

## NONPARAMETRIC ESTIMATION OF PARTIAL TRANSITION PROBABILITIES IN MULTIPLE DECREMENT MODELS<sup>1</sup>

BY ODD AALEN

*University of Oslo and University of California, Berkeley*

Nonparametric estimators are proposed for transition probabilities in partial Markov chains relative to multiple decrement models. The estimators are generalizations of the product limit estimator. We study the bias of the estimators, prove a strong consistency result and derive asymptotic normality of the estimators considered as stochastic processes. We also compute their efficiency relative to the maximum likelihood estimators in the case of constant forces of transition.

**1. Introduction.** The multiple decrement, or competing risks, model is an old tool in actuarial science, demography and medical statistics. Nelson (1969), Altshuler (1970), Hoel (1972), Peterson (1975) and Aalen (1976) have studied empirical, or nonparametric, statistical analyses for such models. The methods applied are related to the product-limit estimator of Kaplan and Meier (1958). The present paper is a continuation of Aalen (1976). We give nonparametric estimators of *general partial transition probabilities*. For theoretical study of these quantities, see e.g., Hoem (1969). Peterson (1975) independently of us suggests the same kind of estimator. Our theoretical results are, however, not contained in his paper.

Note that the assumptions made in our competing risks model correspond to what is often termed "independence" of risks.

Formally a multiple decrement model may be described as a time-continuous Markov chain with one transient state labeled 0 and  $m$  absorbing states numbered from 1 to  $m$ . We define  $P_i(t)$ ;  $i = 0, 1, \dots, m$ ; to be the probability that the process is in state  $i$  at time  $t$  given that it started in state 0 at time 0. The force of transition (see e.g., Hoem (1969)) or infinitesimal transition probability from state 0 to state  $i$  at time  $t$  is given by

$$\alpha_i(t) = P_i'(t)/P_0(t) \quad i = 1, \dots, m$$

provided the derivative exists. We make the following assumption (cf. Feller (1957), Section XVII.9):

ASSUMPTION.  $\alpha_i(t)$  exists and is continuous everywhere for  $i = 1, \dots, m$ .

---

Received August 1975; revised April 1977.

<sup>1</sup> This paper was prepared with the partial support of National Science Foundation Grant GP 38485.

AMS 1970 subject classifications. Primary 62G05, 62M05; Secondary 62E20, 62G20.

Key words and phrases. Multiple decrement model, competing risks, life table, partial Markov chain, partial transition probability, Markov chain estimation, product limit estimator, non-parametric theory.

The word "nonparametric" will in this paper mean that no further assumption is made about the  $\alpha_i(t)$ .

By a partial chain we mean the model that occurs if we put  $\alpha_j(t) \equiv 0$  for all  $j \notin A$  where  $A$  is some subset of  $\{1, \dots, m\}$ . We want to estimate the transition probabilities in the partial chain. The above mentioned papers study such estimation in the case that  $A$  contains only one state.

We introduce some notation. The total forces of transition to the set of states  $\{1, \dots, m\}$  and to a subset  $A$  of states are given by

$$\delta(t) = \sum_{i=1}^m \alpha_i(t) \quad \text{and} \quad \delta_A(t) = \sum_{i \in A} \alpha_i(t)$$

respectively. The cumulative forces of transition are given by

$$\beta_A(t) = \int_0^t \delta_A(s) ds, \quad \beta_i(t) = \int_0^t \alpha_i(s) ds.$$

Let  $q_A(t)$  be the probability, relative to the partial model, of not leaving state 0 in the time interval  $[0, t]$ . We have  $q_A(t) = \exp(-\beta_A(t))$ . Let  $p_A(t) = 1 - q_A(t)$  and  $p_i(t) = p_{(i)}(t)$ . Finally, define  $p(t) = P_0(t)$  and  $r(t) = p(t)^{-1}$ .

The probability of transition  $0 \rightarrow i$  in the time interval  $[0, t]$  in the partial chain corresponding to  $A$  is given by

$$P_i(t, A) = \int_0^t \alpha_i(s) \exp(-\int_0^s \delta_A(u) du) ds.$$

(Of course we must have  $i \in A$ .)

We will make the following assumption about the experiment and the observation: we observe continuously, over the time interval  $[0, 1]$ ,  $n$  independent processes of the kind described above, each with the same set of forces of transition. Every process is assumed to be in state 0 at time 0.

Use the following notation:  $N_i(t)$  is the number of processes in state  $i$  at time  $t$ . We define the  $N_i(t)$  to be right-continuous for  $i > 0$ . Let  $N_A(t) = \sum_{i \in A} N_i(t)$ , and let  $M(t) = N_0(t)$  and define this to be a left-continuous process. Define:

$$\begin{aligned} R(t) &= M(t)^{-1} & \text{if } M(t) > 0, \\ &= 0 & \text{if } M(t) = 0. \end{aligned}$$

As usual "a.s." denotes "almost surely", while  $X_n \rightarrow_p X$  denotes convergence in probability. Let  $L(X)$  denote the distribution of  $X$ , and let  $I(B)$  denote the indicator function of the set  $B$ .

