STATISTICAL RESEARCH REPORT No 6
Institute of Mathematics 1973
University of Oslo

NONPARAMETRIC INFERENCE IN CONNECTION WITH MULTIPLE DECREMENT MODELS

by

Odd Aalen

Summary.

In this paper we define and study certain nonparametric estimators of partial transition probabilities and of integrals of forces of transition in multiple decrement models (which are simple Markov chains). We compute approximately expectations, variances and covariances and prove that the estimators are based on minimal sufficient statistics. We prove that the estimators are strongly consistent and give a generalization of the Glivenko-Cantelli theorem. For some of the estimators we prove asymptotic normality which gives us asymptotic tests. We also give a test of Kolmogorov-Smirnov type. We conclude with a comparison of our estimators with the occurrence-exposure rates. This comparison seems to indicate that even when the forces of transition are constant, the nonparametric estimators are almost equally good as the occurrence-exposure rates.

The theory developed in this paper may alternatively be regarded as a generalization of the theory of empirical distribution functions.

Summary.

In this paper we define and study certain nonparametric estimators of partial transition probabilities and of integrals of forces of transition in multiple decrement models (which are simple Markov chains). We compute approximately expectations, variances and covariances and prove that the estimators are based on minimal sufficient statistics. We prove that the estimators are strongly consistent and give a generalization of the Glivenko-Cantelli theorem. For some of the estimators we prove asymptotic normality which gives us asymptotic tests. We also give a test of Kolmogorov-Smirnov type. We conclude with a comparison of our estimators with the occurrence-exposure rates. This comparison seems to indicate that even when the forces of transition are constant, the nonparametric estimators are almost equally good as the occurrence-exposure rates.

The theory developed in this paper may alternatively be regarded as a generalization of the theory of empirical distribution functions.

Contents:

1. Introduction.

A multiple decrement model (see e.g. Zwinggi (1945), Simonsen (1966), Sverdrup (1967), Hoem (1969)) is a time-continuous Markov chain with one transient state labeled 0 and m absorbing states numbered from 1 to m . We define $P_i(t)$; $i = 0,1,\ldots,m$; 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

$$(1.1) \qquad \alpha_i(t) = P_i'(t)/P_0(t) \qquad i = 1,\ldots,m$$

provided the derivative exists.

We make the following mathematical assumption (see Feller (1957) sec. XVII. 9):

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

In applications of the multiple decrement model one is interested in the socalled partial transition probabilities (see e.g. Hoem (1969)) which occur if one or more of the forces of transition are put identically equal to 0 . Let A be an arbitrary subset of $\{1,\ldots,m\}$. Assume $i \in A$ and let $P_i(t;A)$ be the probability of transition from 0 to i in the time interval $[0,t]$ if $\alpha_j(t) \equiv 0$ for all $j \notin A$. Then we have:

$$(1.2) \qquad P_i(t;A) = \int_0^t \alpha_i(s)\exp\Big(-\sum_{k\in A}\int_0^s \alpha_k(u)du\Big)ds$$

Further we will define for $i = 1,\ldots,m$:

$$(1.3) \qquad p_i(t) = 1-P_i(t;\{i\})$$

so that

$$(1.4) \qquad p_i(t) = \exp(-\int_0^t \alpha_i(s)ds)$$

Let B be a subset of $\{1,\ldots,m\}$. We define:

$$(1.5) \quad p_B(t) = \prod_{k\in B} p_k(t) \ , \ p(t) = \prod_{k=1}^{m} p_k(t) = P_o(t)$$

$$(1.6) \quad \delta_B(t) = \sum_{k\in B} \alpha_k(t) \ , \ \delta(t) = \sum_{k=1}^{m} \alpha_k(t)$$

We also define:

$$(1.7) \qquad \beta_i(t) = \int_0^t \alpha_i(s)ds$$

When the functions $\alpha_i(t)$ are all constant it is well known how to estimate them and test hypotheses about them (see e.g. Sverdrup (1967)). When they are not constant, then different procedures are in use. One is to partition the time interval of observation into smaller intervals and then assume constancy over each subinterval. The estimation may then possibly be followed by some kind of graduation (see e.g. Hoem (1972)). Another method consists in assuming that each $\alpha_i(t)$ equals some parametric function, and then to make inference about the parameters of this function (Grenander (1956)).

When the functions $\alpha_i(t)$ are very irregular or when they have an unknown functional form, the methods mentioned above may be unapplicable or difficult to apply. Such a situation arose in the authors study of statistical methods in connection with projects for determining the security of the Intrauterine Contraceptive Device (Aalen (1972)). The nonparametric theory we will present in this paper is a further development of the methods constructed to tackle that situation. It might be of use also in other situations*.

The quantities which seem to be of greatest interest in practice are the $P_i(t;A)$. We will therefore study the estimation of these quantities when the only assumption which is made about the forces of transition is Assumption 1. We will also study estimation and testing of $\beta_i(t)$.

Our methods are partly inspired by a paper by Kaplan and Meier (1958). Our estimators may partly be regarded as generalizations of their "product-limit" estimator. Their estimator corresponds to our $\hat{p}_i(t)$ which is defined in section 4 . Kaplan and Meier approximately compute the expectation and variance of this estimator and give some other results. Our Theorem 5.2 may be regarded as an extended and somewhat more precise statement of most of their results. Our proof is however quite different from the one in Kaplan and Meier. The other estimators and results which we present are believed to be new.

We will make the following observational assumption:

Assumption 2: We observe continuously n independent processes of the kind described above, each with the same set of forces of transition. The observation of each process starts at time 0 and continues as long as we want to make inference about them. Every process is assumed to be in state 0 at time 0 .

---

* Barlow (1968) and Barlow and van Zwet (1969) have developed other nonparametric methods.

We will use the following notation:

$M_i(t)$ is the number of processes in state i at time t. This definition is made unambiguous if for $i > 0$ we say that a process is in state i at time t if the transition $0 \to i$ is performed exactly at this instant. We put $M(t) = M_0(t)$.

We define $T_\nu(A)$ to be the time of the ν-th transition from 0 to A. We define $M(\nu; i)$ by

$$(1.8) \qquad M(\nu;i) = M(T_\nu(\{i\}))$$

By our definition of $M(t)$, $M(\nu;i)$ will be the number of processes in state 0 just after the ν-th transition from 0 to i has occurred. We also define:

$$(1.9) \qquad M(\nu;A) = M(T_\nu(A))$$

$$(1.10) \qquad M_A(t) = \sum_{i\in A} M_i(t)$$

We also define for positive $M(t)$ and $p(t)$:

$$(1.11) \qquad R(t) = (M(t))^{-1} \;,\; r(t) = (p(t))^{-1}$$

All these quantities will of course be dependent on the number n of processes. If we need to stress this dependence, we will write e.g. $M_n(t)$ instead of $M(t)$, $M_{i,n}(t)$ instead of $M_i(t)$ and similar for the other quantities.

