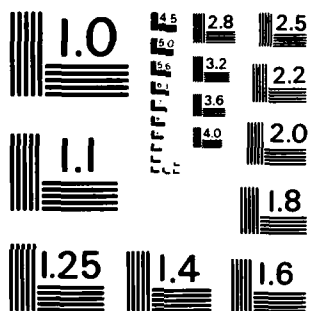


AD-A145 429 MAXIMUM LIKELIHOOD ESTIMATION OF THE SURVIVAL FUNCTIONS 1/1  
OF N STOCHASTICAL..(U) IOWA UNIV IOWA CITY DEPT OF  
STATISTICS AND ACTUARIAL SCIENCE.. C J FELTZ ET AL  
UNCLASSIFIED MAR 84 TR-104 N00014-83-K-0249 F/G 12/1 NL

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |

END  
DATE  
FILMED  
10 84  
DTIC



MICROCOPY RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963-A

12

MAXIMUM LIKELIHOOD ESTIMATION OF THE SURVIVAL FUNCTIONS  
OF N STOCHASTICALLY ORDERED RANDOM VARIABLES

by

Carol J. Feltz  
Western Electric  
Princeton, NJ

and

Richard L. Dykstra\*  
Department of Statistics and Actuarial Science  
University of Iowa, Iowa City, IA

AD-A145 429

DTIC FILE COPY

Technical Report #104  
March 1984

DTIC  
ELECTE  
SEP 12 1984  
B

**DISTRIBUTION STATEMENT A**  
Approved for public release;  
Distribution Unlimited

\*This research was partially supported by ONR Contract N00014-83-K-0249.

84 09 06 052

# ABSTRACT

Very often, populations exist that, logically, should satisfy linear stochastic ordering requirements. For example if a mechanical device is improved through N stages, the corresponding survival functions should be linearly stochastically ordered. Nevertheless, estimates may not reflect this stochastic ordering because of the inherent variability of the observations.

~~Here we~~ <sup>the</sup> characterizes the maximum likelihood estimates of the survival functions subject to linear stochastic ordering requirements. ~~We show our~~ <sup>These</sup> estimates may be expressed in terms of the well-known Kaplan-Meier product limit estimates. ~~We also give~~ <sup>is</sup> an iterative algorithm which we show must converge to the correct solution that depends only upon solving the pairwise problem.

Finally ~~we~~ consider an example concerning survival times for people with squamous carcinoma in the oropharynx when classified by degree of lymph node deterioration at time of discovery.

Key Words and Phrases: Maximum likelihood estimation; survival functions; linear stochastic ordering; censored data; Kaplan-Meier product limit estimator; order restrictions; Kuhn-Tucker vectors.

AMS Classification Numbers: Primary 60G05; Secondary 62N05.

|                       |              |                                            |
|-----------------------|--------------|--------------------------------------------|
| DTIC TAB              |              | <input checked="checked" type="checkbox"/> |
| Unannounced           |              | <input type="checkbox"/>                   |
| Justification         |              | <input type="checkbox"/>                   |
| By <b>PER CALL JC</b> |              |                                            |
| Distribution/         |              |                                            |
| Availability Codes    |              |                                            |
| Dist                  | Avail and/or |                                            |
| <b>A-1</b>            | Special      |                                            |

## 1. INTRODUCTION.

Often one wishes to estimate the survival functions (1-cumulative distribution function) of populations from possibly censored data when nothing is known concerning the functional form of the survival functions. Appealing estimates in this situation have been obtained by Kaplan and Meier (1958) and are usually referred to as Kaplan-Meier product limit estimators. Although Kaplan and Meier restricted themselves to discrete distributions, Johansen (1978) has shown that the product limit estimator is a maximum likelihood estimator (mle) in the class of all distributions under the generalized maximum likelihood framework developed by Kiefer and Wolfowitz (1956).

Many times one may have a situation where, logically, distributions must be stochastically ordered. For example, if a mechanical device is improved, the probability of survival past any given time for the improved device should not be less than that for the original device. In this situation, it seems reasonable to require that estimates of the survival functions should also satisfy this stochastic ordering. Brunk et al. (1966) have given mle's for two stochastically ordered cdf's for uncensored independent random samples. Dykstra (1982) has considered this problem for the case of right censored data and has given the mle's in the form of Kaplan-Meier product limit estimators with modified data. In this paper, we are able to find the mle's of  $N \geq 2$  survival functions when a linear stochastic ordering exists among them. While the solution has a nice characterization, obtaining the actual estimates is quite difficult. An iterative algorithm depending only upon the solution to the pairwise problem is given and is shown to converge to an actual mle.

## 2. THE PROBLEM

We shall assume that we have independent random samples, possibly with right censored observations, from  $N$  discrete populations. The  $N \geq 2$  populations are assumed to be stochastically ordered, so that the corresponding survival functions satisfy

$$(2.1) \quad P_1 \stackrel{st}{\geq} P_2 \stackrel{st}{\geq} \dots \stackrel{st}{\geq} P_N.$$

(We say  $P_i \stackrel{st}{\geq} P_j$  if  $P_i(t) \geq P_j(t)$  for all  $t$ .) The problem is to find nonparametric mle's of the survival functions subject to the constraints in (2.1). We will without loss of generality (WLOG) assume that  $P_1(0) = P_2(0) = \dots = P_N(0) = 1$ , so that all observations are positive.

Complete observations (failures) occur on a subset of the times  $S_1 < S_2 < \dots < S_m$  ( $S_0 = 0, S_{m+1} = \infty$ ). The number of failures from the  $i^{\text{th}}$  population which occur at time  $S_j$  is denoted by  $d_{ij}$ . The number of losses (censored observations) in the interval  $[S_j, S_{j+1})$  from the  $i^{\text{th}}$  population is denoted by  $l_{ij}$ . We assume the  $l_{ij}$  losses occur at times  $L_r^{(ij)}$ ,  $r=1, \dots, l_{ij}$ , where these censoring times are fixed. (The same mle's would obtain for random censoring times which are independent of the times of failure.) Let  $n_{ij} = \sum_{r=j}^m (d_{ir} + l_{ir})$  denote the number of items from the  $i^{\text{th}}$  population surviving to just prior to  $S_j$ . We have assumed that our survival functions are discrete, but this is not really necessary, as one may argue in the context of generalized maximum likelihood (see Johansen (1978)) that our estimates need place probability only on those timepoints at which observations occur.

### 3. REDUCTION OF THE PROBLEM.

Based on the notation in Section 2, the problem is to find survival functions  $P_1, \dots, P_N$  which maximize the likelihood

$$(3.1) \quad \prod_{i=1}^N \left\{ \prod_{r=1}^{L_{i0}} P(L_r^{(i,0)}) \prod_{j=1}^m \left\{ [P_i(S_j-0) - P_i(S_j)]^{d_{ij}} \prod_{r=1}^{L_{ij}} P_i(L_r^{(i,j)}) \right\} \right\},$$

subject to the constraints

$$(3.2) \quad P_i(t) \geq P_{i+1}(t) \quad \forall t, \quad i=1, \dots, N-1.$$

For a given set of survival functions  $P_1, \dots, P_N$  satisfying (3.2), we note that the likelihood cannot be decreased and (3.2) cannot be violated if we replace  $P_1(t)$  by a discrete  $P'_1(t)$  which has possible jumps only at  $S_1, \dots, S_m$  and is such that  $P'_1(S_j) = P_1(S_j)$ . If we now replace  $P_2(t)$  by  $P'_2(t)$  defined to have possible jumps only at  $S_1, \dots, S_m$  and  $P'_2(S_j) = P_2(S_j)$ , the likelihood cannot decrease and (3.2) cannot be violated.