If we want to stress the dependence on  $n$  we will write  $M_n(t)$ ,  $N_{i,n}(t)$  and similarly for the other quantities.

**2. Estimation.** Write:

$$P_i(t, A) = \int_0^t q_A(s) d\beta_i(s).$$

We will estimate this quantity by substituting estimators for the functions  $q_A(s)$  and  $\beta_i(s)$ .

If we think of the set  $A$  as one single state, then we can use Kaplan and Meier's (1958) product limit estimator for estimating  $q_A(t)$ . It can be written

in the following form:

$$(2.1) \quad \hat{q}_A(t) = \exp \int_0^t \log(1 - R(s)) dN_A(s).$$

This integral, and the similar integrals below, are to be taken as Lebesgue-Stieltjes integrals over the time interval  $[0, t]$ .

For  $\beta_i(t)$  we can use the closely related estimator studied by Nelson (1969) and Aalen (1976):

$$(2.2) \quad \hat{\beta}_i(t) = \int_0^t R(s) dN_i(s).$$

We suggest the following estimator for  $P_i(t, A)$ :

$$(2.3) \quad \hat{P}_i(t, A) = \int_0^t \hat{q}_A(s - 0) d\hat{\beta}_i(s).$$

Alternatively we can write the estimator in the form:

$$(2.4) \quad \hat{P}_i(t, A) = \int_0^t \hat{q}_A(s - 0) R(s) dN_i(s).$$

Clearly, one might suggest other versions of this estimator. Instead of substituting  $\hat{q}_A(t)$  for  $q_A(t)$  one could use  $\exp[-\int_0^t R(u) dN_A(u)]$  while  $\beta_i(t)$  might be estimated by  $-\int_0^t \log[1 - R(u)] dN_i(u)$ . (This was suggested by a referee.) When we prefer the estimator (2.3) this has the following reason: firstly, it may be shown that  $\hat{P}_i(t, \{i\})$  coincides with Kaplan and Meier's estimator of  $p_i(t) = P_i(t, \{i\})$ . Clearly, this ought to be the case since our intention is to generalize that estimator. Secondly, consider the case  $A = \{1, \dots, m\}$ . Then  $P_i(t, A) = P_i(t)$ , and hence it is reasonable to require that the estimator  $\hat{P}_i(t, A)$  specializes to  $(1/n)N_i(t)$ . That this is indeed the case may be shown with some computation.

By results of Kaplan and Meier (1958), Breslow and Crowley (1974), Meier (1975) and Aalen (1976)  $\hat{q}_A(t)$  and  $\hat{\beta}_i(t)$  are known to have nice properties, and it is reasonable to assume that these carry over to  $\hat{P}_i(t, A)$ . In this paper we will mainly concentrate on large sample properties, but first we will give the following results:

**PROPOSITION 1.**  $\hat{P}_i(t, A)$  is based on minimal sufficient and complete statistics.

This proposition is an immediate consequence of Theorem 3.1 of Aalen (1976).

**PROPOSITION 2.** Let  $\hat{p}_i(t) = \hat{P}_i(t, \{i\})$ . The following holds:

- (i)  $0 \leq p_i(t) - E\hat{p}_i(t) \leq (1 - p(t))^n \beta_i(t)$ ;
- (ii)  $|E\hat{P}_i(t, A) - P_i(t, A)| \leq (1 - p(t))^n \beta_i(t)(1 + \beta_A(t))$ .

Note that according to this proposition the bias of  $\hat{p}_i(t)$  and  $\hat{P}_i(t, A)$  converges exponentially to 0 when  $n \rightarrow \infty$ . Moreover, Proposition 1 implies that these estimators are uniformly minimum variance estimators for their expectations. These facts coupled together indicate that the estimators should be reasonable candidates.

In the case  $A = \{1, \dots, m\}$  we have  $\hat{P}_i(t, A) = (1/n)N_i(t)$  and hence it is of course unbiased. By Proposition 1 we have in this case a uniformly minimum variance unbiased estimator.

Proposition 2 is easily proved by the technique used in the proof of Theorem 4.1 in Aalen (1976). For completeness the proof will be given in the Appendix. The same method of proof may be used to derive approximate expressions of variances and covariances. However, in this paper we will only give the variances and covariances of the limiting distribution.

**3. Consistency.** We will prove the following consistency result. (All limits below are taken with respect to  $n$ .)

**THEOREM 1.** *When  $n \rightarrow \infty$  the following holds:*

$$\sup_{0 \leq t \leq 1} \frac{n^{\frac{1}{2}}}{\log n} |\hat{P}_{i,n}(t, A) - P_i(t, A)| \rightarrow 0 \text{ a.s.}$$

For the proof we need some intermediate results. The first one is a part of Lemma 2.2 of Barlow and van Zwet (1970).

**LEMMA 1.** *Let  $X_1, X_2, \dots, X_n$  be independent and identically distributed with continuous distribution function  $F(x)$ . Let  $F_n(x)$  be the empirical distribution function. Then*

$$\frac{n^{\frac{1}{2}}}{\log n} \sup_x |F_n(x) - F(x)| \rightarrow 0 \text{ a.s.}$$

We next state Lemma 1 of Breslow and Crowley (1974).