Let $\tau > 0$ be a given point of time. Let X and Y be statistics defined on the total process we observe. Throughout the paper we will use the following notation:

$$(1.12) \begin{cases} E_\tau X = E(X|M(\tau)>0) \;,\; \mathrm{var}_\tau X = \mathrm{var}(X|M(\tau)>0) \\ \mathrm{cov}_\tau(X,Y) = \mathrm{cov}(X,Y|M(\tau) > 0) \end{cases}$$

A uniformly minimum variance unbiased estimator is as usual denoted as an UMVU estimator.

a.s. is as usual used as an abbreviation of the expression "almost surely".

In accordance with standard notation we say that $f(x) = o(x)$ if $\lim \frac{1}{x} f(x) = 0$ when $x \to 0$.

We also put $f(x-) = \lim_{y \uparrow x} f(y)$ and lastly we define $0 \cdot \frac{1}{0} = 0$.

2. The estimators.

We will give heuristic justifications for the estimators we want to propose. We will use an argument which is partly analogous to the one in Kaplan and Meier (1958).

Assume that $[0,t]$ is partitioned into so small intervals that with overwhelming probability there is at most one transition in each subinterval, and so that $\alpha_i(s)$ may be assumed to be constant on each subinterval. (By Assumption 1 $\alpha_i(s)$ is uniformly continuous on $[0,t]$.) Let $\langle s,s+h]$ be one of the subintervals. By (1.1) it is seen that $M(s)h\alpha_i(s)$ is approximately equal to the conditional probability of one transition $0 \to i$ in $\langle s,s+h]$ given $M(s)$. It follows that if there is no transition in $0 \to i$ in $\langle s,s+h]$ then it is natural to "estimate" $h\alpha_i(s)$ by 0, and if there is a transition $0 \to i$ it is natural to "estimate" $h\alpha_i(s)$ by $[M(s)]^{-1}$. With a small modification in order to avoid zero denominator, this leads us to propose the following estimator for $\beta_i(t)$:

$$(2.1)\quad \hat{\beta}_i(t) = \sum_{\nu=1}^{M_i(t)} [M(\nu;i)+1]^{-1}$$

Alternatively we may write:

$$(2.2)\quad \hat{\beta}_i(t) = \int_0^t [M(s)+1]^{-1} dM_i(s)$$

which is by the definition of $M_i(t)$ well defined as a Lebesgue-Stieltjes integral.

In order to estimate $P_i(t;A)$ we will first put $A = \{i\}$ and try to estimate $p_i(t)$ (see (1.3)). We use the same partitioning as above. Since h is small we have:

$$\exp[-\int_s^{s+h} \alpha_i(s)ds] \approx 1-\alpha_i(s)h$$

As above we conclude that this expression may be estimated by 1 or $1-[M(s)]^{-1}$ according as there is 0 or 1 transition $0 \to i$ in $\langle s,s+h]$. Hence we are led to propose the following estimator for $p_i(t)$ (with the same modification as above):

$$(2.3)\quad \hat{p}_i(t) = \prod_{\nu=1}^{M_i(t)} [1-(M(\nu;i)+1)^{-1}]$$

If $M_i(t) = 0$, we put $\hat{p}_i(t) = 1$.

Now we observe that $P_i(t;A)$ may be written

$$P_i(t;A) = \int_0^t \alpha_i(s)p_A(s)ds$$

It is natural to estimate $p_A(t)$ by

$$(2.4) \qquad \hat{p}_A(t) = \prod_{k\in A} \hat{p}_k(t) = \prod_{\nu=1}^{M_A(t)} [1-(M(\nu;A)+1)^{-1}]$$

so that by the same argument as the one leading up to (2.1) and (2.2) we are led to propose the following estimator for $P_i(t;A)$:

$$(2.5) \qquad \hat{P}_i(t;A) = \int_0^t [M(s)+1]^{-1}\hat{p}_A(s-)dM_i(s)$$

By defining suitable quantities one may write this in a more explicit form like (2.1) and (2.3).

By (1.3) we naturally ought to have for all $i \in \{1,\ldots,m\}$:

$$\hat{P}_i(t;\{i\}) = 1-\hat{p}_i(t)$$

This is seen to be a consequence of the following formula which is easily proved by induction:

Let $a_1,a_2,\ldots$ be strictly positive numbers. Then we have for all $k \geq 1$:

$$\sum_{i=1}^{k} \left(\frac{1}{a_i} \prod_{j=1}^{i} \left(1- \frac{1}{a_j+1}\right)\right) = 1- \prod_{i=1}^{k} \left(1 - \frac{1}{a_j+1}\right)$$

If $A = \{1,\ldots,m\}$ we have:

$$\hat{P}_i(t;A) = \frac{1}{n}M_i(t)$$

This shows that our estimators are generalizations of the natural "frequency estimators" of $P_i(t)$. From another point of view we may regard the functions (2.3) and (2.5) as <u>generalized empirical cumulative distribution functions</u>. In the same sense one may regard $\hat{\beta}_i(t)$ as the empirical function corresponding to the theoretical function $\beta_i(t)$. This remark is relevant for the theory developed in section 8.

3. Expectations, variances and covariances.

The following theorems hold:

<u>Theorem 3.1.</u>

(3.1) $$E\,\hat{\beta}_i(t) = \beta_i(t) - \int_0^t \alpha_i(s)[1-p(s)]^n$$

(3.2) $$\operatorname{var}\hat{\beta}_i(t) = \int_0^t \alpha_i(s) E_s(R(s))ds + \rho_1(n;t)$$

(3.3) $$\operatorname{var}\hat{\beta}_i(t) = \frac{1}{n}\int_0^t \alpha_i(s) r(s) ds + \rho_2(n;t)$$ where $n^2\rho_j(n;t)$ for $j=1,2$

is bounded uniformly with respect to n and $t \in [0,1]$.

(3.4) $\hat{\beta}_i(t)$ is weakly consistent.

(3.5) $$\operatorname{cov}(\hat{\beta}_i(s), \hat{\beta}_i(t)) = \operatorname{var}\hat{\beta}_i(s) + \rho_3(n;s,t) \quad \text{for } s \le t$$

where $|\rho_3(n;s,t)| \le n[\beta_j(t) - \beta_j(s)][1-p(t)p(s)]^n$.

(3.6) $$|\text{cov}\ [\hat\beta_i(s),\hat\beta_j(t)]| \leq n[\beta_i(s)+\beta_j(t)][1-p(t)p(s)]^n \text{ for } i\neq j\ .$$

Theorem 3.2.

