The Annals of Statistics
1984, Vol. 12, No. 1, 366–371

A GENERALIZED KAPLAN-MEIER ESTIMATOR[1]

BY JAMES B. ROBERTSON AND V. R. R. UPPULURI

University of California, Santa Barbara and Institute for Energy Analysis

In the theory of competing risks, the nonparametric Kaplan-Meier estimator plays an important role. In this paper, a bivariate nonparametric estimator for the competing risk problem is given, which for the special case of independent causes gives the Kaplan-Meier estimator. This paper also introduces a matrix w, through which dependent models for the competing problem can be studied. These results also indicate the special case on the matrix w for which the bounds to the survival function given by Peterson [4] can be obtained.

1. Introduction. The nonparametric estimation of the survival function from incomplete or censored data was first considered by Kaplan and Meier [3] in 1958. Further ramifications of the two sample problem with censored data were given by Efron (1967). Several authors have remarked (cf. [2]) that the validity of the "independence assumption of the Kaplan-Meier estimator" cannot be checked from the actual data. In this paper we consider the general situation when the time to death and the time to loss (censored part) are not necessarily independent.

In Section 2 we derive a class of nonparametric maximum likelihood estimates (MLE) for the survival function given the censored information. In Section 3 we show that the Kaplan-Meier product limit estimator is a distinguished member of this class. In Section 4 we discuss a general procedure for constructing other ML estimates. Section 5 discusses some special cases and has additional remarks.

The results in this paper provide insights for the Kaplan-Meier estimator and the bounds given by Peterson (1976). It is once again brought out that the "independence" assumption in the use of Kaplan-Meier estimator cannot be checked from the data. The weight matrix w introduced in Section 4 is a possible way to study various dependent models.

2. Maximum likelihood estimators. In this section we shall derive a nonparametric MLE for the survival function. Let D denote the time to death and let L denote the time that a loss occurs. We consider D and L to be nonnegative random variables of which only the first one to occur is observed. Thus an observation consists of a bivariate random vector (T, δ) where $T = \min(D, L)$ is the time of observation and $\delta = 1$ or 2 according as $T = D$ or L indicates the nature of the observation. We suppose that δ is well defined, that

Received October 1982; revised July 1983.

[1] This paper is based on work partially performed under Contract No. DE-AC05-760R00033 between the Department of Energy and Oak Ridge Associated Universities.

AMS 1980 *subject classifications.* Primary, 62G05; secondary, 62N05.

Key words and phrases. Competing risks, Kaplan-Meier estimator, Peterson's bounds on the survival function.

is, that $P[D = L] = 0$. The basic problem is to estimate the survival probabilities $S(d) = P[D > d]$. For the purpose of estimation, we suppose that we are given a random sample of (T, δ) of size n. That is, $(T_1, \delta_1), (T_1, \delta_1), \cdots, (T_n, \delta_n)$ are independent random vectors, each with the same distribution as (T, δ) given above. A MLE $\hat{P}$ will be concentrated only on values (d, l) that give rise to actual observations (T_i, δ_i). The largest observation, $T_{(n)}$, requires special attention since it implies the existence of a death or loss that has not been observed. Let $\mathscr{D} = \{d_1, d_2, \cdots, d_r\}$ be the observed deaths, let $\mathscr{L}\ \{l_1, l_2, \cdots, l_s\}$ be the observed losses, and let $\mathscr{T} = \mathscr{D} \cup \mathscr{L} = \{T_1, T_2, \cdots, T_n\}$ be the set of all observations. Let $\hat{\Omega} = (\mathscr{D} \cup \{T_{(n)}\}) \times (\mathscr{T} \cup \{T_{(n)}\})$. Thus, if $\delta_{(n)} = 1$ (resp. $\delta_{(n)} = 2$), $\hat{\Omega}$ has $r(s + 1)$ (resp. $(r + 1)s$) points. $\hat{\Omega}$ will be the support of the MLE measure $\hat{P}$. This is illustrated for $n = 7$ and deaths at times 1.0, 3.1, 5.4, 12.1 and losses at times 0.8, 2.7, 9.2 in Figure 1.