**LEMMA 2.** *Let  $\hat{\beta}_A(t) = \sum_{i \in A} \hat{\beta}_i(t)$ . If  $M(1) > 0$ , then*

$$0 < -\log \hat{q}_A(t) - \hat{\beta}_A(t) < \frac{n - M(1)}{nM(1)}.$$

The first part of the next proposition is a strengthening of Theorem 6.1 in Aalen (1976). The second part gives a strong consistency result for the product limit estimator.

**PROPOSITION 3.** *The following limits hold when  $n \rightarrow \infty$ :*

- (i)  $\sup_{0 \leq t \leq 1} (n^{\frac{1}{2}}/\log n) |\hat{\beta}_{i,n}(t) - \beta_i(t)| \rightarrow 0 \text{ a.s.,}$
- (ii)  $\sup_{0 \leq t \leq 1} (n^{\frac{1}{2}}/\log n) |\hat{q}_{A,n}(t) - q_A(t)| \rightarrow 0 \text{ a.s.}$

**PROOF.** Put  $c_n = n^{\frac{1}{2}}/\log n$ . The supremas below are all taken over the set  $0 \leq t \leq 1$ . We have

$$\hat{\beta}_{i,n}(t) - \beta_i(t) = \int_0^t (nR_n(s) - r(s)) d\left(\frac{1}{n} N_{i,n}(s)\right) + \int_0^t r(s) d\left(\frac{1}{n} N_{i,n}(s) - P_i(s)\right).$$

Hence:

$$\sup c_n |\hat{\beta}_{i,n}(t) - \beta_i(t)| \leq X_n + Y_n$$

where

$$X_n = \sup c_n |nR_n(t) - r(t)|$$

$$Y_n = \sup \left| \int_0^t r(s) d\left[ c_n \left( \frac{1}{n} N_{i,n}(s) - P_i(s) \right) \right] \right|.$$

It is enough to show that  $X_n \rightarrow 0$  and  $Y_n \rightarrow 0$  a.s. We have:

$$X_n \leq \frac{\sup c_n \left| \frac{1}{n} M_n(t) - p(t) \right|}{\frac{1}{n} M_n(1)p(1)} I(M_n(1) > 0) + 2n^2 r(1) I(M_n(1) = 0).$$

The first part goes to 0 a.s. by Lemma 1. The second part goes to 0 a.s. by the Borel–Cantelli lemma and the fact that  $M_n(1)$  is binomial  $(n, p(1))$ . Partial integration gives us for  $Y_n$ :

$$\begin{aligned} Y_n &= \sup \left| r(t)c_n \left( \frac{1}{n} N_{i,n}(t) - P_i(t) \right) - \int_0^t c_n \left( \frac{1}{n} N_{i,n}(s) - P_i(s) \right) dr(s) \right| \\ &\leq 2r(1) \sup c_n \left| \frac{1}{n} N_{i,n}(t) - P_i(t) \right|. \end{aligned}$$

Hence, by Lemma 1, also  $Y_n \rightarrow 0$  a.s. This proves (i). For the proof of (ii) we first use (i) and Lemma 2 to establish the following:

$$c_n I(M_n(1) > 0) \sup_t |\log \hat{q}_{A,n}(t) - \log q_A(t)| \rightarrow 0 \text{ a.s.}$$

By applying the mean value theorem we get:

$$c_n I(M_n(1) > 0) \sup_t |\hat{q}_{A,n}(t) - q_A(t)| \rightarrow 0 \text{ a.s.}$$

Using the fact that  $M_n(1)$  is binomially distributed and applying the Borel–Cantelli lemma we get:

$$c_n I(M_n(1) = 0) \sup_t |\hat{q}_{A,n}(t) - q_A(t)| \rightarrow 0 \text{ a.s.} \quad \square$$

PROOF OF THEOREM 1. We can write:

$$\begin{aligned} \hat{P}_{i,n}(t, A) - P_i(t, A) &= \int_0^t \hat{q}_{A,n}(s - 0) d\hat{\beta}_{i,n}(s) - \int_0^t q_A(s) d\beta_i(s) \\ &= \int_0^t (\hat{q}_{A,n}(s - 0) - q_A(s)) d\hat{\beta}_{i,n}(s) + \int_0^t q_A(s) d(\hat{\beta}_{i,n}(s) - \beta_i(s)). \end{aligned}$$

By treating separately the two integrals of the last expression Theorem 1 may be proved in the same way as part (i) of Proposition 3.  $\square$