Continuing this reasoning, we see it will suffice to maximize the expression

$$\prod_{i=1}^N \prod_{j=1}^m [P_i(S_{j-1}) - P_i(S_j)]^{d_{ij}} P_i(S_j)^{L_{ij}}$$

subject to  $P_i(S_j) \geq P_{i+1}(S_j) \quad \forall j, \quad i=1, \dots, N-1$  among those survival functions which place probability only at the points  $S_1, S_2, \dots, S_m$ .

(We note that maximum likelihood estimates need not be uniquely defined if the last observation from a population is a loss. We will avoid this ambiguity by requiring our maximum likelihood estimates of the survival functions to be as small as possible.) Equivalently, we wish to maximize

$$\prod_{i=1}^N \prod_{j=1}^m \left[ 1 - \frac{P_i(S_j)}{P_i(S_{j-1})} \right]^{d_{ij}} \left[ \frac{P_i(S_{j-1})}{P_i(S_{j-2})} \cdots P_i(S_1) \right]^{d_{ij}} \left[ \frac{P_i(S_j)}{P_i(S_{j-1})} \cdots P_i(S_1) \right]^{l_{ij}},$$

or letting  $p'_{ij} = \frac{P_i(S_j)}{P_i(S_{j-1})}$ , to maximize

$$\prod_{i=1}^N \prod_{j=1}^m (1 - p'_{ij})^{d_{ij}} p_{ij}'^{l_{ij}} \prod_{r < j} p_{ir}'^{d_{ij} + l_{ij}}$$

subject to  $\prod_{r=1}^j p_{ir}' \geq \prod_{r=1}^j p_{i+1,r}', \quad j=1, \dots, m; \quad i=1, \dots, N-1.$

Finally, recalling that  $n_{ij} = \sum_{r=j}^m (d_{ir} + l_{ir})$ , letting  $p_{ij} = \ln p_{ij}'$ , and considering the log of the likelihood, it will suffice to maximize

$$(3.3) \quad f(p_1, \dots, p_N) = \sum_{i=1}^N \sum_{j=1}^m d_{ij} \ln(1 - e^{p_{ij}}) + (n_{ij} - d_{ij}) p_{ij}$$

subject to the constraints

$$(3.4) \quad \sum_{r=1}^j p_{ir} \geq \sum_{r=1}^j p_{i+1,r}, \quad 0 \geq p_{ij} \geq -\infty$$

for  $j=1, \dots, m; \quad i=1, \dots, N-1.$

The problem has been reduced to maximizing a concave function subject to linear inequality constraints.

Since we are maximizing a bounded concave function over a closed convex region, there must exist a solution  $\hat{p} = (\hat{p}_1, \dots, \hat{p}_N)$  which satisfies the constraints in equation (3.4) and maximizes equation (3.3). If the active constraints (those where equality holds) of equation (3.4) needed for a solution  $\hat{p}$  were known, the number of variables could be reduced by using certain of the expressions in (3.4) with equality holding. Then expression (3.3) could be maximized by setting the appropriate partial derivatives equal to zero, or by other means, and solving for the remaining independent variables. Of course, determining which constraints are the active ones is, in general, a very difficult problem. However, by noting that certain equality constraints must hold, and then maximizing equation (3.3) subject to an arbitrary fixed set of equality constraints in (3.4), the general form of the solution  $\hat{p}$  can be determined.

Notice that the constraints  $\sum_{l=1}^{a_1} p_{il} = \sum_{l=1}^{a_1} p_{i+1,l}, \sum_{l=1}^{a_2} p_{il} = \sum_{l=1}^{a_2} p_{i+1,l}, \dots,$

$\sum_{l=1}^{a_k} p_{il} = \sum_{l=1}^{a_k} p_{i+1,l}$  can be written as  $\sum_{l=1}^{a_1} p_{il} = \sum_{l=1}^{a_1} p_{i+1,l}, \sum_{l=a_1+1}^{a_2} p_{il} =$

$\sum_{l=a_1+1}^{a_2} p_{i+1,l}, \dots, \sum_{l=a_{k-1}+1}^{a_k} p_{il} = \sum_{l=a_{k-1}+1}^{a_k} p_{i+1,l}$ , so that an arbitrary  $p_{ij}$

need be present in at most two active constraints. Thus suppose  $p_{ij}$  is present in

$$p_{i+1,b} = \sum_{l=a}^b p_{il} - \sum_{l=a}^{b-1} p_{i+1,l},$$

and

$$p_{i,s} = \sum_{l=r}^s p_{i-1,l} - \sum_{l=r}^{s-1} p_{i,l}, \quad a \leq j \leq b \quad \text{and} \quad r \leq j \leq s.$$

Also suppose that  $d_{ij} > 0$ , and no other parameter equals  $p_{ij}$ . Then, with appropriate substitution, the partial derivative with respect to  $p_{ij}$  of equation (3.3) set equal to zero yields

$$(3.5) \quad -\frac{d_{ij} e^{p_{ij}}}{(1-e^{p_{ij}})} + n_{ij} - d_{ij} = h_{ij} - k_{ij}$$

where

$$k_{ij} = -\frac{d_{i+1,b} e^{p_{i+1,b}}}{(1-e^{p_{i+1,b}})} + (n_{i+1,b} - d_{i+1,b})$$

and

$$h_{ij} = -\frac{d_{is} e^{p_{is}}}{(1-e^{p_{is}})} + (n_{is} - d_{is}).$$

Of course it doesn't matter which variables we choose as the dependent variables. It follows that  $k_{ij}$  must have constant value for  $a \leq j \leq b$ , providing  $d_{i+1,j} > 0$ , and  $h_{ij}$  must have constant value for  $r \leq j \leq s$  if  $d_{ij} > 0$ . Solving the equation for  $p_{ij}$  yields a solution of the form

$$(3.6) \quad p_{ij}^* = \ln\left(\frac{n_{ij} - d_{ij} + k_{ij} - h_{ij}}{n_{ij} + k_{ij} - h_{ij}}\right).$$

Inspection of  $h_{ij}$  and  $k_{ij}$  will reveal that  $h_{ij} = k_{i-1,j}$  if both  $d_{ij}$  and  $d_{i-1,j}$  are positive. If  $d_{ij} = 0$ , that part of the objective function (3.3) involving  $p_{ij}$  is linear in  $p_{ij}$  with nonnegative slope.

Intuitively,  $p_{ij}^*$  should be as large as possible, and hence equal to zero unless our constraints will not allow this. This would happen for example if  $d_{11} > 0$ , but  $d_{21} = d_{22} = 0$ . Then our constraints would require that  $p_{11} = p_{21}$  and  $p_{12} = p_{22}$ . We make these substitutions, and then solve by derivative methods. We will show later that these solutions are still of the form (3.6) for appropriate  $k_{ij}$ , and appropriate definitions of indeterminate forms. Once  $d_{ij} > 0$ ,  $p_{il}^* = 0$  if  $d_{il} = 0$ ,  $l > j$ . Note that if  $p_{il}^* = 0$  and  $d_{il} = 0$ , then  $h_{il}$  is indeterminate, and hence does not violate our claim of constant  $h_{ij}$  over  $r \leq j \leq s$ . In any event,  $p_{ij}^*$  can always be written in the form of (3.6) if indeterminate values are properly defined.

Since the true solution,  $\hat{p}$ , is given by  $p^*$  for some set of active constraints,  $\hat{p}$  is necessarily of the form

$$\begin{aligned}
 \hat{p}_{1j} &= \ln \left( \frac{n_{1j} - d_{1j} + k_{1j}}{n_{1j} + k_{1j}} \right) \\
 (3.7) \quad \hat{p}_{ij} &= \ln \left( \frac{n_{ij} - d_{ij} + k_{ij} - k_{i-1,j}}{n_{ij} + k_{ij} - k_{i-1,j}} \right) \quad 2 \leq i \leq N-1 \\
 \hat{p}_{Nj} &= \ln \left( \frac{n_{Nj} - d_{Nj} - k_{N-1,j}}{n_{Nj} - k_{N-1,j}} \right)
 \end{aligned}$$

for appropriate values of  $k_{ij}$  and appropriate definitions of indeterminate forms.