(3.7) $$|E\ \hat p_i(t) - p_i(t)| \leq \exp[(1-p(t))^n]-1\ .$$

(3.8) $$\text{var}\ \hat p_i(t) = \tfrac{1}{n}p_i^2(t)\int_0^t \alpha_i(s)r(s)ds+o(\tfrac{1}{n})\ .$$

(3.9) $\hat p_i(t)$ is weakly consistent.

(3.10) $$\text{cov}\ (\hat p_i(s),\hat p_i(t)) = \frac{p_i(t)}{p_i(s)}\ \text{var}\ \hat p_i(s)+o(\tfrac{1}{n}) \quad \text{for} \quad s \leq t\ .$$

(3.11) $$\text{cov}\ [\hat p_i(s),\hat p_j(t)] = o(\tfrac{1}{n}) \quad \text{for } i \neq j\ .$$

Theorem 3.3.

(3.12) $$|E\ \hat P_i(t;A) - P_i(t;A)| \leq \beta_i(t)\{\exp[(1-p(t))^n]-1\}\ .$$

(3.13) $$\begin{cases} \text{Let}\quad \sigma_A(t) = p_A^2(t)\int_0^t \delta_A(s)r(s)ds. \quad \text{Then:} \\ \text{var}\ \hat P_i(t;A) = \frac{1}{n}\int_0^t \alpha_i(s)p_A^2(s)r(s)ds \\ + \frac{2}{n}\int_0^t [\alpha_i(s)p_A(s)\int_0^s \alpha_i(u)(\sigma_A(u)p_A(u)^{-1}-p_A(u)r(u))du]ds \\ + o(\frac{1}{n}) \end{cases}$$

(3.14) $\hat P_i(t;A)$ is weakly consistent.

(3.15) $A \cap B = \phi \Rightarrow \mathrm{cov}\,[\hat{P}_i(t;A),\hat{P}_j(s;B)] = o(\frac{1}{n})$ for all $i \in A$, $j \in B$

For the proof of these theorems we need a sequence of lemmas:

<u>Lemma 3.1.</u> Let X and $U(h)$ be random variables, the latter defined for each $h \geq 0$, such that the following conditions are fulfilled:

(i) There exists a and b such that $\Pr(|X| \leq a) = 1$, $\Pr(|U(h)| \leq b) = 1$ for all $h > 0$.

(ii) $\Pr(U(h) \neq 0) = o(h)$.

Then we have:

$$EU(h) = o(h)\ ,\ E(XU(h)) = o(h)$$

$$\mathrm{var}(X+U(h)) = \mathrm{var}X+o(h)\ ,\ \mathrm{var}(XU(h)) = o(h)$$

$$\mathrm{cov}(X,U(h)) = o(h)$$

The proof of this lemma is trivial and will not be given.

<u>Lemma 3.2.</u> Let X be binomial $(n;p)$ with $p_o \leq p \leq 1$ for a given $p_o > 0$. Let us denote $E(f(X)|X>0)$ by $E^*(f(X))$. Then we have:

(i) There exists a constant k independent of n and $p \in [p_o,1]$ so that

$$|E^*(\frac{n}{X}) - \frac{1}{p}| \leq \frac{k}{n}$$

$$|E^*(\frac{n^2}{X^2}) - \frac{1}{p^2}| \leq \frac{k}{n}$$

The following statements hold uniformly on $[p_o,1]$:

(ii) $E^*|\frac{n}{X} - \frac{1}{p}| \to 0$

(iii) $E^*(\frac{n^\nu}{X^2}) \to 0$ for $\nu < 2$.

The proof is given in the Appendix.

<u>Lemma 3.3.</u> Let for $i = 1,\ldots,m$ $N_i(t,h)$ be the number of transitions $0 \to i$ in $[t,t+h\rangle$. Then we have

$$Pr(N_i(t,h) = 0\,|M(t)) = 1-M(t)\cdot\alpha_i(t)h+o(h)$$

$$Pr(N_i(t,h) = 1\,|M(t)) = M(t)\;\alpha_i(t)h+o(h)$$

$$Pr((N_i(t,h)=1)\cap \bigcap_{\substack{j=1\\ j\neq i}}^{m} (N_j(t,h)=0)\,|M(t)) = M(t)\alpha_i(t)h+o(h)$$

$$Pr(\sum_{j=1}^{m} N_j(t,h) \geq 2\,|M(t)) = o(h)$$

<u>Proof:</u> The lemma is an easy consequence of the assumptions in section 1.

<u>Lemma 3.4.</u> Let $g(x)$, $h(x)$, $f_n(x)$, $g_n(x)$; $n = 1,2,\ldots$; be real continuous functions defined on $[0,a]$, $a < \infty$, and such that $f_n(0) = c$ for all n. Assume that the $f_n(x)$ are all differentiable on $[0,a]$ and that the following limit relations hold uniformly on $[0,a]$:

$$f_n'(x)+g_n(x)f_n(x) \to h(x)$$

$$g_n(x) \to g(x)$$

when $n \to \infty$. Then:

$$f_n(x) \to \exp(-\int_0^x g(y)dy)\int_0^x h(y)\exp(\int_0^y g(t)dt)dy+c\cdot\exp(-\int_0^x g(y)dy)$$

uniformly on $[0,a]$.

The proof is given in the Appendix. We now proceed to the proofs of our theorems:

Proof of theorem 3.1. We assume $0 \leq t \leq 1$ and put $X(t) = \hat{\beta}_i(t)$. Let $I(t,h)$ be equal to 1 if there is at least one transition $0 \to i$ in $[t,t+h\rangle$, and 0 if there is none. We can then write

$$(3.16) \quad X(t+h) = X(t)+I(t,h)R(t)+U(t,h)$$

where by lemma 3.3 we have:

$$(3.17) \quad \begin{cases} Pr(I(t,h)=1 \mid M(t)) = M(t)\cdot h\cdot\alpha_i(t)+o(h) \\ Pr(I(t,h)=0 \mid M(t)) = 1-M(t)h\alpha_i(t)+o(h) \\ Pr(U(t,h)\neq 0) = o(h) \end{cases}$$

The following inequalities obviously hold:

$$(3.18) \quad 0 \leq X(t) \leq n \;,\; 0 \leq U(t,h) \leq n \;.$$

Proof of (3.1): Let $f(t) = E\,X(t)$. By (3.16), (3.17), (3.18) and lemma 3.1 we have:

$$f(t+h) = f(t)+h\alpha_i(t)Pr(M(t)>0)+o(h)$$

Since $f(0) = 0$ this gives us

$$f(t) = \int_0^t \alpha_i(s)Pr(M(s)>0)ds$$

and hence (3.1) is proved.