**4. Weak convergence.** In this section we will study convergence in distribution of the stochastic processes  $\hat{P}_i(t, A)$ . Throughout the section  $t$  will be limited to the interval  $[0, 1]$ . Let  $D$  be the function space considered in Section 14 of Billingsley (1968) and let  $\rho$  be the metric  $d_0$  defined there. In this section the term “weak convergence” will be used with respect to the product metric  $\rho_k$  on the product space  $D^k$  for appropriate values of  $k$ . Let  $C$  be the subset of  $D$  consisting of all continuous functions on  $[0, 1]$ , and let  $\lambda$  be the usual uniform metric on this space. Let  $\lambda_k$  be the product metric on the product space  $C^k$ . It is well known that  $\lambda_k$  and  $\rho_k$  coincide on  $C^k$ . It is also known that if  $x_n \in D^k$  and  $x \in C^k$ , then  $x_n \rightarrow x$  in the  $\lambda_k$ -metric if and only if  $x_n \rightarrow x$  in the  $\rho_k$ -metric. For both these facts see e.g., Billingsley (1968, page 112). They will be repeatedly

used below without further mentioning. A consequence of the last mentioned fact is that if  $Z_n$  and  $Z$  are random elements of  $D$  and  $C$  respectively, then  $\rho_k(Z_n, Z) \rightarrow_p 0$  if and only if  $\lambda_k(Z_n, Z) \rightarrow_p 0$ .

Let  $X_1, \dots, X_m$  be independent Gaussian processes on the time interval  $[0, 1]$ , each with independent increments and expectation 0. Let the variances be given by  $\text{Var } X_i(t) = \int_0^t \alpha_i(s)r(s) ds$ . We choose versions of  $X_1, \dots, X_m$  with continuous sample paths.

**THEOREM 2.** *The vector consisting of all processes of the form  $Y_{i,n}(\cdot, A) = n^{\frac{1}{2}}(\hat{P}_{i,n}(\cdot, A) - P_i(\cdot, A))$  for  $i \in A$  and  $A \subset \{1, \dots, m\}$  converges weakly to the vector consisting of the Gaussian processes  $Y_i(\cdot, A)$  defined by:*

$$-Y_i(t, A) = \int_0^t [\int_s^t q_A(u)\alpha_i(u) du - q_A(s)] dX_i(s) + \sum_{j \in A - \{i\}} \int_0^t \int_s^t q_A(u)\alpha_i(u) du dX_j(s),$$

where the integrals are stochastic integrals in quadratic mean.

**REMARK.** Stochastic integrals in quadratic mean are defined e.g., in Cramér and Leadbetter (1967, Section 5.3) where their properties are discussed. One should note that the representation of the  $Y$ -processes as stochastic integrals over the  $X$ -processes makes it very simple to compute moments. For instance we have:

$$-EY_i(t, A) = \int_0^t [\int_s^t q_A(u)\alpha_i(u) du - q_A(s)] dEX_i(s) + \sum_{j \in A - \{i\}} \int_0^t \int_s^t q_A(u)\alpha_i(u) du dEX_j(s) = 0.$$

Since the  $X_i$  are independent processes and have independent increments, we have:

$$\begin{aligned} \text{Var } Y_i(t, A) &= \int_0^t [\int_s^t q_A(u)\alpha_i(u) du - q_A(s)]^2 d(\text{Var } X_i(s)) \\ &\quad + \sum_{j \in A - \{i\}} \int_0^t [\int_s^t q_A(u)\alpha_i(u) du]^2 d(\text{Var } X_j(s)) \\ &= \int_0^t [\int_s^t q_A(u)\alpha_i(u) du - q_A(s)]^2 \alpha_i(s)r(s) ds \\ &\quad + \sum_{j \in A - \{i\}} \int_0^t [\int_s^t q_A(u)\alpha_i(u) du]^2 \alpha_j(s)r(s) ds. \end{aligned}$$

It is clear how covariances between pairs of  $Y_i(t, A)$  for different  $i, t$  and  $A$  can be computed in the same easy way. The expressions will not be given here, we will just note that if  $i \in A$  and  $j \in B$ , then:

$$A \cap B = \emptyset \implies Y_i(\cdot, A) \text{ is independent of } Y_j(\cdot, B).$$

**PROOF.** The proof will follow the Pyke and Shorack (1968) approach. Define  $X_{i,n}(t) = n^{\frac{1}{2}}(\hat{\beta}_{i,n}(t) - \beta_i(t))$  and  $Z_{A,n}(t) = n^{\frac{1}{2}}(\hat{q}_{A,n}(t) - q_A(t))$ .

We can write:

$$(4.1) \quad Y_{i,n}(t, A) = \int_0^t Z_{A,n}(s) d\beta_i(s) + \int_0^t q_A(s) dX_{i,n}(s) + n^{-\frac{1}{2}} \int_0^t Z_{A,n}(s) dX_{i,n}(s).$$

(These integrals and the integrals below are still defined as Lebesgue–Stieltjes or Lebesgue integrals until otherwise stated.)

For proving the theorem we will use the following proposition which is proved as Theorem 8.2 in Aalen (1976). Put  $\mathbf{X} = [X_1, \dots, X_m]$  and  $\mathbf{X}(n) = [X_{1,n}, \dots, X_{m,n}]$ .

PROPOSITION 4.  $\mathbf{X}(n)$  converges weakly to  $\mathbf{X}$ .