Now, the problem is to identify the  $k_{ij}$ 's. Heuristically, since the  $k_{ij}$ 's may be interpreted as playing the same role as the  $n_{ij}$ 's, the estimates are obtained by transferring unfailed items from one sample to the

next stochastically larger sample, and then using the Kaplan-Meier (maximum likelihood) estimate for each survival function. It will then turn out that the  $k_{ij}$ 's will never be negative, which seems reasonable in light of our one-sided restrictions. Of course probability could be transferred from one population to several larger populations, but (3.7) indicates that we do not need to think of things this way.

Even if the active constraints were known, a large number of simultaneous nonlinear equations would have to be solved to find  $\hat{p}$ . In the following sections, the  $k_{ij}$ 's will be characterized and a method using this characterization will be given to estimate the  $N$  survival functions, avoiding the simultaneous nonlinear equations problem.

#### 4. KUHN-TUCKER CHARACTERIZATION

The characterization of the  $k_{ij}$ 's requires some theory from convex analysis. The problem of estimating  $N$  survival functions subject to the stochastic orderings given by equation (3.2) may be written in terms of a particular ordinary convex programming problem. An ordinary convex programming problem is to find the vector  $p$  which minimizes a given convex function  $f_0(p)$  subject to the constraints  $f_j(p) \leq 0$ ,  $j=1, \dots, n$  where each  $f_j$  is a convex function on  $\mathbb{R}^m$ . By letting  $f_0(p) = -f(p)$  from equation (3.3), and letting  $f_{ij}(p) = \sum_{l=1}^j p_{i+1,l} - \sum_{l=1}^j p_{il}$ ,  $i=1, \dots, N-1$ ;  $j=1, \dots, m$ , our problem is couched in the terms of an ordinary convex programming problem. We call  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$  a Kuhn-Tucker vector if  $\lambda_j \geq 0$  for  $j=1, \dots, n$ , and if the unrestricted infimum of  $h(p) = f_0(p) + \lambda_1 f_1(p) + \dots + \lambda_n f_n(p)$  is equal to the restricted infimum of  $f_0(p)$ . The following theorem from Rockafellar (1970) characterizes Kuhn Tucker vectors, and will be useful in characterizing the  $k_{ij}$ 's.

Theorem 4.1. Let  $\hat{\lambda}$  and  $\hat{p}$  be vectors in  $\mathbb{R}^n$  and  $\mathbb{R}^m$  respectively. In order that  $\hat{\lambda}$  is a Kuhn-Tucker vector for our problem and  $\hat{p}$  is an optimal solution, the following conditions are necessary and sufficient.

$$(4.1) \quad \begin{aligned} &\text{a. } \hat{\lambda}_j \geq 0, \quad f_j(\hat{p}) \leq 0, \quad \text{and} \quad \hat{\lambda}_j f_j(\hat{p}) = 0 \quad \text{for } j = 1, \dots, n, \\ &\text{b. } 0 \in [\partial f_0(\hat{p}) + \hat{\lambda}_1 \partial f_1(\hat{p}) + \dots + \hat{\lambda}_n \partial f_n(\hat{p})] \end{aligned}$$

where  $\partial f_j(p)$  is the set of all subgradients of  $f_j$  at  $\hat{p}$ .

Since the true solution and a Kuhn-Tucker vector must satisfy the

subgradient equation in (4.1b), the general form of a Kuhn-Tucker vector can be found for our problem using the general form of the solution given in equation (3.7).

Theorem 4.2. If  $\hat{p}_{ij}$  denotes the solution to the problem specified in equations (3.3) and (3.4), then a Kuhn-Tucker vector corresponding to the constraints  $f_{ij} = \sum_{l=1}^j p_{i+1,l} - \sum_{l=1}^j p_{il} \leq 0$  is given by

$$\lambda_{ij} = \begin{cases} k_{ij}^{-k_{i,j+1}}, & 1 \leq j < m \\ k_{im}, & j = m \end{cases}$$

when the solution  $\hat{p}$  is specified in the form given by (3.7).

Proof: In the ordinary convex program characterization of our problem,

$$f_0(p) = - \left\{ \sum_{i=1}^N \sum_{j=1}^m d_{ij} \ln(1 - e^{p_{ij}}) + (n_{ij} - d_{ij}) p_{ij} \right\}$$

which we want to minimize subject to the constraints

$$f_{ij}(p) = \sum_{l=1}^j p_{i+1,l} - \sum_{l=1}^j p_{il} \leq 0 \text{ for } i=1, \dots, N-1; j=1, \dots, m.$$

Notice that  $f_0(p)$  and  $f_{ij}(p)$ ,  $i=1, \dots, N-1; j=1, \dots, m$  are differentiable functions of  $p_{i'j'}$  for  $i'$  and  $j'$  so that equation (4.1b) can be written in matrix notation. (Of course the  $n$  in Theorem 4.1 is now  $m(N-1)$ .) Equation (4.1b) then becomes  $0 = F_0 + F\lambda$ , where  $F_0$  is the  $Nm \times 1$  vector of derivatives of  $f_0(p)$  with respect to  $p_{ij}$ ,  $\lambda$  is an

$(N-1)m \times 1$  Kuhn-Tucker vector and  $F$  is the  $Nm \times (N-1)m$  matrix of partial derivatives of  $f_{ij}$  with respect to  $p_{i,j}$ , and  $\tilde{0}$  is an  $Nm \times 1$  vector of zeros.

Equivalently, we must have  $F\lambda = -F_0$ , and since  $F$  is of full column rank which is less than the number of rows, any solution to our equation must be unique. We may express  $F$  as

$$\begin{pmatrix} -T & 0 & 0 & \cdots & 0 \\ T & -T & 0 & \cdots & \vdots \\ 0 & T & -T & \cdots & \\ \vdots & \vdots & & & \\ \vdots & \vdots & & T & -T \\ 0 & 0 & \cdots & & T \end{pmatrix}$$

where  $T^{m \times m}$  is an upper triangular matrix of ones. We note that  $F_0$  has the form

$$\begin{bmatrix} -k_{11} \\ \vdots \\ -k_{1m} \\ \vdots \\ k_{i1} - k_{i+1,1} \\ \vdots \\ k_{im} - k_{i+1,m} \\ \vdots \\ k_{N-1,1} \\ \vdots \\ k_{N-1,m} \end{bmatrix} \quad \text{where } i = 1, \dots, N-2.$$

It can be verified that a conditional inverse of  $F$  is given by

$$F^C = \begin{pmatrix} T^{-1} & 0 & 0 & \cdots \\ T^{-1} & T^{-1} & 0 & \cdots \\ T^{-1} & T^{-1} & T^{-1} & \cdots \\ \vdots & \vdots & \vdots & \\ T^{-1} & T^{-1} & T^{-1} & \cdots \end{pmatrix}$$

where

$$T^{-1} = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -1 & & \\ \vdots & \vdots & \vdots & \cdots & 1 & -1 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

Careful calculation will verify that  $\lambda = -F^C F_0$  is of the desired form and a solution to equation (4.1b).

Conditions (4.1a) must hold if the  $\lambda_{ij}$ 's are to form a Kuhn-Tucker vector. If  $\lambda_{ij} \geq 0$ , the  $k_{ij}$ 's occurring in equation (3.7) must be such that

$$(4.2) \quad k_{i1} \geq k_{i2} \geq \cdots \geq k_{im} \geq 0 \quad \text{for all } i.$$

Conversely, if equation (4.2) holds, the  $\lambda_{ij}$  will be nonnegative.

The condition that  $\lambda_i f_i(\hat{p}) = 0$  in Theorem 4.1, together with the monotonicity requirement given in equation (4.2) leads to a characterization for the solution provided certain indeterminate forms are properly defined. Thus we have the following theorem.