Any of the individual observations (T_i, δ_i) could be realized on the sample space $\hat{\Omega}$. If $\delta_i = 1$ (resp. $\delta_i = 2$), the observation (T_i, δ_i) corresponds to the set

$$A_i = \{(d_i, l_j): l_j > d_i\} \cup \{(d_i, T_{(n)})\}$$

$$(\text{resp. } A_i = \{(d_j, l_i): d_j > l_i\} \cup \{(T_{(n)}, l_i)\}).$$

In particular $A_{(n)} = \{(T_{(n)}, T_{(n)})\}$. The sets $\{A_1, A_2, \cdots, A_n\}$ form a partition of $\hat{\Omega}$. Let $p_i = \hat{P}(A_i)$. Then the likelihood function

$$L(p_1, p_2, \cdots, p_n) = \ln \hat{P}(\cap_{i=1}^n [(D_i, L_i) \in A_i]) = \ln \textstyle\prod_{i=1}^n p_i$$

is easily seen to be maximized by $p_i = 1/n$ for all i. Thus, we have proved:

THEOREM 1. *A probability measure $\hat{P}$ on $\hat{\Omega}$ is a* MLE *for the distribution P of (D, L) if and only if $\hat{P}(A_i) = 1/n$ for all $i = 1, 2, \cdots, n$.*

We remark that if $\delta_{(n)} = 2$, we are assigning mass to the event $[D = T_{(n)}]$. This

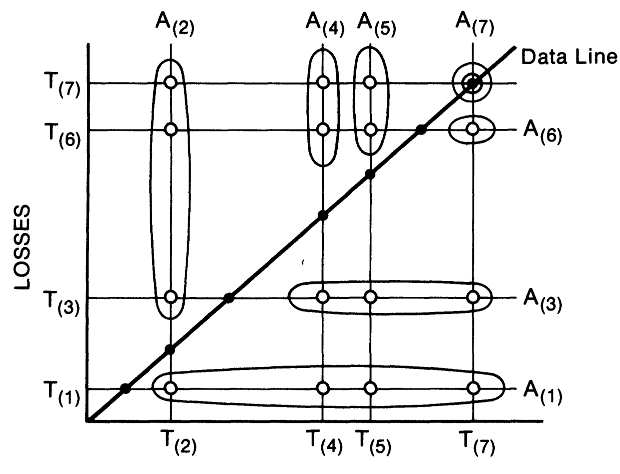

FIG. 1. $\hat{\Omega}$ and the partition.

should be interpreted to say that $D > T_{(n)}$, and the precise value of D cannot be determined (cf. Peterson (1976) for a discussion).

3. Independence. In this section we prove the following:

THEOREM 2. *There is one and only one* MLE *probability* $\hat{P}$ *on* $\hat{\Omega}$ *such that* D *and* L *on* $(\hat{\Omega}, \hat{P})$ *are independent. This is given by the Kaplan-Meier estimator.*

PROOF. Suppose first that $\hat{P}$ is given by

$$\hat{P}(d_i, l_j) = p(d_i)q(l_i). \tag{3.1}$$

If $\delta_i = 1$ (resp. $\delta_i = 2$), we then have

$$\begin{gathered}(1/n) = P(A_i) = p(T_i)[1 - \textstyle\sum_{l_j<T_i} q(l_j)] \\ (\text{resp. } (1/n) = P(A_i) = q(T_i)[1 - \textstyle\sum_{d_j<T_i} p(d_j)]).\end{gathered} \tag{3.2}$$

Let $T_{(1)} < T_{(2)} < \cdots < T_{(n)}$ be the natural ordering of the observations. Proceeding by induction on $i = 1, 2, \cdots, n$, if $\delta_{(i)} = 1$ (resp. $\delta_{(i)} = 2$), we get

$$\begin{gathered}p(T_{(i)}) = \{n[1 - \textstyle\sum_{l_j<T_{(i)}} q(l_j)]\}^{-1} \\ (\text{resp. } q(T_{(i)}) = \{n[1 - \textstyle\sum_{d_j<T_{(i)}} p(d_j)]\}^{-1}).\end{gathered} \tag{3.3}$$

Since at the ith step the right hand side of (3.3) is uniquely determined, the distributions p and q are uniquely determined.

Conversely, if $p(d_i)$ and $q(l_j)$ are defined inductively by (3.3) and if (3.1) holds, then (3.2) implies that $\hat{P}(A_i) = 1/n$ for all i. Thus, $\hat{P}$ is a MLE by Theorem 1.

Kaplan and Meier (1958) derived their estimator under the assumption of independence and so by the uniqueness results, it is given by (3.3) above. □

4. Distribution of mass to the right. We now describe a general procedure for calculating the marginal distribution $\hat{P}(D = d_i)$ of a MLE. This procedure, in the case of independence, was first considered by Efron (1967).