To exploit this result we must express the  $Y_{i,n}(\cdot, A)$  in terms of the  $X_{j,n}$ . We will write  $U_n = o_p(1)$  for random elements  $U_n$  of  $D^k$  when  $\rho_k(U_n, 0)$  (or equivalently  $\lambda_k(U_n, 0)$ ) converges in probability to 0.

By the arguments used in the proof of Proposition 3, part (ii) we have:

$$\begin{aligned} Z_{A,n} &= -q_A n^{\frac{1}{2}}(\hat{\beta}_{A,n} - \beta_A) + o_p(1) \\ &= -q_A \sum_{j \in A} X_{j,n} + o_p(1). \end{aligned}$$

Using this (4.1) can be rewritten in the following way:

$$(4.2) \quad \begin{aligned} Y_{i,n}(t, A) &= -\int_0^t q_A(s) \sum_{j \in A} X_{j,n}(s) \alpha_i(s) ds \\ &\quad + \int_0^t q_A(s) dX_{i,n}(s) - S_{i,n}(t, A) + o_p(1) \end{aligned}$$

where

$$S_{i,n}(t, A) = n^{-\frac{1}{2}} \sum_{j \in A} \int_0^t q_A(s) X_{j,n}(s) dX_{i,n}(s).$$

We now want to prove that  $S_{i,n}(\cdot, A) = o_p(1)$ . Here we follow Pyke and Shorack (1968) in using item 3.1.1 in Skorohod (1956).  $D$  is complete and separable with metric  $\rho$  and the same holds therefore for  $D^k$  with metric  $\rho_k$ . Thus the mentioned result of Skorohod ensures us that there exists a probability space representation of the processes

$$\mathbf{X} = [X_1, \dots, X_m] \quad \text{and} \quad \mathbf{X}(n) = [X_{1,n}, \dots, X_{m,n}]$$

such that  $\rho_k(\mathbf{X}(n), \mathbf{X}) \rightarrow 0$  a.s., and we use this representation in the following.

All supremas below are taken over  $t \in [0, 1]$ . We have:

$$\begin{aligned} \sup |S_{i,n}(t, A)| &\leq \sum_{j \in A} n^{-\frac{1}{2}} \sup \left| \int_0^t q_A(s) X_{j,n}(s) dX_{i,n}(s) \right| \\ &\leq \sum_{j \in A} n^{-\frac{1}{2}} \sup \left| \int_0^t q_A(s) (X_{j,n}(s) - X_j(s)) dX_{i,n}(s) \right| \\ &\quad + \sum_{j \in A} n^{-\frac{1}{2}} \sup \left| \int_0^t q_A(s) X_j(s) dX_{i,n}(s) \right| \\ &\leq \sum_{j \in A} \sup |X_{j,n}(t) - X_j(t)| (\hat{\beta}_{i,n}(1) + \beta_i(1)) \\ &\quad + \sum_{j \in A} \sup \left| \int_0^t q_A(s) X_j(s) d(\hat{\beta}_{i,n}(s) - \beta_i(s)) \right|. \end{aligned}$$

The first term in the last expression converges in probability to 0. For the second term we use the method described in the proof of Theorem 4 in Breslow and Crowley (1974). Consider a subset  $\Omega_0$  of the underlying probability space such that  $P(\Omega_0) = 1$ , and such that for  $\omega \in \Omega_0$  all  $X_j$  are uniformly continuous on  $[0, 1]$ , and  $\rho(\hat{\beta}_{i,n}, \beta_i)$  converges to 0. Choose a partition (depending on  $\omega$ ) of  $[0, 1]$  into  $K$  intervals  $I_k = (\xi_{k-1}, \xi_k]$  such that

$$\sup_j \sup_{t \in I_k} |X_j(t)q_A(t) - X_j(\xi_k)q_A(\xi_k)| < \varepsilon$$

for  $k = 1, \dots, K$ . Then the second term above is bounded by

$$\varepsilon(\hat{\beta}_{i,n}(1) + \beta_i(1)) + 2K\lambda(X_j q_A, 0)\lambda(\hat{\beta}_{i,n}, \beta_i),$$

which tends to  $2\beta_i(1)\varepsilon$  when  $n \rightarrow \infty$ . Since  $\varepsilon$  is arbitrary, this shows that the second term also converges in probability to 0. Hence by (4.2) we can write:

$$Y_{i,n}(t, A) = -\sum_{j \in A} \int_0^t q_A(s) X_{j,n}(s) \alpha_i(s) ds + \int_0^t q_A(s) dX_{i,n}(s) + o_p(1).$$

By partial integration rewrite the second term above and get:

$$Y_{i,n}(t, A) = - \sum_{j \in A} \int_0^t q_A(s) X_{j,n}(s) \alpha_i(s) ds + q_A(t) X_{i,n}(t) - \int_0^t X_{i,n}(s) d[q_A(s)] + o_p(1).$$