Theorem 4.3. A solution to the problem of maximizing  $f(p)$  given in equation (3.3) under the constraints in equation (3.4) are given by the  $\hat{p}_{ij}$ 's described in equation (3.7) iff the  $k_{ij}$ 's are such that:

- a) the  $\hat{p}_{ij}$ 's in (3.7) satisfy the constraints in (3.4) and  $-\infty \leq \hat{p}_{ij} \leq 0$ ,
- b)  $k_{i1} \geq k_{i2} \geq \dots \geq k_{im} \geq 0$  for  $i = 1, 2, \dots, N-1$ , and
- c) whenever  $k_{ij} > k_{i,j+1}$ ,  $\sum_{l=1}^j \hat{p}_{il} = \sum_{l=1}^j \hat{p}_{i+1,l}$

provided some indeterminate forms are appropriately defined.

Indeterminate forms occur when  $d_{ij} = 0$  for  $i_0 < i$  and  $j \leq j_0$  and  $d_{i_0, j_0} > 0$ . In this case,

$$p_{i_0, j_0} = p_{i_0+1, j_0} = \dots = p_{N, j_0} = \ln \left[ \frac{\sum_{i=i_0}^N n_{i, j_0}^{-d_{i_0, j_0}} - k_{i_0-1, j_0}}{\sum_{i=i_0}^N n_{i, j_0}^{-k_{i_0-1, j_0}}} \right]$$

as mentioned earlier. Notice that this is consistent with (3.7) where we take

$$k_{ij} = \sum_{l=i+1}^N n_{lj}, \quad i > i_0, j \leq j_0$$

since this leads to indeterminate forms.

## 5. OBTAINING THE $\hat{p}_{ij}$

Even though Theorem 4.3 characterizes the solution to the problem, it does not provide any means of obtaining the solution in general. However, for the case  $N = 2$ , Dykstra (1982) has found closed form expressions for the  $\hat{p}_{ij}$  as well as an efficient algorithm for calculating them. We will elaborate on this procedure, and then give an iterative procedure based only on this pairwise scheme which will converge to the general solution. This will give us a means of solving the general problem.

To solve the pairwise problem, one need first identify the integer  $j_0$  such that  $d_{2j} = 0$ ,  $j < j_0$ , but  $d_{2,j_0} > 0$ . Then

$$(5.1) \quad \hat{p}_{1j} = \hat{p}_{2j} = \ln \left( \frac{n_{1j} + n_{2j} - d_{1j}}{n_{1j} + n_{2j}} \right)$$

(which requires that  $k_{1j} = n_{2j}$ ) for  $j < j_0$ . For  $j_0 \leq a \leq b$ , we let  $k_{a,b}$  denote the solution to

$$\sum_{j=a}^b \ln \left( \frac{n_{1j} - d_{1j} + k}{n_{1j} + k} \right) = \sum_{j=a}^b \ln \left( \frac{n_{2j} - d_{2j} - k}{n_{2j} - k} \right)$$

if it exists positively and 0 otherwise. Then Dykstra has shown the  $\hat{p}_{ij}$  which maximize (3.3) subject to the linear constraints  $\sum_{i=1}^j p_{1j} \geq \sum_{i=1}^j p_{2j}$  are of the form

$$(5.2) \quad \hat{p}_{1j} = \ln \left( \frac{n_{1j} - d_{1j} + \hat{k}_{1j}}{n_{1j} + \hat{k}_{1j}} \right); \quad \hat{p}_{2j} = \ln \left( \frac{n_{2j} - d_{2j} - \hat{k}_{1j}}{n_{2j} - \hat{k}_{1j}} \right)$$

where  $\hat{k}_{1j} = \min_{a \leq j} \max_{j \leq b} k_{a,b}$  for  $j \geq j_0$ . The restricted mle's of  $P_1(t)$  and  $P_2(t)$  are then given by

$$(5.3) \quad \hat{P}_i(t) = \exp \left[ \sum_{j; S_j \leq t} \hat{p}_{ij} \right], \quad i = 1, 2.$$

We shall take  $\hat{P}_i(t)$  to be as small as possible subject to the order restrictions so as to ensure uniqueness of the estimates. Using the fact that  $k_{11} \geq k_{12} \geq \dots \geq k_{1m}$ , Dykstra has derived an algorithm which can be used to compute the mle of  $p_{1j}$  and  $p_{2j}$  for  $j = 1, \dots, m$ . This algorithm is given below:

#### Pairwise Algorithm

Algorithm for two survival functions  $P_1(t) \geq P_2(t)$ :

1. Define  $\hat{p}_{ij}$  (and  $k_{1j}$ ) as in (5.1) for  $j < j_0$ .
2. Find the largest  $j_1 \geq j_0$  such that the  $k_{1j_1} > 0$  which solves

$$\sum_{j=j_0}^{j_1} \ln \left( \frac{n_{1j} - d_{1j} + k}{n_{1j} + k} \right) = \sum_{j=j_0}^{j_1} \ln \left( \frac{n_{2j} - d_{2j} - k}{n_{2j} - k} \right)$$

is maximized. Then let  $k_{1j} = k_{1j_1}$  for  $j_0 \leq j \leq j_1$ .

3. Find the largest  $j_2 > j_1$  such that the  $k_{1j_2} > 0$  which solves

$$\sum_{j=j_1+1}^{j_2} \ln \left( \frac{n_{1j} - d_{1j} + k}{n_{1j} + k} \right) = \sum_{j=j_1+1}^{j_2} \ln \left( \frac{n_{2j} - d_{2j} - k}{n_{2j} - k} \right)$$

is maximized. Let  $k_{1j} = k_{1j_2}$  for  $j_1 < j \leq j_2$ .

4. Continue until either  $j_h = m$  or no such positive  $k_{1j_h}$  exists, in which case let  $k_{1j} = 0$  for  $j > j_{h-1}$ . Then  $\hat{p}_{ij}$  and  $\hat{P}_i(t)$  are given as in (5.2) and (5.3) respectively.

The  $N$  sample problem can be solved using an iterative procedure based upon the pairwise algorithm given previously. The pairwise method is used to massage (readjust) the data until an equilibrium point is reached. Letting  $k_{ij}^{(l)}$  be the estimate of  $k_{ij}$  found during the  $l^{\text{th}}$  iteration of the algorithm, the  $N$  sample algorithm is given as follows:

1. Find the  $k_{1,j}^{(1)}$ 's corresponding to  $P_1 \geq P_2$  using the pairwise algorithm given previously.
2. Incorporate  $k_{1,j}^{(1)}$  into  $n_{1,j}$  and  $n_{2,j}$  by replacing  $n_{1,j}$  by  $n_{1,j} + k_{1,j}^{(1)}$  and  $n_{2,j}$  by  $n_{2,j} - k_{1,j}^{(1)}$ .
3. Using the adjusted data, find the  $k_{2,j}^{(1)}$ 's corresponding to  $P_2 \geq P_3$  using the pairwise algorithm.
4. Replace  $n_{2,j} - k_{1,j}^{(1)}$  by  $n_{2,j} + k_{2,j}^{(1)} - k_{1,j}^{(1)}$  and  $n_{3,j}$  by  $n_{3,j} - k_{2,j}^{(1)}$ .
5. Continue down the chain of survival functions until the  $N-1^{\text{th}}$  and  $N^{\text{th}}$  samples have been compared, and  $n_{N-1,j}$  has been replaced by  $n_{N-1,j} - k_{N-2,j}^{(1)} + k_{N-1,j}^{(1)}$  and  $n_{N,j}$  has been replaced by  $n_{N,j} - k_{N-1,j}^{(1)}$ .
6. The procedure starts again at the top of the chain. After setting  $k_{1,j}^{(1)} = 0$  for  $j=1, \dots, m$ , find the  $k_{1,j}^{(2)}$  corresponding to  $P_1 \geq P_2$  using the pairwise algorithm.
7. Incorporate  $k_{1,j}^{(2)}$  into the data by replacing  $n_{1,j}$  by  $n_{1,j} + k_{1,j}^{(2)}$  and  $n_{2,j} + k_{2,j}^{(1)}$  by  $n_{2,j} + k_{2,j}^{(1)} - k_{1,j}^{(2)}$ .
8. After setting  $k_{2,j}^{(1)} = 0$  for  $j=1, \dots, m$ , use the pairwise algorithm to find  $k_{2,j}^{(2)}$  for  $P_2 \geq P_3$  and incorporate  $k_{2,j}^{(2)}$  by replacing

$$n_{2,j} - k_{1,j}^{(2)} \text{ by } n_{2,j} + k_{2,j}^{(2)} - k_{1,j}^{(2)} \text{ and } n_{3,j} + k_{3,j}^{(1)} \text{ by } n_{3,j} + k_{3,j}^{(1)} - k_{2,j}^{(2)}.$$

9. Continue down the chain, taking out the effect of the comparison  $(k_{ij}^{(1)})$  in the previous iteration, recomparing  $P_i \geq P_{i+1}$ , and then reincorporating the new  $k_{ij}^{(2)}$ 's.
10. When the bottom of the chain is reached, start with this procedure again at the top of the chain and continue until achieving convergence.

