equivalently set $t_{h+1} = t_h$. (Note that if in addition $y_i = 0$, $i = 1, \cdots, h$, so that the likelihood function is (4.1), this gives us the version of the KM estimator that always assigns zero mass to the right of t_h.)

If at least one of the w_i's is positive then we should choose a very large M ($\gg t_h$), and interpret the fixed point of the RT algorithm as an M-restricted MLE. (Of course, there are data sets for which this will be an unrestricted MLE.) Note, however, that if p_M denotes the algorithm's fixed point then, since

$$\mu_M = \sum_{i=1}^{h+1} t_i p_M(t_i) < \text{const.} < \infty \quad \text{for all} \quad M\text{'s}, \qquad \text{we have} \quad p_M(t_{h+1}) = O(M^{-1}).$$

Therefore one should not attribute much importance to the particular value of $p_M(t_{h+1})$ but rather interpret $p_M(t_h) + p_M(t_{h+1})$ as an estimate of the tail probability and $t_h p_M(t_h) + t_{h+1} p_M(t_{h+1})$ as an estimate of the tail expectation.

4.4 *A numerical example.* The following example, based on $n_y + n_w = 3$ windows, has $h = 10$ observation values. The data are given in Table 1. Five different values of M are used in the RT algorithm, which took approximately 20 iterations to converge for all M. The required accuracy in Step D of the algorithm was 10^{-7}. Resulting estimates $p_M(t_i)$ are given in Table 1, together with estimates μ_M of the mean.

REMARK. One should not deduce from this example that the RT estimator always assigns zero mass to t_i's for which $x_i = 0$ and $i < h$. The reader can verify that for $(t_1, t_2) = (1, 2)$, $(x_1, x_2) = (0, 1)$, and $(y_1, y_2) = (z_1, z_2) = (3, 0)$, the NPMLE is $p(1) = p(2) = \frac{1}{2}$.

More complicated examples are considered in an ongoing study by L. Denby and the author.

TABLE 1
A numerical example of the RT *estimates for varying M.*

Data					Estimates $p_M(t)$				
t_i	x_i	y_i	z_i	w_i	$M = 10^2$	$M = 10^3$	$M = 10^4$	$M = 10^5$	$M = 10^6$
3	0	1	0	0	0	0	0	0	0
7	1	0	0	0	.1082	.1098	.1101	.1101	.1101
8	0	0	1	0	0	0	0	0	0
9	2	0	0	0	.2361	.2411	.2417	.2418	.2418
10	0	1	0	0	0	0	0	0	0
13	1	0	0	0	.1307	.1354	.1360	.1361	.1361
14	0	0	1	0	0	0	0	0	0
16	1	0	0	0	.1592	.1692	.1705	.1707	.1707
17	0	0	0	1	0	0	0	0	0
19	2	0	0	0	.3303	.3393	.3411	.3413	.3413
M	0	0	0	0	.0355	.0052	†ε_1	ε_2	ε_3
					$\mu_M = 16.954$	19.027	19.328	19.359	19.362

† $10^2\,\varepsilon_3 \simeq 10\,\varepsilon_2 \simeq \varepsilon_1 \simeq 5.4 \times 10^{-4}$

5. A few applications. (i). In Vardi (1981) the problem of estimating the distribution of periods between successive absences from work (show-up periods) for telephone operators was considered. Here "failures" are absences from work. Since attendance records are kept on a calendar basis, rather than "failure time" basis, the sampling periods usually started in the middle of a show-up period and continued for a fixed period of time. Our algorithm was then needed in order to estimate the distribution of the show-up periods.

(ii). In studying maintenance policies for vehicles, Brosh et al. (1975) sampled 100 vehicles over a fixed period of 27 months. The dates and types of components that failed (immediate replacement) were recorded and most vehicles had only one or two failures per component. The authors considered the data "insufficient to establish directly the life distribution of each component with a high degree of confidence." This is a situation where we have a relatively few x_i's and w_i's, and about 100 y_i's and z_i's. This is a sufficiently large number of observations to consider the RT estimator before one tries to fit a parametric family.

(iii). In studying the performance of a telephone network, Denby et al. (1975) sampled about one minute of each of many weekday hours of system operation. Here the window lengths are about one minute and the RT algorithm can be used to estimate the "lifetime" (time between consecutive errors in transmission) distribution.

(iv). Consider a medical study in which the interest is in estimating the distribution of the time periods between consecutive epilepsy attacks (the "lifetimes") for a homogeneous group of epileptics, and the data is obtained by asking each member of the group to record the dates of attacks (if any) during a specified period. The RT estimator could then be used to estimate the lifetime distribution.

In a fashion similar to the last example, the need for the RT estimator often occurs when a study is initiated on the basis of existing records which are kept on a calendar basis, rather than a failure time basis.

The above examples give the impression that in most applications the windows are all of equal lengths. This, however, need not be the case. For instance, in (i) above the data come from different sources, and attendance records from different sources covered different periods. Also, in applications such as (ii) above, it is more appropriate to measure the components' "lifetimes" in miles, rather than in time units, in which case the windows would be of different lengths, because over a given period of time different vehicles usually accumulate different mileages.

Final remark. The applicability of the RT estimator to continuous time renewal processes, with nonarithmetic lifetime distribution, can be justified using a limiting argument of the type described in Kalbfleish and Prentice (1980, pages 12–13); see also Johansen (1978). A treatment of the case where the lifetime distribution belongs to an absolutely continuous, parametric, family, and asymptotic results which are relevant to the RT algorithm will be given in a future paper.