We are now ready to use Proposition 4. Let  $Y(n)$  be the vector consisting of all  $Y_{i,n}(\cdot, A)$  in some order. It is easily figured out that the vector has  $l = m2^{m-1}$  components. If we regard  $Y(n)$  as a function of  $X(n)$ , then we have a mapping from  $D^m$  to  $D^l$ . If we use the metrics  $\lambda_m$  and  $\lambda_l$  on these spaces then the mapping is obviously continuous. This is also the case for the metrics  $\rho_n$  and  $\rho_k$  if the function is restricted to  $C^m$ . Since now  $X$  is a.s. an element of  $C^m$ , it follows from Theorem 5.1 of Billingsley (1968) and from Proposition 4 above that  $Y(n)$  converges weakly to the vector consisting of the components:

$$U_i(t, A) = - \sum_{j \in A} \int_0^t q_A(s) X_j(s) \alpha_i(s) ds + q_A(t) X_i(t) - \int_0^t X_i(s) d[q_A(s)].$$

So far all our integrals have been Lebesgue–Stieltjes or Lebesgue integrals. However, as shown on page 90 in Cramér and Leadbetter (1967), the integrals in the last expression can alternatively be taken as stochastic integrals in quadratic mean without changing their value. By using the partial integration formula 5.3.7 in Cramér and Leadbetter (1967) we conclude that  $U_i(t, A) = Y_i(t, A)$ . □

**5. Estimation of the asymptotic variance.** We will suggest an estimator of  $\text{Var } Y_i(t, A)$ , which is given in the remark after Theorem 2. We first rewrite the expression for the variance in the following form:

$$\text{Var } Y_i(t, A) = \int_0^t [P_i(t, A) - P_i(s, A) - q_A(s)]^2 r(s) d\beta_i(s) + \sum_{j \in A - \{i\}} \int_0^t [P_i(t, A) - P_i(s, A)]^2 r(s) d\beta_j(s).$$

Estimators of  $q_A(t)$ ,  $\beta_j(t)$  and  $P_i(t, A)$  are given by (2.1), (2.2) and (2.3). Since  $M(t)$  is binomial  $(n, r(t)^{-1})$ , it is reasonable to estimate  $r(t)$  by  $nR(t)$ . Hence, by the same principle as was used for estimating  $P_i(t, A)$ , we assert that the following is a reasonable estimator of  $\text{Var } Y_i(t, A)$ .

$$n \int_0^t [\hat{P}_i(t, A) - \hat{P}_i(s, A) - \hat{q}_A(s)]^2 R(s) d\hat{\beta}_i(s) + \sum_{j \in A - \{i\}} n \int_0^t [\hat{P}_i(t, A) - \hat{P}_i(s, A)]^2 R(s) d\hat{\beta}_j(s).$$

Alternatively we can write:

$$n \int_0^t [\hat{P}_i(t, A) - \hat{P}_i(s, A) - \hat{q}_A(s)]^2 R^2(s) dN_i(s) + \sum_{j \in A - \{i\}} n \int_0^t [\hat{P}_i(t, A) - \hat{P}_i(s, A)]^2 R^2(s) dN_j(s).$$

By the same kind of arguments as in Section 3 it is easily shown that this estimator converges almost surely to  $\text{Var } Y_i(t, A)$ , uniformly in  $t$ .

Of course, relevant covariance functions of different kinds can be estimated in a similar way.

**6. Asymptotic relative efficiency.** In this section we will assume that all forces of transition are constant on the time interval  $[0, t]$ , i.e.,

$$\alpha_i(s) = \alpha_i \quad \forall s \in [0, t], \quad i = 1, \dots, m, \quad 0 \leq t \leq 1,$$



where the  $\alpha_i$  are positive numbers. Let  $\alpha_{i,t}^*$  be the maximum likelihood estimator of  $\alpha_i$  based on complete observation over the time interval  $[0, t]$ . The  $\alpha_{i,t}^*$  are the so-called "occurrence/exposure" rates for the  $\alpha_i$ . They are given by:

$$\alpha_{i,t}^* = N_i(t) / \int_0^t M(s) ds,$$

(see e.g., Hoem (1971)).

Let  $\hat{\delta} = \sum_{j=1}^m \alpha_j$ ,  $\hat{\delta}_A = \sum_{j \in A} \alpha_j$ ,  $\hat{\delta}_i^* = \sum_{j=1}^m \alpha_{j,t}^*$ ,  $\hat{\delta}_{A,t}^* = \sum_{j \in A} \alpha_{j,t}^*$ . The maximum likelihood estimators of the  $P_i(t, A)$  are given by:

$$P_i^*(t, A) = \frac{\alpha_{i,t}^*}{\hat{\delta}_{A,t}^*} (1 - e^{-\hat{\delta}_{A,t}^* t}).$$

In this section we will study the asymptotic efficiency of  $\hat{P}_i(t, A)$  relative to  $P_i^*(t, A)$  in the sense that we will compare the variances of the asymptotic distributions. We will denote these by  $asVar \hat{P}_i(t, A)$  and  $asVar P_i^*(t, A)$  respectively. We regard  $i$  and  $A$  as fixed and introduce parameters  $a$  and  $b$  by  $\delta_A = a\hat{\delta}$  and  $\alpha_i = b\hat{\delta}_A$ . One notes that  $0 \leq a \leq 1$  and  $0 \leq b \leq 1$ .