The estimated  $\hat{p}_{ij}$ 's are then obtained from (3.7) and the latter part of Theorem 4.3 with the  $k_{ij}^{(L)}$  obtained above entered in. We will let  $\hat{p}_{ij}^{(L)}$  indicate (3.7) with the  $k_{ij}^{(L)}$ 's entered.

Theorem 5.1. The procedure discussed above must yield values of  $\hat{p}_{ij}^{(L)}$  which converge for every  $i$  and  $j$  as  $L \rightarrow \infty$ . Moreover, these limiting values solve the problem of maximizing  $f(p)$  subject to the constraints  $\sum_{l=1}^j p_{il} \geq \sum_{l=1}^j p_{i+1,l}$  for  $j=1, \dots, m$ ;  $i=1, \dots, N-1$ .

Proof: Consider the first step of the second cycle. We obtain  $k_{1j}^{(2)}$  by using the pairwise algorithm for  $P_1 \geq P_2$  with the data  $n_{1j}$ , and  $n_{2j} + k_{2j}^{(1)}$  in place of  $n_{1j}$  and  $n_{2j}$ . We need to consider two cases:

1. If  $d_{2L} = 0$  for  $L \leq j$ , then  $k_{1j}^{(2)} = n_{2j} + k_{2j}^{(1)} \geq n_{2j} = k_{1j}^{(1)}$ .
2. Otherwise  $k_{1j}^{(2)} = \min_{a \leq j} \max_{j \leq b} k'_{a,b}$  where  $k'_{a,b}$  is positive and solves

$$\sum_{j=a}^b \ln \left( \frac{n_{1j} - d_{1j} + k}{n_{1j} + k} \right) = \sum_{j=a}^b \ln \left( \frac{n_{2j} + k_{2j}^{(1)} - d_{2j} - k}{n_{2j} + k_{2j}^{(1)} - k} \right)$$

or is zero.

If  $k_{a,b}$  is positive and solves

$$\sum_{j=a}^b \ln \left( \frac{n_{1j} - d_{1j} + k}{n_{1j} + k} \right) = \sum_{j=a}^b \ln \left( \frac{n_{2j} - d_{2j} - k}{n_{2j} - k} \right),$$

it easily follows that  $k'_{a,b} \geq k_{a,b}$  for all  $a, b$ , and hence

$$k_{1j}^{(2)} \geq k_{1j}^{(1)}.$$

Similar arguments suffice for other  $i$ , and we may conclude that  $k_{ij}^{(l)}$  is nondecreasing in  $l$  for all  $i, j$ . Since the  $k_{ij}^{(l)}$  are uniformly bounded (say by  $\sum_{i=1}^N n_{il}$ ), we know there exists  $k_{ij}$  such that  $k_{ij}^{(l)} \nearrow k_{ij}$ .

Finally, noting that the  $\hat{p}_{ij}$  in (3.7) are continuous functions of the  $k_{ij}$ , and that the  $k_{ij}^{(l)}$  correspond to solutions of the pairwise problems, we can argue that the  $\hat{p}_{ij}$  defined by (3.7) for the limiting values  $k_{ij}$  satisfy the conditions of Theorem 4.3, and hence solve the general problem.

We mention again the appealing form of the estimators. They are still of the Kaplan-Meier form, where, heuristically, unfailed items from a population are transferred to the next stochastically larger population.

## 6. EXAMPLE

This example involves a large clinical trial carried out by the Radiation Therapy Oncology Group cited in Kalbfleisch and Prentice (1980). The analysis was done on only a small part of the data, and females were deleted to make the data more homogeneous. Patients diagnosed with squamous carcinoma of the oropharynx were classified by the degree to which the regional lymph nodes were affected by this disease.

We let  $P_1$  be the survival function corresponding to those patients for whom the disease had no effect on the lymph nodes. Correspondingly, we let  $P_2$ ,  $P_3$ , and  $P_4$  be the survival functions of patients with increased effects of the disease on the lymph nodes. The data for each survival function is given in Table 1.

Because of the nature of the disease, one would certainly expect a linear ordering with respect to lymphatic involvement. Thus an ordering of the form  $P_1 \stackrel{st}{\geq} P_2 \stackrel{st}{\geq} P_3 \stackrel{st}{\geq} P_4$  seems inherently reasonable.

The unrestricted mle's (Kaplan-Meier) for these survival functions are given in Table 3. Notice that with respect to the expected orderings, there are reversals so that estimates requiring these orderings will be different. The algorithm given in Section 5 was applied to the problem. The  $k_{ij}$  values found in the first two iterations, in the order in which they were found, along with the essentially limiting values obtained after 20 cycles are given in Table 2. (The procedure used here actually worked pairwise from bottom to top rather than from top to bottom as indicated in the paper. Of course, the limiting values will be identical.) The values of these  $k_{ij}^{(L)}$  indicate how much "smoothing" is occurring between populations.

The mle's subject to the stochastic ordering  $P_1 \stackrel{st}{\geq} P_2 \stackrel{st}{\geq} P_3 \stackrel{st}{\geq} P_4$ , obtained via our algorithm are displayed in Table 4. Note that the stochastic ordering requirement has brought substantial change in our estimates.

TABLE 1

Survival Times in Days of Patients with Squamous Carcinoma in the Oropharynx  
with Various Degrees of Lymph Node Deterioration (+ denotes censored observation).

Degrees of deterioration of lymph nodes.

| i = 1                                      | i = 2 | i = 3 | i = 4 |
|--------------------------------------------|-------|-------|-------|
| 167                                        | 216   | 105   | 94    |
| 238                                        | 324   | 222   | 99    |
| 276+                                       | 338   | 279   | 112   |
| 296                                        | 347   | 395   | 127   |
| 324                                        | 599   | 465   | 134   |
| 351                                        | 763   | 546   | 147   |
| 372                                        | 929   | 915   | 192   |
| 374                                        | 1086+ | 918   | 219   |
| 404                                        | 1092  | 1058+ | 255   |
| 445+                                       | 1317+ | 1455+ | 262   |
| 541                                        | 1609+ | 1644+ | 274   |
| 560                                        |       |       | 293   |
| 943                                        |       |       | 307   |
| 998+                                       |       |       | 327   |
| 1234+                                      |       |       | 334   |
| 1460+                                      |       |       | 363   |
| 1823+                                      |       |       | 407   |
|                                            |       |       | 413+  |
|                                            |       |       | 459   |
| (The degree of deterioration (i) equals    |       |       | 517   |
| N-stage tumor classification in            |       |       | 532   |
| Kalbfleisch and Prentice (1980) plus one.) |       |       | 544   |
|                                            |       |       | 672   |
|                                            |       |       | 696   |
|                                            |       |       | 800   |
|                                            |       |       | 914   |
|                                            |       |       | 1312+ |
|                                            |       |       | 1446+ |
|                                            |       |       | 1472+ |

TABLE 2

Values of the  $k_{ij}^{(2)}$  obtained for survival data in Table 1.