We say that w on $\mathscr{T} \times \mathscr{T}$ is a redistribution to the right (RR) matrix if

(1) $w_{ij} \geq 0$ for all i, j
(2) $\sum_j w_{ij} = 1$, and
(3) $w_{ij} = 0$ if $j < i$ or if $j = i < n$.

w_{ij} represents the proportion of the mass of $T_{(i)}$ that is transferred to $T_{(j)}$.

We will be given an initial distribution m on $\mathscr{T}$, $m_i = m\{T_{(i)}\}$. The expression, "redistribute the mass of $T_{(k)}$, $k < n$ according to w," will signify the linear operation acting on the "vector" m given by

$$\begin{aligned} m_i' &= m_i && \text{if} \quad i < k, \\ m_k' &= 0, \quad \text{and} \\ m_i' &= m_i + m_k w_{ki} && \text{if} \quad i > k. \end{aligned}$$

We shall denote this operation in matrix notation by $m' = m \cdot w_k$. Thus the matrix w_k is obtained by replacing the kth row of the identity matrix by the kth row of the w matrix.

We shall apply the above procedure in two different cases.

Procedure 1. Start initially with m equal to the uniform distribution on $\mathcal{T}$, i.e., $m_i = 1/n$ for all i. Inductively redistribute the mass of each loss to the right according to w. That is

$$m' = m \cdot w_1(\delta_{(1)}) \cdots w_n(\delta_{(n)})$$

where $w_k(\delta_{(k)}) = I$ (the identity matrix) or w_k according as $\delta_{(k)} = 1$ or 2. (In Fig. 1 $m' = mw_1w_3w_6$.) The results of this operation is a probability distribution $\hat{p}(T_{(k)}) = m'_k$ on $\mathcal{D} \cup \{T_{(n)}\}$. (In Fig. 1, $\hat{p}(T_{(2)}) = 1/6$, $\hat{p}(T_{(4)}) = \hat{p}(T_{(5)}) = 5/24$, and $\hat{p}(T_{(7)}) = 5/12$.)

Procedure 1 will be described by saying that $\hat{p}$ is obtained by *redistributing the masses of the losses to the right for the uniform distribution on* $\mathcal{T}$ *according to the* RR *matrix* w.

Procedure 2. For each $T_{(i)}$ let m be defined by $m_j = 0$ if $i \neq j$ and $1/n$ if $i = j$. (This distribution is not a probability.) For simplicity we assume that $\delta_{(i)} = 1$, i.e., that $T_{(i)}$ is a death. If $\delta_{(i)} = 2$ then a similar result will be obtained by interchanging "losses" and "deaths." Now redistribute the masses of the deaths to the right according to w, i.e.,

$$m' = m\bar{w}_1(\delta_{(1)}) \cdots \bar{w}_n(\delta_{(n)})$$

where now $\bar{w}_k(\delta_{(k)}) = w_k$ or I according as $\delta_{(k)} = 1$ or 2. We define $\hat{P}$ on $A_{(i)}$ by $\hat{P}(T_{(i)}, l_j) = m'_j$. (In Fig. 1, with $i = 2$, $\hat{P}(T_{(2)}, T_{(3)}) = 1/35$ and $\hat{P}(T_{(2)}T_{(6)}) = \hat{P}(T_{(2)}, T_{(7)}) = 2/35$.) Carrying out the above procedure for each i results in a probability distribution $\hat{P}$ on Ω such that $\hat{P}(A_{(i)}) = 1/n$ for all i.

The crucial connection between these procedures is given by the following

LEMMA. *Let* w *be an* RR *matrix, then the distribution* $\hat{p}$ *on* $\mathcal{D} \cup \{T_{(n)}\}$ *obtained in Procedure* 1 *is the marginal distribution of* $\hat{P}$ *defined on* $\hat{\Omega}$ *as in Procedure* 2.

PROOF. Define the usual basis vectors e_i by $e_{ij} = 0$ if $j \neq i$ and $= 1$ if $j = i$. Fix an i such that $\delta_{(i)} = 1$. Then the marginal distribution of $\hat{P}$ is given by

$$\sum_j \hat{P}(T_{(i)}, T_{(j)}) = \{\sum_{\delta_{(j)}=2, j<i} e_j\bar{w}_1(\delta_{(1)}) \cdots \bar{w}_n(\delta_{(n)})e_i^T + 1\}/n$$

where e_i^T denotes the transpose of the vector e_i.

If $\delta_{(j)} = 1$, then $e_j\bar{w}_k(\delta_{(k)}) = e_j$ and $e_je_i^T = 0$ if $j \neq i$ and 1 if $j = i$. If $\delta_{(j)} = 2$ and $i > j$, then all the components of $e_j\bar{w}_1(\delta_{(1)}) \cdots \bar{w}_n(\delta_{(n)})$ are 0, so $e_j\bar{w}_1(\delta_{(1)}) \cdots \bar{w}_n(\delta_{(n)})e_i^T = 0$.