Proof of (3.2): Let $g(t) = \text{var}\,X(t)$. By (3.16), (3.18) and lemma 3.1 we have:

$$(3.19) \quad g(t+h) = \text{var}\,(X(t)+I(t,h)R(t))+o(h) \;.$$

We define the function $\tilde{J}(x)$ to be equal to 0 if $x = 0$ and 1 if $x \neq 0$.

By the independence of the individual processes and their Markov property we have that if $M(t)$ is given, then $I(t,h)$ is stochastically independent of $X(t)$. This gives us:

$$\operatorname{var} E\,[X(t)+I(t,h)R(t)\,|X(t),M(t)] =$$

$$= \operatorname{var}[X(t)+h\alpha_i(t)\tilde{J}(M(t))]+o(h) =$$

$$= \operatorname{var} X(t)+2h\alpha_i(t)\operatorname{cov}[X(t),\tilde{J}(M(t))]+o(h)$$

and

$$E \operatorname{var}\,[X(t)+I(t,h)R(t)\,|X(t),M(t)] =$$

$$= E \operatorname{var}\,[I(t,h)R(t)\,|X(t),M(t)] =$$

$$= E\,E\,[I(t,h)R(t)^2\,|X(t),M(t)]+o(h) =$$

$$= \alpha_i(t)hE_t(R(t))\Pr(M(t)>0)+o(h)$$

The two last results together with (3.19) gives us:

$$g(t+h) = g(t)+\alpha_i(t)hE_t(R(t))\Pr(M(t)>0)+2h\alpha_i(t)\operatorname{cov}[X(t),\tilde{J}(M(t))] +o(h)$$

and hence (3.2) follows by the following easy inequality:

$$(3.19a)\qquad \operatorname{cov}[X(t),\tilde{J}(M(t))] \leq n\Pr(M(t)=0)$$

<u>Proof of (3.3):</u> (3.3) follows from (3.2) by lemma 3.2, part (i).

<u>Proof of (3.4):</u> (3.4) follows immediately from (3.1) and (3.3).

<u>Proof of (3.5):</u> We have for $s \leq t$:

$$\operatorname{cov}\,(X(s),X(t)) = \operatorname{var}\,(X(s))+\operatorname{cov}\,(X(s),X(t)-X(s)) .$$

We have:

$$(3.20)\quad \begin{cases} \operatorname{cov}\,[X(s),X(t)-X(s)] = E\,\operatorname{cov}\,[X(s),X(t)-X(s)\,|M(s)] \\ +\operatorname{cov}\,[E\,(X(s)\,|M(s)),E\,(X(t)-X(s)\,|M(s))] \end{cases}$$

It follows from the definition of the X-process and from (3.1) that

$$(3.21)\quad E\,[X(t)-X(s)\,|M(s)] = \int_s^t \alpha_i(u)[1-(1-p(u))^{M(s)}]du$$

It is also clear that when $M(s)$ is given then $X(s)$ and $X(t)-X(s)$ are independent. Hence:

$$(3.22)\quad \operatorname{cov}\,[X(s),X(t)-X(s)\,|M(s)] = 0$$

(3.20), (3.21) and (3.22) gives us:

$$(3.23)\quad |\operatorname{cov}\,[X(s),X(t)-X(s)]| \leq n[\beta_i(t)-\beta_i(s)]E[1-p(t)]^{M(s)} =$$

$$= n[\beta_i(t)-\beta_i(s)][1-p(t)p(s)]^n$$

Hence (3.5) is proved.

<u>Proof of (3.6):</u> We put $Y(t) = \hat{\beta}_j(t)$ for a j different from i. In analogy to (3.16) we write:

$$Y(t+h) = Y(t)+J(t,h)R(t)+V(t,h)\ .$$

We define $h(t) = \operatorname{cov}_T[X(t),Y(t)]$. By lemma 3.1 we have:

$$(3.24)\quad \begin{cases} g(t+h) = g(t)+\operatorname{cov}\,[X(t),J(t,h)R(t)] \\ +\operatorname{cov}\,[Y(t),I(t,h)R(t)]+\operatorname{cov}\,[I(t,h)R(t),J(t,h)R(t)]+o(h) \end{cases}$$

We have: $\operatorname{cov}\,[X(t),J(t,h)R(t)] =$

$\operatorname{cov}\,\{E\,[X(t)\,|X(t),M(t)],E\,[J(t,h)R(t)\,|X(t),M(t)]\}$

$+E\,\operatorname{cov}\,[X(t),J(t,h)M(t)\,|X(t),M(t)] = \operatorname{cov}\,[X(t),\alpha_j(t)h\tilde{J}(M(t))]$

$+\ o(h)\ .$

Analogously: $\operatorname{cov}\,[Y(t),I(t,h)R(t)] = \operatorname{cov}[Y(t),\alpha_i(t)h\tilde{J}(M(t))]$

$+\ o(h)\ .$

By lemma 3.3 we have:

$$\Pr[I(t,h)J(t,h)\neq 0] = o(h)$$

so that:

$$\operatorname{cov}\,[I(t,h)R(t), J(t,h)R(t)] = o(h)\ .$$

The last computations together with (3.24) gives us:

$$h'(t) = \alpha_j(t)\operatorname{cov}[X(t),\tilde{J}(M(t))]+\alpha_i(t)\operatorname{cov}[Y(t),\tilde{J}(M(t))]$$

which together with $h(0) = 0$ and (3.19a) gives:

$$|h(t)|=|\operatorname{cov}\,[X(t),Y(t)]|\leq n[\beta_i(t)+\beta_j(t)][1-p(t)]^n$$

Now assume $s \leq t$. Then:

$$\operatorname{cov}\,[X(s),Y(t)] = \operatorname{cov}\,[X(s),Y(s)]+\operatorname{cov}\,[X(s),Y(t)-Y(s)]$$

$$= |\operatorname{cov}[X(s),Y(t)-Y(s)]|\leq n[\beta_j(t)-\beta_j(s)][1-p(t)p(s)]^n$$

The last inequality is proved in the same way as (3.23). Hence (3.6) and theorem 3.1 is proved.

Q.E.D.

<u>Proof of theorem 3.2:</u> We fix i and define $X(t) = \hat{p}_i(t)$. Similarly to (3.16) we can write:

$$(3.25)\quad X(t+h) = X(t)[1-I(t,h)R(t)][1-U(t,h)]$$

where the distribution of $I(t,h)$ is given by (3.17) while $\Pr[U(t,h)\neq 0] = o(h)$. We obviously have:

$$0 \leq X(t) \leq 1 \qquad\qquad 0 \leq U(t,h) \leq 1$$

<u>Proof of (3.7):</u> (3.7) is easily proved similarly to (3.1).

In the following proof $h_n(t) = o(n^{-\nu})$ will mean that $\lim_{n\to\infty} n^{\nu}h_n(t) = 0$ <u>uniformly</u> on each finite interval.