First cycle:

|                  |        |                  |                    |
|------------------|--------|------------------|--------------------|
| $k_{3j}^{(1)} =$ | 2.333  | $j \leq 3$       | $T \leq 105$       |
|                  | 0      | $j > 3$          | $T > 105$          |
| $k_{2j}^{(1)} =$ | 1.853  | $29 < j \leq 29$ | $374 < T \leq 374$ |
|                  | 1.205  | $j \leq 53$      | $T \leq 1609$      |
|                  | 0      | $j > 53$         | $T > 1609$         |
| $k_{1j}^{(1)} =$ | 12.853 | $9 < j \leq 9$   | $192 < T \leq 192$ |
|                  | 4.380  | $j \leq 40$      | $T \leq 560$       |
|                  | 0      | $j > 40$         | $T > 560$          |

Second cycle:

|                  |          |                  |                     |
|------------------|----------|------------------|---------------------|
| $k_{3j}^{(2)} =$ | 3.569    | $j \leq 3$       | $T \leq 105$        |
|                  | 0        | $j > 3$          | $T > 105$           |
| $k_{2j}^{(2)} =$ | 5.481    | $11 < j \leq 11$ | $219 < T \leq 219$  |
|                  | 3.752    | $j \leq 29$      | $T \leq 374$        |
|                  | 1.205    | $29 < j \leq 53$ | $374 < T \leq 1609$ |
|                  | 0        | $j > 53$         | $T > 1609$          |
| $k_{1j}^{(2)} =$ | 12.853   | $9 < j \leq 9$   | $192 < T \leq 192$  |
|                  | 5.847    | $j \leq 40$      | $T \leq 560$        |
|                  | 0        | $j > 40$         | $T > 560$           |
|                  | $\vdots$ |                  |                     |

Twentieth cycle:

|                   |        |                  |                     |
|-------------------|--------|------------------|---------------------|
| $k_{3j}^{(20)} =$ | 10.525 | $3 < j \leq 3$   | $105 < T \leq 105$  |
|                   | 1.438  | $j \leq 17$      | $T \leq 279$        |
|                   | 0      | $j > 17$         | $T > 279$           |
| $k_{2j}^{(20)} =$ | 12.290 | $11 < j \leq 11$ | $219 < T \leq 219$  |
|                   | 5.763  | $j \leq 29$      | $T \leq 374$        |
|                   | 1.205  | $29 < j \leq 53$ | $374 < T \leq 1609$ |
|                   | 0      | $j > 53$         | $T > 1609$          |
| $k_{1j}^{(20)} =$ | 12.853 | $9 < j \leq 9$   | $192 < T \leq 192$  |
|                   | 10.371 | $j \leq 20$      | $T \leq 307$        |
|                   | 7.253  | $20 < j \leq 40$ | $307 < T \leq 560$  |
|                   | 0      | $j > 40$         | $T > 560$           |

TABLE 3  
Kaplan-Meier MLE's of survival functions from data

| Index | Time | P <sub>1</sub> | P <sub>2</sub> | P <sub>3</sub> | P <sub>4</sub> |
|-------|------|----------------|----------------|----------------|----------------|
| 0     | 0.0  | 1.0000         | 1.0000         | 1.0000         | 1.0000         |
| 1     | 94   | 1.0000         | 1.0000         | 1.0000         | .9655          |
| 2     | 99   | 1.0000         | 1.0000         | 1.0000         | .9310          |
| 3     | 105  | 1.0000         | 1.0000         | .9091          | .9310          |
| 4     | 112  | 1.0000         | 1.0000         | .9091          | .8966          |
| 5     | 127  | 1.0000         | 1.0000         | .9091          | .8621          |
| 6     | 134  | 1.0000         | 1.0000         | .9091          | .8276          |
| 7     | 147  | 1.0000         | 1.0000         | .9091          | .7931          |
| 8     | 167  | .9412          | 1.0000         | .9091          | .7931          |
| 9     | 192  | .9412          | 1.0000         | .9091          | .7586          |
| 10    | 216  | .9412          | .9091          | .9091          | .7586          |
| 11    | 219  | .9412          | .9091          | .9091          | .7241          |
| 12    | 222  | .9412          | .9091          | .8182          | .7241          |
| 13    | 238  | .8824          | .9091          | .8182          | .7241          |
| 14    | 255  | .8824          | .9091          | .8182          | .6897          |
| 15    | 262  | .8824          | .9091          | .8182          | .6552          |
| 16    | 274  | .8824          | .9091          | .8182          | .6207          |
| 17    | 279  | .8824          | .9091          | .7273          | .6207          |
| 18    | 293  | .8824          | .9091          | .7273          | .5862          |
| 19    | 296  | .8193          | .9091          | .7273          | .5862          |
| 20    | 307  | .8193          | .9091          | .7273          | .5517          |
| 21    | 324  | .7563          | .8182          | .7273          | .5517          |
| 22    | 327  | .7563          | .8182          | .7273          | .5172          |
| 23    | 334  | .7563          | .8182          | .7273          | .4828          |
| 24    | 338  | .7563          | .7272          | .7273          | .4828          |
| 25    | 347  | .7563          | .6364          | .7273          | .4828          |
| 26    | 351  | .6933          | .6364          | .7273          | .4828          |
| 27    | 363  | .6933          | .6364          | .7273          | .4483          |
| 28    | 372  | .6303          | .6364          | .7273          | .4483          |
| 29    | 374  | .5672          | .6364          | .7273          | .4483          |
| 30    | 395  | .5672          | .6364          | .6364          | .4483          |
| 31    | 404  | .5042          | .6364          | .6364          | .4483          |
| 32    | 407  | .5042          | .6364          | .6364          | .4138          |
| 33    | 459  | .5042          | .6364          | .6364          | .3762          |
| 34    | 465  | .5042          | .6364          | .5455          | .3762          |
| 35    | 517  | .5042          | .6364          | .5455          | .3386          |
| 36    | 532  | .5042          | .6364          | .5455          | .3009          |
| 37    | 541  | .4322          | .6364          | .5455          | .3009          |
| 38    | 544  | .4322          | .6364          | .5455          | .2633          |
| 39    | 546  | .4322          | .6364          | .4546          | .2633          |
| 40    | 560  | .3601          | .6364          | .4546          | .2633          |
| 41    | 599  | .3601          | .5455          | .4546          | .2633          |
| 42    | 672  | .3601          | .5455          | .4546          | .2257          |
| 43    | 696  | .3601          | .5455          | .4546          | .1881          |
| 44    | 763  | .3601          | .4546          | .4546          | .1881          |
| 45    | 800  | .3601          | .4546          | .4546          | .1505          |
| 46    | 914  | .3601          | .4546          | .4546          | .1129          |
| 47    | 915  | .3601          | .4546          | .3636          | .1129          |
| 48    | 918  | .3601          | .4546          | .2727          | .1129          |
| 49    | 929  | .3601          | .3636          | .2727          | .1129          |
| 50    | 943  | .2881          | .3636          | .2727          | .1129          |
| 51    | 1092 | .2881          | .2424          | .2727          | .1129          |
| 52    | 1472 | .2881          | .2424          | .2727          | .0000          |
| 53    | 1609 | .2881          | .0000          | .2727          | .0000          |
| 54    | 1644 | .2881          | .0000          | .0000          | .0000          |
| 55    | 1823 | .0000          | .0000          | .0000          | .0000          |

TABLE 4

24

MLE's of survival functions under order restrictions from data in Table 1.