Thus

$$\sum_j \hat{P}(T_{(i)}, T_{(j)}) = \sum_j e_j\bar{w}_1(\delta_{(1)}) \cdots \bar{w}_n(\delta_{(n)})e_i^T/n$$
$$= (1/n, \cdots, 1/n)\bar{w}_1(\delta_{(1)}) \cdots \bar{w}_n(\delta_{(n)})e_i^T$$
$$= \hat{p}(T_{(i)}). \quad \square$$

THEOREM 3. *A probability distribution $\hat{p}$ defined on $\mathscr{D} \cup \{T_{(n)}\}$ is a* MLE *if and only if there exists an* RR *matrix w such that $\hat{p}$ is obtained by redistributing the masses of the losses to the right for the uniform distribution on $\mathscr{T}$ according to w.*

PROOF. Suppose that RR matrix w is given. Then $\hat{p}$ obtained by Procedure 2 above is a MLE for P by Theorem 1 since by constriction $\hat{P}(A_i) = 1/n$ for all $i \in \mathscr{T}$.

Conversely let $\hat{p}$ be a MLE. By Theorem 1 there exists a MLE $\hat{P}$ on $\hat{\Omega}$ for P such that $\hat{p}$ is the marginal distribution of $\hat{P}$. Lt i be given such that $\delta_i = 1$. Then $A_i = \{(T_i, l_j): l_j > T_i\}$ and $\sum_{A_i} \hat{P}(T_i, l_j) = \hat{P}(A_i) = 1/n$. Let m be concentrated on $\{T_{(i)}\}$ with $m\{T_{(i)}\} = 1/n$. Define $w_{ij} = n \cdot \hat{P}(T_{(i)}, T_{(j)})$ if $i \le j$ and $T_{(j)} \in \mathscr{L} \cup \{T_{(n)}\}$ and zero otherwise. If $\delta_i = 2$, then replace $\mathscr{L}$ by $\mathscr{D}$ in the above sentence. It is clear that w so defined is an RR matrix and that an application of Procedure 2 above will produce the probability measure. $\hat{P}$. $\square$

We remark that for a given MLE $\hat{p}$ there may be many different MLE $\hat{P}$ on Ω with this marginal, and for every MLE $\hat{P}$ on Ω there may be many different RR matrices w for which Procedure 2 will yield $\hat{P}$.

5. Special cases and remarks.

(a) If $w_{ij} = w_{ik}$ for all j, $k > j$, then we are redistributing the mass "equally" to the right. This is the kind of procedure considered by Efron (1967) and yields the Kaplan-Meier estimator.

(b) If $w_{i,i+1} = 1$ (resp. $w_{in} = 1$), then we get the largest (resp. smallest) MLE for the survival function $S(t)$. These bounds have been considered by Peterson (1976).

(c) *Maximum entropy.* It is frequently useful to ask for distributions that maximize entropy (cf. e.g. Rao (1973), page 162–163, page 172–173, page 217). This technique can be used in our situation to select a specific ML estimator from the class of all such estimates. Subject to the constraint that $\hat{P}(A_i) = 1/n$ the entropy of $\hat{P}$ is maximized by assigning each of the points within A_i equal probability. This corresponds to an RR matrix which redistributed the mass of each loss (death) uniformly on the deaths (losses) to the right. This procedure will put less probability on the longer lifetimes than the Kaplan-Meier estimate.

Acknowledgment. The authors would like to express their sincere thanks to Peter Groer for his comments.

REFERENCES

EFRON, B. (1967). The two sample problem with censored data. *Proc. 5th Berkeley Symp. Math. Statist. and Probab.* **4** 831–853.

GROER, P. G. (1978). Introduction. *Environmental International* **1** 289–290.

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

PETERSON, A. V. (1976). Bounds for a joint distribution function with fixed sub-distribution functions: applications to competing risks. *Proc. Natl. Acad. Sci. USA* **73** 11–13.

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

RAO, C. R. (1973). *Linear Statistical Inference and its Applications*, 2nd edition. Wiley, New York.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
SANTA BARBARA, CALIFORNIA 93106

MATHEMATICS AND STATISTICS
RESEARCH DEPARTMENT
COMPUTER SCIENCES DIVISION
P.O. BOX Y
OAK RIDGE, TENNESSEE 37830