By applying a Taylor series development and using results in Hoem (1971)

TABLE 1  
Asymptotic efficiency

		$\delta t = .5$				
		$a = .20$	$a = .40$	$a = .60$	$a = .80$	$a = 1.00$
$b = .25$		.985	.990	.993	.996	.997
$b = .50$		.983	.987	.989	.991	.992
$b = .75$		.981	.983	.985	.986	.987
$b = 1.00$		.979	.979	.979	.979	.979
		$\delta t = 1.0$				
		$a = .20$	$a = .40$	$a = .60$	$a = .80$	$a = 1.00$
$b = .25$		.942	.961	.976	.987	.991
$b = .50$		.936	.950	.963	.973	.979
$b = .75$		.929	.937	.945	.953	.958
$b = 1.00$		.921	.921	.921	.921	.921
		$\delta t = 2.0$				
		$a = .20$	$a = .40$	$a = .60$	$a = .80$	$a = 1.00$
$b = .25$		.795	.866	.928	.970	.988
$b = .50$		.775	.834	.892	.940	.967
$b = .75$		.752	.789	.835	.882	.920
$b = 1.00$		.724	.724	.724	.724	.724
		$\delta t = 3.0$				
		$a = .20$	$a = .40$	$a = .60$	$a = .80$	$a = 1.00$
$b = .25$		.610	.746	.874	.960	.992
$b = .50$		.580	.697	.827	.928	.976
$b = .75$		.542	.622	.736	.856	.935
$b = 1.00$		.496	.496	.496	.496	.496

we find:

$$\text{asVar } P_i^*(t, A) = (1 - e^{-\delta t})^{-1} a^{-1} b [a^2 b \delta^2 t^2 e^{-2a\delta t} + (1 - b)(1 - e^{-a\delta t})^2].$$

From the variance formula in the remark to Theorem 2 we can compute:

$$\text{asVar } \hat{P}_i(t, A) = (1 - 2a)^{-1} ab [(1 - 2ab)e^{\delta t(1-2a)} - b(1 - 2a)e^{-2a\delta t} - 1 + b].$$

The efficiency is the quotient between  $\text{asVar } P_i^*(t, A)$  and  $\text{asVar } \hat{P}_i(t, A)$ :

$$e(a, b, \delta, t) = \frac{(1 - 2a)[a^2 b \delta^2 t^2 e^{-2a\delta t} + (1 - b)(1 - e^{-a\delta t})^2]}{a^2(1 - e^{-\delta t})[(1 - 2ab)e^{\delta t(1-2a)} - b(1 - 2a)e^{-2a\delta t} - 1 + b]}.$$

Table 1 gives values of this function for some values of  $a, b$  and  $\delta t$ .

It is seen from the expression above that

$$\begin{aligned} f(a, b, \delta) &= \lim_{t \rightarrow \infty} e(a, b, \delta, t) = 0 && a < \frac{1}{2} \quad \text{or} \quad b = 1 \\ &= \frac{2a - 1}{a^2} && a \geq \frac{1}{2} \quad \text{and} \quad b < 1. \end{aligned}$$

One should note the discontinuity of  $f$  at  $b = 1$ . In particular  $f(1, b, \delta) = 1$  for  $b < 1$  while  $f(1, 1, \delta) = 0$ . This complements results given by Sverdrup (1965, page 195) for a single decrement model, where of course  $b = 1$ .

Table 1 together with the asymptotic results in the last paragraph seem to indicate that the relative efficiency of the nonparametric estimators is good when either  $\delta t$  is small or when  $b < 1$  and  $a$  is relatively close to 1.

**Acknowledgment.** This work grew out of a part of my master's thesis (Aalen, 1972) which I wrote under the supervision of Professor Jan M. Hoem. Professor L. Le Cam has been very helpful in answering questions about the weak convergence theory. Also, Professor P. Bickel has made some useful comments. Finally I am very grateful to the referees for correcting some mistakes and pointing out several obscurities.

APPENDIX

PROOF OF PROPOSITION 2. We will first show:

$$(A.1) \quad 0 \leq E\hat{q}_A(t) - q_A(t) \leq \beta_A(t)(1 - p(t))^n.$$

In Section 1  $M(t)$ , and hence  $R(t)$ , was defined to be left-continuous. By using this fact together with (2.1) we can write:

$$\hat{q}_A(t + h) = \hat{q}_A(t)(1 - R(t)I(t, h))(1 - U(t, h))$$

where  $I(t, h)$  is 1 if there is at least one transition to  $A$  in the time interval  $(t, t + h)$  and 0 otherwise, and

$$\Pr(U(t, h) \neq 0) = o(h), \quad \Pr(0 \leq U(t, h) \leq 1) = 1.$$

Hence, if we put  $f(t) = E\hat{q}_A(t)$ , we get

$$f(t + h) = f(t) - E(\hat{q}_A(t)R(t)I(t, h)) + o(h).$$

Now  $\Pr [I(t, h) = 1 | M(t)] = M(t)\delta_A(t)h + o(h)$ , hence:

$$\begin{aligned} f(t+h) &= f(t) - h\delta_A(t)E[\hat{q}_A(t)R(t)M(t)] + o(h); \\ f(t) &= f(t-h) - h\delta_A(t-h)E[\hat{q}_A(t-h)R(t-h)M(t-h)] + o(h). \end{aligned}$$

Hence:  $f'(t) = -\delta_A(t)E[\hat{q}_A(t)R(t)M(t)]$ .