APPENDIX: PROOF OF LEMMA 2.4

Let $f \in Q$, $A = S_f(t_h + 1)$ and $B = \mu_f(t_h + 1)$. Since we assumed that at least one failure has been observed, the number of *first* failures is positive and so $n_y > 0$. Thus if $B = \infty$ we get from (2.8) that $P_f(\text{data}) = 0$ and for every $M > t_h$ we can choose $p_M(t_i) = (h + 1)^{-1}$, $i = 1, \cdots, h + 1$, to get the desired inequality. It follows then that we can assume $B < \infty$. Suppose now that $B > 0$, and hence $A > 0$. Then if M is any integer larger than B/A, the probability function g given by

$$g(j) = f(j), \qquad j = 1, \cdots, t_h, \qquad g(t_h + 1) = \frac{AM - B}{M - t_h - 1} \quad \text{and} \quad g(M) = \frac{B - A(t_h + 1)}{M - t_h - 1}$$

satisfies $P_g(\text{data}) = P_f(\text{data})$ and, of course, $g \in Q_M$. Suppose now that there exists a k such that $g(k) > 0$ and $t_i < k < t_{i+1}$, $0 \le i \le h - 1$ ($t_0 \equiv 1$). Then by moving $(t_h - t_i)^{-1}(t_h - k)g(k)$ of the mass at k to t_i, and the remainder of the mass to t_h, we obtain a new probability function p which satisfies $p(k) = 0$,

$$\text{(A.1)} \qquad \sum_{j=t_m}^{t_h} p(j) \ge \sum_{j=t_m}^{t_h} g(j), \quad \text{and} \quad \sum_{j=t_m}^{t_h} (j - t_m + 1)p(j) \ge \sum_{j=t_m}^{t_h} (j - t_m + 1)g(j),$$

with equalities for $m = 0, \cdots, i$ and strict inequalities for $m = i + 1, \cdots, h$. Substituting this in (2.8) we get $P_p(\text{data}) > P_g(\text{data}) = P_f(\text{data})$. Repeating the above argument, if necessary, for other k's, we conclude that there exists $p \in Q_M$ such that $P_p(\text{data}) > P_f(\text{data})$ and p satisfies $p(j) = 0$ for $j \notin \{1, t_1, \cdots, t_h, t_h + 1, M\}$. For such a p we write (2.8) as

$$\text{(A.2)} \qquad P_p(\text{data}) = \{\textstyle\sum_{i=0}^{h+2} t_i p(t_i)\}^{-(n_y+n_w)} \prod_{i=1}^{h} [\{p(t_i)\}^{x_i} \{\sum_{j=i}^{h+2} p(t_j)\}^{y_i+z_i} \{\sum_{j=i}^{h+2} (t_j - t_i + 1)p(t_j)\}^{w_i}]$$

where $t_0 = 1$, $t_{h+1} = t_h + 1$, $t_{h+2} = M$. Suppose now that $p(t_{h+1}) > 0$, then by redistributing the mass between t_h and $t_{h+2} = M$, so that $\sum_{j=0}^{2} t_{h+j} p(t_{h+j})$ remains unchanged, we shall increase (A.2) if $x_h > 0$ and leave it unchanged otherwise. Thus we can assume that $p(t_{h+1}) = 0$. Suppose now that $p(t_0) > 0$ and that $t_1 > t_0 = 1$. (Note that if $t_1 = t_0$ the proof is completed.) Then by applying a single iteration of the RT algorithm to the points t_0, t_1, $\cdots$, t_h, M with p being p^{old} we get a probability function p^{new} which satisfies $p^{\text{new}}(t_0) = 0$, and, because of the strict monotonicity of the algorithm (Lemma 3.2), $P_{p^{\text{new}}}(\text{data}) > P_p(\text{data})$. (Note that the proof of Lemma 3.2 is independent of Lemma 2.4.) Thus we can also assume $p(t_0) = 0$. This completes the proof for the case $B > 0$. The proof for the case $B = 0$ is similar and will be omitted.

Acknowledgments. This paper is dedicated to the memory of Professor Jack Kiefer.

REFERENCES

BROSH, I., SHLIFER, E., and ZEIRA, Y. (1975). Optimal maintenance policy for a fleet of vehicles. *Management Science* **22** 4, 401–410.

COX, D. R. (1969). Some sampling problems in technology. In: Johnson, N. L., and Smith, H. Jr. (Eds.) *New Developments in Survey Sampling*, 506–527. Wiley-Interscience, New York.

DEMPSTER, A. P., LAIRD, N. M., and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc.* **39** 1–37.

DENBY, L., GABBE, J. D., and MCRAE, J. E. (1975). T1-carrier, San Francisco Bay area, September 1973: Errors and Network. Bell Laboratories, unpublished technical memorandum.

EFRON, B. (1967). The two sample problem with censored data. *Proc. Fifth Berkeley Symposium in Math. Statist.* IV. 831–853. Prentice-Hall, New York.

FELLER, W. (1968). *Introduction to Probability Theory and its Applications.* **1** 3d ed. Wiley, New York.

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

KALBFLEISCH, J. D., and PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data.* Wiley, New York.

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

TURNBULL, B. W. (1976). The empirical distribution function with arbitrarily grouped censored and truncated data. *J. Roy. Statist. Soc. B.* **38** 290–295.

TURNBULL, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.* **69** 169–173.

VARDI, Y. (1981). Absenteeism of operators: A statistical study with managerial applications. *Bell System Tech. J.* **60** 1, 13–38.

VARDI, Y. (1982). Nonparametric estimation in the presence of length bias. *Ann. Statist.* **10** 616–620.

ZANGWILL, W. I. (1969). *Nonlinear Programming: A Unified Approach.* Prentice-Hall, New York.

BELL LABORATORIES
600 MOUNTAIN AVE.
MURRAY HILL, NEW JERSEY 07974