| Index | Time | P <sub>1</sub> | P <sub>2</sub> | P <sub>3</sub> | P <sub>4</sub> |
|-------|------|----------------|----------------|----------------|----------------|
| 0     | 0.0  | 1.0000         | 1.0000         | 1.0000         | 1.0000         |
| 1     | 94   | 1.0000         | 1.0000         | 1.0000         | .9459          |
| 2     | 99   | 1.0000         | 1.0000         | 1.0000         | .8918          |
| 3     | 105  | 1.0000         | 1.0000         | .8917          | .8918          |
| 4     | 112  | 1.0000         | 1.0000         | .8917          | .8569          |
| 5     | 127  | 1.0000         | 1.0000         | .8917          | .8220          |
| 6     | 134  | 1.0000         | 1.0000         | .8917          | .7871          |
| 7     | 147  | 1.0000         | 1.0000         | .8917          | .7522          |
| 8     | 167  | .9665          | .9665          | .8917          | .7522          |
| 9     | 192  | .9665          | .9665          | .8917          | .7173          |
| 10    | 216  | .9665          | .8917          | .8917          | .7173          |
| 11    | 219  | .9665          | .8917          | .8917          | .6824          |
| 12    | 222  | .9665          | .8917          | .7346          | .6824          |
| 13    | 238  | .9299          | .8917          | .7346          | .6824          |
| 14    | 255  | .9299          | .8917          | .7346          | .6476          |
| 15    | 262  | .9299          | .8917          | .7346          | .6127          |
| 16    | 274  | .9299          | .8917          | .7346          | .5778          |
| 17    | 279  | .9299          | .8917          | .5775          | .5778          |
| 18    | 293  | .9299          | .8917          | .5775          | .5457          |
| 19    | 296  | .8917          | .8917          | .5775          | .5457          |
| 20    | 307  | .8917          | .8917          | .5775          | .5136          |
| 21    | 324  | .8477          | .7869          | .5775          | .5136          |
| 22    | 327  | .8477          | .7869          | .5775          | .4815          |
| 23    | 334  | .8477          | .7869          | .5775          | .4494          |
| 24    | 338  | .8477          | .6821          | .5775          | .4494          |
| 25    | 347  | .8477          | .5773          | .5775          | .4494          |
| 26    | 351  | .8036          | .5773          | .5775          | .4494          |
| 27    | 363  | .8036          | .5773          | .5775          | .4173          |
| 28    | 372  | .7596          | .5773          | .5775          | .4173          |
| 29    | 374  | .7156          | .5773          | .5775          | .4173          |
| 30    | 395  | .7156          | .5773          | .4925          | .4173          |
| 31    | 404  | .6716          | .5773          | .4925          | .4173          |
| 32    | 407  | .6716          | .5773          | .4925          | .3852          |
| 33    | 459  | .6716          | .5773          | .4925          | .3502          |
| 34    | 465  | .6716          | .5773          | .4075          | .3502          |
| 35    | 517  | .6716          | .5773          | .4075          | .3152          |
| 36    | 532  | .6716          | .5773          | .4075          | .2801          |
| 37    | 541  | .6245          | .5773          | .4075          | .2801          |
| 38    | 544  | .6245          | .5773          | .4075          | .2451          |
| 39    | 546  | .6245          | .5773          | .3225          | .2451          |
| 40    | 560  | .5773          | .5773          | .3225          | .2451          |
| 41    | 599  | .5773          | .5070          | .3225          | .2451          |
| 42    | 672  | .5773          | .5070          | .3225          | .2101          |
| 43    | 696  | .5773          | .5070          | .3225          | .1751          |
| 44    | 763  | .5773          | .4366          | .3225          | .1751          |
| 45    | 800  | .5773          | .4366          | .3225          | .1401          |
| 46    | 914  | .5773          | .4366          | .3225          | .1051          |
| 47    | 915  | .5773          | .4366          | .2375          | .1051          |
| 48    | 918  | .5773          | .4366          | .1526          | .1051          |
| 49    | 929  | .5773          | .3662          | .1526          | .1051          |
| 50    | 943  | .4619          | .3662          | .1526          | .1051          |
| 51    | 1092 | .4619          | .2791          | .1526          | .1051          |
| 52    | 1472 | .4619          | .2791          | .1526          | .0000          |
| 53    | 1609 | .4619          | .1525          | .1526          | .0000          |
| 54    | 1644 | .4619          | .0000          | .0000          | .0000          |
| 55    | 1823 | .0000          | .0000          | .0000          | .0000          |

FIGURE 2.