<u>Proof of (3.8):</u> Let $f(t) = \text{var}\, X(t)$. By computations analogous to the ones in the proof of (3.2) we easily get:

$$(3.26) \quad f'(t) = -2\alpha_i(t)f(t)+\alpha_i^2(t)E_t[X^2(t)R(t)]+o(n^{-1})$$

One easily shows:

$$E_t[X^2(t)R(t)] = \tfrac{1}{n}f(t)r(t)+\tfrac{1}{n}p_i^2(t)r(t)+E_t[X^2(t)(R(t)-\tfrac{1}{n}r(t))]+o(n^{-1})$$

If we put $g_n(t) = nf(t)$, (3.26) takes the form:

$$g_n'(t)+\alpha_i(t)[2-\tfrac{1}{n}r(t)]g_n(t) = \alpha_i(t)p_i^2(t)r(t)$$

$$+E_t[X^2(t)(nR(t)-r(t))]+o(1) \; .$$

By lemma 3.2, part (ii), lemma (3.4) and assumption 1 this permits us to conclude that

$$(3.27) \quad g_n(t) \to p_i^2(t)\int_0^t \alpha_i(s)r(s)ds$$

when $n \to \infty$,(uniformly on $[0,1]$). This proves (3.8).

<u>Proof of (3.9):</u> (3.9) is an immediate consequence of (3.7) and (3.8).

<u>Proof of (3.10):</u> By conditioning with respect to $(M(s),\hat{p}_i(s))$ (3.10) is easily proved similarly to (3.5).

<u>Proof of (3.11):</u> (3.11) is proved similarly to (3.6).

Q.E.D.

<u>Proof of theorem 3.3:</u> (3.12) and (3.15) are easily proved similarly to (3.1) and (3.6). (3.14) is an immediate consequence of (3.12) and (3.13). It remains to prove (3.13). For this purpose the following lemma is useful:

Lemma 3.5.

$$\lim_{n\to\infty} n \operatorname{cov}[\hat{P}_i(t;A),\hat{p}_A(t)] = p_A(t)\int_0^t \alpha_i(s)[\sigma_A(s)p_A(s)^{-1}-p_A(s)r(s)]ds$$

where $\sigma_A(t) = p_A^2(t)\int_0^t \delta_A(s)r(s)ds$.

This lemma may be proved in a way similar to the proof of (3.8). The same methods may then be used to derive (3.13) by means of lemma 3.5.

4. Sufficient and complete statistics. Proof that the estimators are based on minimal sufficient statistics.

Let us assume that we observe each process in the time interval $[0,1]$. Our observation of the ν-th process is completely described by the pair (U_ν,Y_ν) where U_ν is the time (in $[0,1]$) the process stays in state 0 , while Y_ν is the state at time $t = 1$. The likelihood of (U_ν,Y_ν) is given by

$$\alpha_{Y_\nu}(U_\nu)p(U_\nu)$$

where we define $\alpha_0(t) \equiv 1$. By the independence of the individual processes the total likelihood of all pairs (U_ν,Y_ν) may be written

$$(4.1) \qquad \prod_{\nu=1}^{n} [\alpha_{Y_\nu}(U_\nu)p(U_\nu)]$$

Let N be the number of processes which leave state 0 during $[0,1]$. Let $T_1 < T_2 < \ldots < T_N$ be the ordered times transition of these processes and let $X_1,\ldots,X_N$ be the corresponding final states. It is easily shown that the likelihood of $Z = [N,(T_1,X_1),\ldots,(T_N,X_N)]$ is given by

$$(4.2) \qquad \frac{n!}{(n-N)!} \prod_{\nu=1}^{n} [\alpha_{X_\nu}(T_\nu)p(T_\nu)]p^{n-N}$$

where $p = p(1)$. By dividing (4.1) by (4.2) we get the conditional likelihood of $[(U_1,Y_1),\ldots,(U_n,Y_n)]$ given Z equal to

$$(4.3) \qquad \frac{(n-N)!}{n!}$$

if $[U_1,Y_1),\ldots,(U_n,Y_n)]$ is a permutation of $[(T_1,X_1),\ldots,(T_N,X_N),(1,0),\ldots,(1,0)]$. (The last vector should of course contain n elements.) Hence we have the following lemma:

Lemma 4.1. Z is a sufficient statistic.

(For a more stringent proof of this than the one given above one would have to use the measure-theoretic definition of conditional probability. This is not difficult, but we will not do it in this paper.)

We will now turn to the question of the completeness of Z. We will make the following definitions for $t \in [0,1]$:

$$(4.4) \qquad \gamma_i(t) = \begin{cases} \dfrac{\alpha_i(t)}{\delta(t)} & \text{if } \delta(t) > 0 \\ \dfrac{1}{m} & \text{if } \delta(t) = 0 \end{cases} \quad ; \; i = 1,\ldots,m$$

$$(4.5) \qquad \gamma_0(t) \equiv 1$$

$$(4.6) \qquad f(t) = (1-p)^{-1}\delta(t)p(t)$$

The likelihood (4.2) of Z may then be put in the form:

$$(4.7)\quad \frac{n!}{(n-N)!}\prod_{\nu-1}^{N}[\gamma_{X_\nu}(T_\nu)f(T_\nu)]p^{n-N}(1-p)^N$$

Let α_i be the function $\alpha_i(t)$ restricted to $[0,1]$. We define the following sets:

$$(4.8)\quad \mathfrak{A}=\{\alpha=(\alpha_1,\dots,\alpha_m)\,|\,\alpha_i \text{ satisfies A.1 for } i=1,\dots,m\}$$

(A.1 denotes the assumption 1 of section 1.)

$$(4.9)\quad \Gamma=\{\gamma=(\gamma_1,\dots,\gamma_m)\,|\,\alpha\in\mathfrak{A}\}$$

(4.10) $\mathfrak{F}$ is the class of continuous densities on $[0,1]$

$$(4.11)\quad I=\langle 0,1]$$

We recall that p may be written in the following way

$$(4.12)\quad p=\exp(-\int_0^1 \delta(t)dt)$$

The following lemma is easily proved:

<u>Lemma 4.2.</u> (4.4), (4.6) and (4.12) constitute a one-to-one mapping from $\mathfrak{A}$ onto $\Gamma\times\mathfrak{F}\times I$.

We are now ready to prove the completeness of Z. <u>Let us fix $\gamma=\gamma^{(0)}$ and $f=f_0$</u>. By applying the factorization criterion for sufficiency to (4.7) it is seen that N is in this case sufficient for Z.