Define  $K$  by  $K(x) = 1$  if  $x = 0$  and  $K(x) = 0$  otherwise. We then get

$$f'(t) = -\delta_A(t)f(t) + \delta_A(t)E[\hat{q}_A(t)K(M(t))].$$

Solving for  $f(t)$  and exploiting the condition  $f(0) = 1$  gives us:

$$f(t) = q_A(t) + \int_0^t E[\hat{q}_A(s)K(M(s))] \exp(-\int_s^t \delta_A(u) du) \delta_A(s) ds.$$

Hence:

$$0 \leq f(t) - q_A(t) \leq EK(M(t))\beta_A(t)$$

which is equivalent to (A.1). By now using a method similar to the one above we can easily show:

$$E\hat{P}_i(t, A) - P_i(t, A) = \int_0^t (E\hat{q}_A(s) - q_A(s))\alpha_i(s) ds - \int_0^t E[\hat{q}_A(s)K(M(s))]\alpha_i(s) ds.$$

Hence by (A.1):

$$\begin{aligned} |E\hat{P}_i(t, A) - P_i(t, A)| &\leq \int_0^t (1 - p(s))^n \beta_A(s) \alpha_i(s) ds + \int_0^t E[\hat{q}_A(s)K(M(s))]\alpha_i(s) ds \\ &\leq (1 - p(t))^n \beta_A(t) \beta_i(t) + E(K(M(t)))\beta_i(t) \\ &= (1 - p(t))^n \beta_i(t)(1 + \beta_A(t)). \end{aligned}$$

This proves Proposition 2.

#### REFERENCES

- [1] AALEN, O. (1972). Estimering av risikorater for prevensjons-midlet "spiralen." (In Norwegian). Graduate thesis in statistics, University of Oslo.
- [2] AALEN, O. (1976). Nonparametric inference in connection with multiple decrement models. *Scand. J. Statist.* 3 15-27.
- [3] ALTSHULER, B. (1970). Theory for the measurement for competing risks in animal experiments. *Math. Biosciences* 6 1-11.
- [4] BARLOW, R. B. and VAN ZWET, W. R. (1970). Asymptotic properties of isotonic estimators for the generalized failure rate function. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.). Cambridge Univ. Press.
- [5] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- [6] BRESLOW, N. and CROWLEY, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.* 2 437-453.
- [7] CHIANG, C. L. (1968). *Introduction to Stochastic Processes in Biostatistics*. Wiley, New York.
- [8] CRAMÉR, H. and LEADBETTER, M. R. (1967). *Stationary and Related Stochastic Processes*. Wiley, New York.
- [9] DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* 27 642-669.
- [10] FELLER, W. (1957). *An Introduction to Probability Theory and its Applications*. I. 2nd Ed. Wiley, New York.
- [11] HOEL, D. G. (1972). A representation of mortality data by competing risks. *Biometrics* 28 475-488.

- [12] HOEM, J. M. (1969). Purged and partial Markov chains. *Skand. Aktuarietidskr.* **52** 147-155.
- [13] HOEM, J. M. (1971). Point estimation of forces of transition in demographic models. *J. Roy. Statist. Soc. Ser. B* **33** 275-289.
- [14] KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457-481.
- [15] MEIER, P. (1975). Estimation of a distribution function from incomplete observations. In *Perspectives in Probability* (J. Gani, ed.) 67-87. Applied Probability Trust.
- [16] NELSON, W. (1969). Hazard plotting for incomplete failure data. *J. Qual. Tech.* **1** 27-52.
- [17] PETERSON, A. V. (1975). Nonparametric estimation in the competing risks problem. Technical report No. 73, Stanford Univ.
- [18] PYKE, R. and SHORACK, G. (1968). Weak convergence of a two sample empirical process and a new approach to the Chernoff-Savage theorems. *Ann. Math. Statist.* **39** 755-771.
- [19] SIMONSEN, W. (1966). *Forsikringsmatematik*. Hefte I. Københavns Universitets Fond til Tilvejebringelse af Laeremidler.
- [20] SKOROHOD, A. V. (1956). Limit theorems for stochastic processes. *Theor. Probability Appl.* **1** 261-290.
- [21] SVERDRUP, E. (1965). Estimates and test procedures in connection with stochastic models for deaths, recoveries and transfers between different states of health. *Skand. Aktuarietidskr.* **48** 184-211.
- [22] ZWINGGI, E. (1945). *Versicherungsmathematik*. Verlag Birkhäuser, Basel.

DEPT. OF MATHEMATICAL AND  
PHYSICAL SCIENCES  
UNIVERSITY OF TROMSØ  
N-9001 TROMSØ  
NORWAY