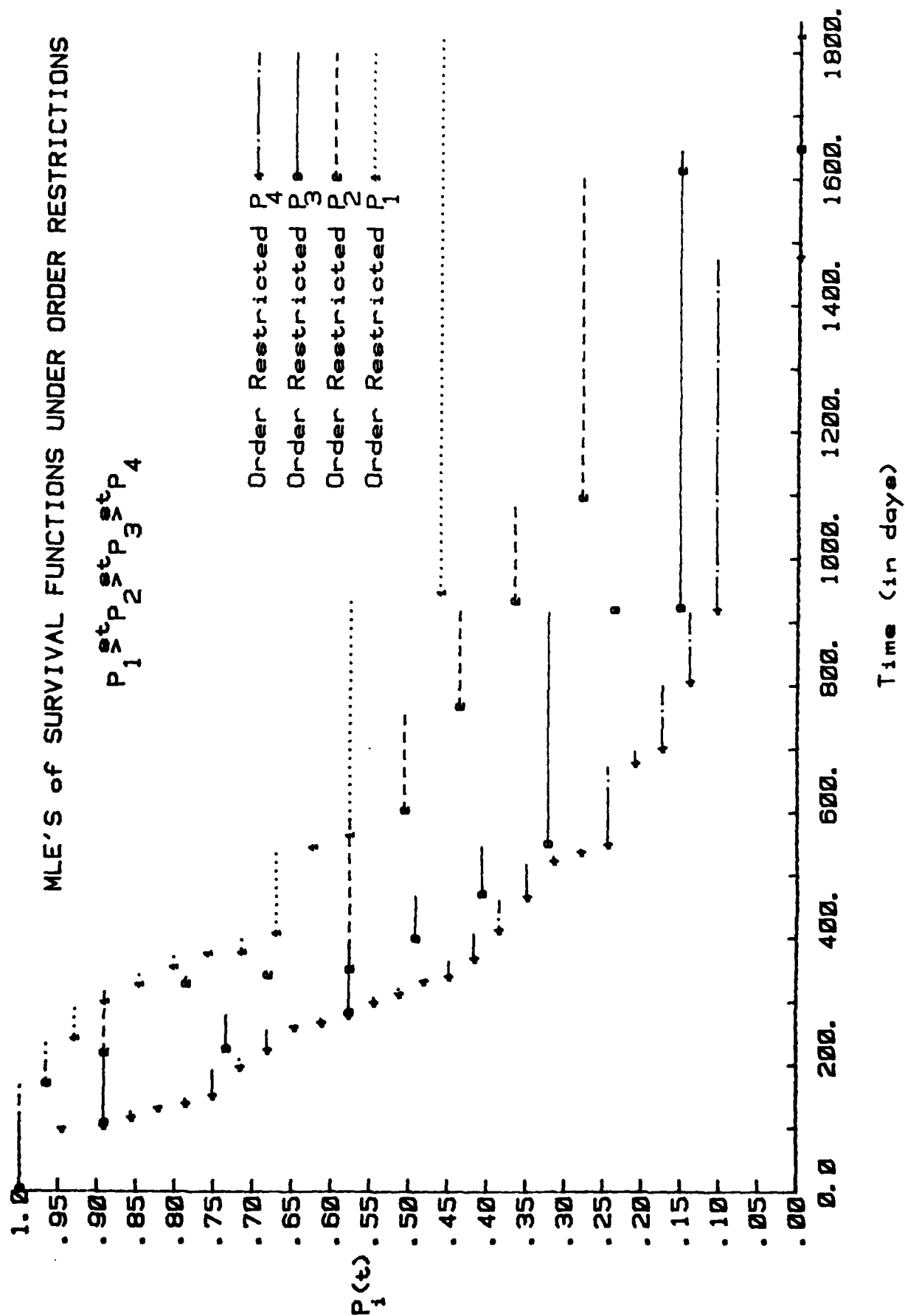
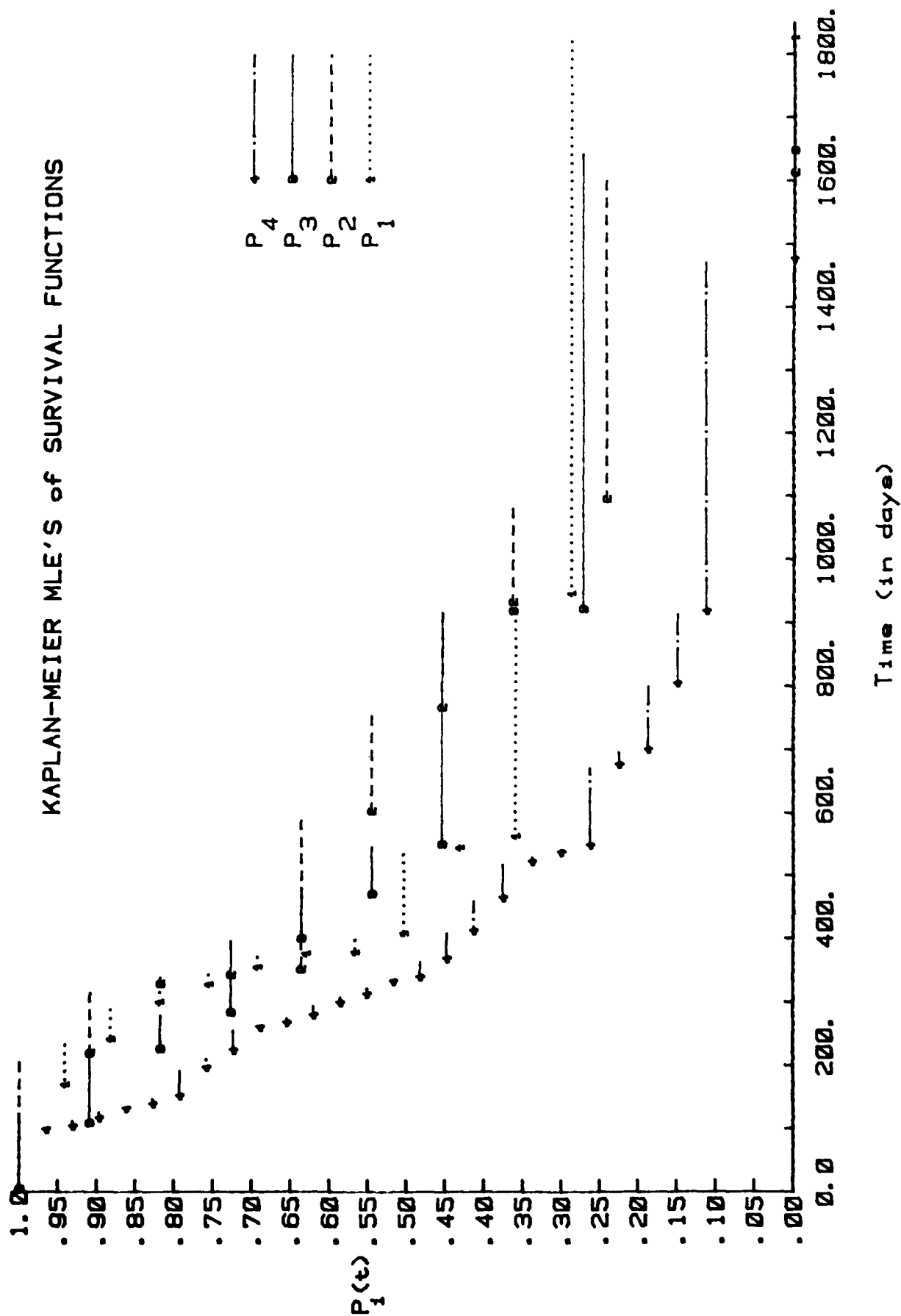


FIGURE 1.



## REFERENCES

1. Brunk, H.D., Franck, W.E., Hanson, D.L. and Hogg, R.V. (1966). Maximum likelihood estimation of the distributions of two stochastically ordered random variables. J. Amer. Statist. Assoc. 61, 1067-1080.
2. Dykstra, Richard L. (1982). Maximum likelihood estimation of the survival functions of stochastically ordered random variables. J. Amer. Statist. Assoc. 77, 621-628.
3. Johansen, Søren (1978). The product limit estimator as maximum likelihood estimator. Scand. Jour. Statist. 5, 195-199.
4. Keifer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Ann. Math. Statist. 27, 887-906.
5. Kalbfleisch, J.D. and Prentice, R.L. (1980). The Statistical Analysis of Failure Time Data. Wiley, New York.
6. Kaplan, E.L. and Meier, Paul (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457-481.
7. Rockafeller, R. Tyrrell (1970). Convex Analysis. Princeton University Press, Princeton, NJ.

| REPORT DOCUMENTATION PAGE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                      | READ INSTRUCTIONS<br>BEFORE COMPLETING FORM                   |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------|---------------------------------------------------------------|
| 1. REPORT NUMBER<br>104                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 2. GOVT ACCESSION NO.<br>AD-A145 429 | 3. RECIPIENT'S CATALOG NUMBER                                 |
| 4. TITLE (and Subtitle)<br>Maximum Likelihood Estimation of the Survival Functions of N Stochastically Ordered Random Variables.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                      | 5. TYPE OF REPORT & PERIOD COVERED<br>Technical - August 1984 |
| 7. AUTHOR(s)<br>Carol J. Feltz<br>Richard L. Dykstra                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                      | 6. PERFORMING ORG. REPORT NUMBER<br>104                       |
| 8. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Department of Statistics and Actuarial Science<br>The University of Iowa<br>Iowa City, Iowa 52242                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                      | 9. CONTRACT OR GRANT NUMBER(s)<br>N00014-83-K-0249            |
| 10. CONTROLLING OFFICE NAME AND ADDRESS<br>Office of Naval Research<br>800 N. Quincy<br>Arlington, VA 22217 -                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                      | 11. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS   |
| 12. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                      | 13. <del>MARCH</del><br>August 1984                           |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                      | 14. NUMBER OF PAGES<br>25                                     |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                      | 15. SECURITY CLASS. (of this report)<br>Unclassified          |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                      | 16. DECLASSIFICATION/DOWNGRADING SCHEDULE                     |
| 17. DISTRIBUTION STATEMENT (of this Report)<br><br>APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                      |                                                               |
| 18. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                      |                                                               |
| 19. SUPPLEMENTARY NOTES                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                      |                                                               |
| 20. KEY WORDS (Continue on reverse side if necessary and identify by block number)<br>Maximum likelihood estimation; survival functions; linear stochastic ordering; censored data; Kaplan-Meier product limit estimator; order restrictions; Kuhn-Tucker vectors.                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                      |                                                               |
| 21. ABSTRACT (Continue on reverse side if necessary and identify by block number)<br>Very often, populations exist that, logically, should satisfy linear stochastic ordering requirements. For example if a mechanical device is improved through N stages, the corresponding survival functions should be linearly stochastically ordered. Nevertheless, estimates may not reflect this stochastic ordering because of the inherent variability of the observations.<br>Here we characterize the maximum likelihood estimates of the survival functions subject to linear stochastic ordering requirements. We show our estimates may be expressed in terms of the well-known Kaplan-Meier product limit |                                      |                                                               |

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 68 IS OBSOLETE  
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

#20.

estimates. We also give an iterative algorithm which we show must converge to the correct solution that depends only upon solving the pairwise problem.

Finally we consider an example concerning survival times for people with squamous carcinoma in the oropharynx when classified by degree of lymph node deterioration at time of discovery.

ND

ATE

LMED

-84

TIC