Let us denote by $\mathfrak{K}(\gamma^{(o)},f_o)$ the part of $\mathfrak{K}$ in which $\gamma = \gamma^{(o)}$ and $f = f_o$. By lemma 4.2 p runs through the whole of I when α runs through $\mathfrak{K}(\gamma^{(o)},f_o)$. Since N is binomial $(n,1-p)$ it follows that N is also complete. Hence for an arbitrary integrable g we have:

$$(4.13)\qquad \begin{cases} Eg(T_1,\dots,T_N,N) = 0 \quad \forall\alpha \in \mathfrak{K}(\gamma^{(o)},f_o) \\ \qquad\Downarrow \\ E[g(T_1,\dots,T_k,k)\,|N=k] = 0 \\ \text{for } k = 0,1,\dots,n \text{ and } \gamma = \gamma^{(o)},\ f = f_o \end{cases}$$

(If $k = 0$ $\{T_1,\dots,T_k\}$ is taken to be the empty set.)

It is easily established that the likelihood of $(T_1,\dots,T_N,N)$ is given by:

$$(4.14)\qquad \frac{n!}{(n-N)!}\prod_{\nu=1}^{N} f(T_\nu)\ p^{n-N}(1-p)^N$$

Since N is binomial $(n,1-p)$ it follows that the conditional likelihood of $T_1,\dots,T_k$ given $N = k$ (≥ 1) is given by:

$$(4.15)\qquad k!\prod_{\nu=1}^{k} f(T_\nu)$$

This is equal to the likelihood of the order statistics corresponding to k independent random variables with the density f. Hence by the well known results about the completeness of the order statistics in the nonparametric case (see e.g. Lehmann (1959) p. 133) it follows that:

$$(4.16)\quad \begin{cases} E[g(T_1,\dots,T_k,k)\,|\,N=k] = 0 \qquad \forall f \in \mathfrak{F} \\ \Downarrow \\ g(T_1,\dots,T_k,k) = 0 \quad \text{a.s. for all} \quad f \in \mathfrak{F} \\ \text{for} \quad k = 0,1,\dots,n \end{cases}$$

Let $\mathfrak{K}(\gamma^{(o)})$ be the part of $\mathfrak{K}$ in which $\gamma = \gamma^{(o)}$. By lemma 4.2 f runs through the whole of $\mathfrak{F}$ when α runs through $\mathfrak{K}(\gamma^{(o)})$. Since f_o is arbitrary in (4.13) it follows by (4.13) and (4.16) that

$$(4.17)\quad \begin{cases} Eg(T_1,\dots,T_N,N) = 0 \qquad \forall \alpha \in \mathfrak{K}(\gamma^{(o)}) \\ \Downarrow \\ g(T_1,\dots,T_N,N) = 0 \qquad \text{a.s. for all } \alpha \in \mathfrak{K}(\gamma^{(o)}) \end{cases}$$

By dividing (4.7) with (4.14) we get the conditional likelihood of $(X_1,\dots,X_N)$ given $(T_1,\dots,T_N,N)$:

$$(4.18)\qquad \prod_{\nu=1}^{N} [\gamma_{X_\nu}(T_\nu)]$$

It follows that $(T_1,\dots,T_N,N)$ is sufficient when γ is given. Together with (4.17) this gives us for any integrable h :

$$Eh(Z) = 0 \qquad \forall \alpha \in \mathfrak{K}$$
$$\Downarrow$$
$$(4.19)\quad E[h(Z)\,|\,T_1,\dots,T_N,N] = 0 \quad \text{a.s. for any} \quad \alpha \in \mathfrak{K}$$

By (4.14) the distribution of $(T_1,\dots,T_N,N)$ is independent of γ . This means that the set with probability 1 where

(4.19) holds is independent of γ . Let $(t_1,\ldots,t_k,k)$ be a point in this set. With probability 1 the t's are distinct, hence we assume this to be the case. We define a set of numbers $x_1,\ldots,x_k$ so that

$$x_i \in \{1,\ldots,m\} \quad \text{for} \quad i = 1,\ldots,k \ .$$

We now let γ be an element of Γ for which

$$\gamma_{x_i}(t_i) = 1 \qquad i = 1,\ldots,k \ .$$

By (4.18) we have:

$$\Pr[\bigcap_{\nu=1}^{k}(X_\nu=x_\nu)\,|\bigcap_{\nu=1}^{k}(T_\nu=t_\nu)\cap N=k] = 1$$

for our choice of γ . If (4.19) is true then this implies

$$h(z) = 0 \quad \text{for} \quad z = [k,(t_1,x_1),\ldots,(t_k,x_k)]$$

and the completeness of Z is proved. Together with lemma 4.1 this gives us the following theorem:

<u>Theorem 4.1.</u> $[N,(T_1,X_1),\ldots,(T_N,X_N)]$ is a sufficient and complete statistic with respect to the class of distributions which is generated by the set of those α that satisfy assumption 1.

We will prove the following corollary

<u>Corollary.</u> $[N,(T_1,X_1),\ldots,(T_N,X_N)]$ is minimal sufficient.

Proof: Let μ be the Lebesgue measure. We define the measure ρ on the Borel sets of $[0,1]$ in the following way:

$$\rho(A) = \mu(A) \quad \text{for} \quad A \subset [0,1\rangle$$

$$\rho(\{1\}) = 1$$

We define the measure $\varkappa$ on all subsets of $\{0,1,\dots,m\}$ in the following way:

$$\varkappa(\{i\}) = P_i(1) \quad \text{for all} \quad i \in \{0,1,\dots,m\}$$

We define the function $f(t;i)$ in the following way:

$$f(t;0) = \begin{cases} 0 & \text{if} \quad 0 \leq t < 1 \\ 1 & \text{if} \quad t = 1 \end{cases}$$

$$f(t;i) = \begin{cases} P_i'(t)[P_i(1)]^{-1} & \text{if} \quad i > 0 \;,\; 0 \leq t < 1 \\ 0 & \text{if} \quad i > 0 \;,\; t = 1 \end{cases}$$

Then it is easily seen that we can write:

$$\Pr(U_\nu \leq u \cap Y_\nu \leq y) = \int_0^y \int_0^u f(t;i)d\rho(t)d\varkappa(i)$$

Hence our class of probability distributions of $[(U_1,Y_1),\dots,(U_n,Y_n)]$ is dominated by a σ-finite measure. The corollary then follows by Theorem 9 in Sverdrup (1966).

Q.E.D.

It is easily seen that all our estimators are functions of Z. Hence they are based on a minimal sufficient statistic. Besides, it follows from theorem 4.1 that our estimators are UMVU estimators of their expectations. Taken together with (3.1), (3.7) and (3.12) this shows us that our estimators should be quite good when n are not too small.

5. Strong consistency. A generalization of the Glivenko-Cantelli theorem.

For the Glivenko-Cantelli theorem, see e.g. Chung (1968, p. 124).

We will prove the following theorems. All limits are taken with respect to n.

Theorem 5.1. $\Pr[\sup_{0\le t\le 1} |\hat{\beta}_{i,n}(t)-\beta_i(t)|\to 0] = 1$.

Theorem 5.2. $\Pr[\sup_{t\ge 0} |\hat{p}_{i,n}(t)-p_i(t)|\to 0] = 1$.

Theorem 5.3. $\Pr[\sup_{0\le t\le 1} |\hat{P}_{i,n}(t;A)-P_i(t;A)|\to 0] = 1$.

Theorem 5.2 is seen to be a generalization of the Glivenko-Cantelli theorem. The following lemmas will be needed for the proofs:

Lemma 5.1. Let $F_0, F_1, F_2, \ldots$ be right-continuous increasing functions on $[0,1]$ such that $F_n(0) = 0$ and $F_n(1) \le 1$ for all n. Let $g_0, g_1, g_2, \ldots$ be stepwise continuous functions on $[0,1]$ such that $0 \le g_0(x) \le a < \infty$ for all $x \in [0,1]$. Assume that

$$F_n(x) \underset{n\to\infty}{\to} F_0(x) \quad , \quad g_n(x) \underset{n\to\infty}{\to} g_0(x) \quad \text{uniformly on } [0,1] .$$

Then: $\int_0^x g_n(y)dF_n(y) \to \int_0^x g_o(y)dF_o(y)$ uniformly on $[0,1]$.

Lemma 5.2. Assume that $f_n(x)$ (≥ 0) converges towards $f(x)$ uniformly on $[0,1]$ when $n \to \infty$. Assume that there exists an $a > 0$ such that $f(x) > a$ on $[0,1]$. Then the following limits hold uniformly on $[0,1]$:

$$f_n(x)^{-1} \to f(x)^{-1}$$

$$[f_n(x)+\tfrac{1}{n}]^{-1} \to f(x)^{-1}$$

$$-n\log[1-(nf_n(x)+1)^{-1}] \to f(x)^{-1}$$

These lemmas may easily be proved by standard methods. For completeness the proofs are given in the Appendix.

Proof of theorem 5.1. We define

$$X_n(t) = \frac{1}{n} M_{i,n}(t) \;,\; Y_n(t) = n[M_n(t)+1]^{-1}$$

Hence we may write:

$$\hat{\beta}_{i,n}(t) = \int_0^t Y_n(s)dX_n(s)$$

With probability 1 X_n satisfies the assumption about F_n in lemma 5.1 , while Y_n satisfies the assumption about g_n . By the Glivenko-Cantelli theorem and lemma 5.2 the following statements are true with probability 1 when $n \to \infty$:

$$X_n(t) \to P_i(t) \quad \text{uniformly on} \quad [0,1]$$

$$Y_n(t) \to r(t) \quad \text{uniformly on} \quad [0,1]$$

$r(t)$ satisfies the assumptions about $g_o(x)$ in lemma 5.1. Hence with probability 1:

$$\hat{\beta}_{i,n}(t) \to \int_0^t r(s)P_i'(s)ds = \beta_i(t)$$

uniformly on $[0,1]$.

Q.E.D.

Proof of theorem 5.2. We may write:

$$\hat{p}_{i,n}(t) = \exp\left\{\int_0^t \log[1-(M_n(s)+1)^{-1}]dM_{i,n}(t)\right\}$$

We define:

$$X_n(t) = \frac{1}{n} M_{i,n}(t) \text{ , } Y_n(t) = -n\log[1-(M_n(t)+1)^{-1}]$$

By the Glivenko-Cantelli theorem and lemma 5.2 the following statements are true with probability 1 for an arbitrary finite interval $[0,a]$:

$$X_n(t) \to P_i(t) \text{ uniformly on } [0,a]$$

$$Y_n(t) \to r(t) \text{ uniformly on } [0,a]$$

By lemma 5.1 we may then conclude:

$$\Pr[\sup_{0\leq t\leq a} |\hat{p}_{i,n}(t)-p_i(t)| \to 0] = 1$$

Let $\epsilon > 0$ be given. Let $p_i = \lim_{t\to\infty} p_i(t)$. Choose a so large that $|p_i(t)-p| \leq \epsilon$ for $t \geq a$. Let $z_{i,n}(t)$; $n = 1,2,\ldots$; be a realization of $\hat{p}_{i,n}(t)$ for which $\sup_{0\leq t\leq a} |z_{i,n}(t)-p_i(t)| \to 0$. Choose N so large that

$$n \geq N \Rightarrow \sup_{0\leq t\leq a} |z_{i,n}(t)-p_i(t)| \leq \epsilon$$

Then, since $z_{i,n}(t)$ is decreasing in t, we have:

$$n \geq N \Rightarrow \sup_{t \geq 0} |z_{i,n}(t) - p_i(t)| \leq 2\epsilon \ .$$

Q.E.D.

Proof of theorem 5.3. Of course theorem 5.2 remains true when i is replaced by a set A of states. This fact together with lemmas 5.1 and 5.2 immediately yields theorem 5.3.

Q.E.D.

6. Estimators of var $\hat{\beta}_i(t)$ and var $\hat{p}_i(t)$.

By considerations of the same kind as in section 2 it is easily deduced from (3.2) that the following is a reasonable estimator of var $\hat{\beta}_i(t)$:

$$(6.1) \qquad \hat{\sigma}_i(t) = \sum_{\nu=1}^{M_i(t)} [M(\nu;i)+1]^{-2}$$

We put

$$(6.2) \qquad \sigma_i(t) = \text{var}\ \hat{\beta}_i(t)$$

The following theorem holds:

Theorem 6.1.

(i) $E\hat{\sigma}_i(t) = \int_0^t \alpha_i(s) E_s(R(s)) \Pr(M(s) > 0) ds$

(ii) $\Pr\{\sup_{0 \leq t \leq 1} [n|\hat{\sigma}_{i,n}(t) - \sigma_{i,n}(t)|] \underset{n \to \infty}{\to} 0\} = 1$

Proof: Similarly to (3.1) we easily prove (i).
(3.2) shows the significanse of (i).
(ii) is proved similarly to theorem 5.1.

Q.E.D.

We put:

(6.3) $\rho_i(t) = \text{var } \hat{p}_i(t)$

Let $T_o = 0$ and let $T_1 < T_2 < \ldots$ be the ordered times of transition $0 \to i$ in $[0,\tau]$. We define a stochastic process $\hat{\rho}_i(t)$ on $[0,\tau]$ in the following way:

$$(6.4)\quad \begin{cases} \hat{\rho}_i(0) = 0 \\ \hat{\rho}_i(T_\nu) = \hat{\rho}_i(T_{\nu-1})[1-\frac{2}{M(T_\nu)+1}] + \frac{\hat{p}_i^2(T_\nu)}{M^2(T_\nu)} \quad \text{for} \\ \qquad\qquad \nu = 1,2,\ldots,M_i(\tau) \\ \hat{\rho}_i(t) = \hat{\rho}_i(T_\nu) \text{ if } T_\nu \leq t < T_{\nu+1} \text{ and } \nu = 1,2,\ldots,M_i(\tau)-1 \\ \hat{\rho}_i(t) = \rho_i(T_\nu) \text{ if } T_\nu \leq t \leq \tau \text{ and } \nu = M_i(\tau) . \end{cases}$$

The following theorem holds:

Theorem 6.2.

(i) $E\hat{\rho}_i(t) = \rho_i(t)+o(\frac{1}{n})$

(ii) $\Pr\{\sup_{0\leq t\leq 1} [n|\hat{p}_{i,n}^2(t)\hat{\sigma}_{i,n}(t)-\rho_{i,n}(t)|] \underset{n \to \infty}{\to} 0\} = 1$

Proof: We may write:

$$\hat{\rho}_i(t+h) = \hat{\rho}_i(t)[1-2I(t,h)R(t)]+\hat{p}_i^2(t)R^2(t)I(t,h)+U(t,h)$$

where the distribution of $I(t,h)$ is given by (3.17) while $Pr[U(t,h)\neq 0] = o(h)$. Let $f(t) = E\,\hat{p}_i(t)$. It is easily shown that $f(t)$ satisfies equation (3.26) with the same side-condition $f(0) = 0$. Hence (i) is true.

Part (ii) of the theorem is an easy consequence of theorems 5.1 and 5.2 and the fact that (3.27) holds uniformly on $[0,1]$.

Q.E.D.

7. Asymptotic normality.

I have found no central limit theorem in the literature by which it is possible to prove the asymptotic normality (AN) of our estimators directly. In this section I will give a proof of the AN of $[\hat{\beta}_1(t),\dots,\hat{\beta}_m(t)]$ by means of characteristic functions. Then I will use that result to deduce the AN of $[\hat{p}_1(t),\dots,\hat{p}_m(t)]$. I have not yet studied the question of AN for $\hat{P}_j(A;t)$ in general, but I believe that AN for these estimators may be proved in a similar way to that of $\hat{\beta}_j(t)$.

We define for $j = 1,\dots,m$:

$$(7.1) \qquad X_{j,n}(t) = n^{\frac{1}{2}}[\hat{\beta}_{j,n}(t)-\beta_j(t)]$$

We will prove the following theorem:

Theorem 7.1. $X_{j,n}(t)$; $j = 1,\dots,m$; are asymptotically independent and normally distributed with means 0 and variances

$$\int_0^t \alpha_j(s)r(s)ds \qquad j = 1,\dots,m$$

respectively.

For the proof we need the following lemma:

<u>Lemma 7.1.</u> Let X_n be a statistic defined on the process we observe. Then:

$$\Pr[X_n \leq x \mid M_n(\tau) > 0] \underset{n \to \infty}{\to} F(x)$$

$$\Updownarrow$$

$$\Pr(X_n \leq x) \underset{n \to \infty}{\to} F(x)$$

The proof is trivial and will not be given.

<u>Proof of the theorem:</u>

Let $\lambda_1, \ldots, \lambda_m$ be specified real numbers. Define

$$\varphi_n(t) = E \exp[i \sum_{j=1}^{m} \lambda_j X_{j,n}(t)]$$

where $i = \sqrt{-1}$. We may write equation (3.16) in the form:

$$(7.2) \quad \hat{\beta}_{j,n}(t+h) = \hat{\beta}_{j,n}(t) + I_{j,n}(t,h) R_n(t) + U_{j,n}(t,h)$$

Lemma 3.3 gives us the simultaneous distribution of the $I_{j,n}(t,h)$ for $j = 1, \ldots, m$ and also permits us to conclude that

$$(7.3) \quad \Pr[\bigcup_{j=1}^{m} (U_{j,n}(t,h) \neq 0)] = o(h)$$

In this proof $f(n;t) = o(n^{-1})$ will mean that $\lim_{n \to \infty} nf(n;t) = 0$ <u>uniformly</u> on every finite interval.

We define

$$(7.4)\qquad W_{1,n}(t) = \exp[i\sum_{j=1}^{m}\lambda_j X_{j,n}(t)]$$

$$(7.5)\qquad W_{2,n}(t,h) = \exp\{i\sum_{j=1}^{m}\lambda_j n^{\frac{1}{2}}[I_{j,n}(t,h)R_n(t)-\int_t^{t+h}\alpha_j(s)ds]\}$$

$$(7.6)\qquad W_{3,n}(t,h) = \exp[i\sum_{j=1}^{m}\lambda_j n^{\frac{1}{2}}U_{j,n}(t,h)]$$

By (7.2) we have

$$\varphi_n(t+h) = E\,[W_{1,n}(t)W_{2,n}(t,h)W_{3,n}(t,h)]\ .$$

By (7.3) and lemma 3.1 we have:

$$\varphi_n(t+h) = E\,[W_{1,n}(t)W_{2,n}(t,h)]+o(h)$$

If $M_n(t)$ is given then $(I_{1,n}(t,h),\dots,I_{m,n}(t,h))$ is independent of $(X_{1,n}(t),\dots,X_{m,n}(t))$ so that

$$(7.7)\qquad \varphi_n(t+h) = E\{E[W_{1,n}(t)\,|M_n(t)]E[W_{2,n}(t,h)\,|M_n(t)]\}$$

For a moment we will drop n from the notation. By lemma 3.3 we have for $M(t) > 0$:

$$(7.8)\qquad \begin{cases} E\,[W_2(t,h)\,|M(t)]=[1-M(t)\delta(t)h]\exp[-in^{\frac{1}{2}}\sum_{j=1}^{m}\lambda_j\int_t^{t+h}\alpha_j(s)ds] \\ +h\sum_{j=1}^{m}\{M(t)\alpha_j(t)\exp[i\lambda_j n^{\frac{1}{2}}(R(t)-\int_t^{t+h}\alpha_j(s)ds)]\}+o(h) \end{cases}$$

From the Taylor development of e^{ix} one has:

$$(7.9) \qquad e^{ix} = 1+ix-\tfrac{1}{2}x^2+\tfrac{1}{3}x^3\xi(x)$$

with $|\xi(x)| \le 1$, $\xi(x)$ continuous, for all real x . By applying (7.9) to the exponential functions in (7.8) we get for $M(t) > 0$:

$$(7.10) \begin{cases} E\,[W_2(t,h)|M(t)] = 1-hnR(t)\sum_{j=1}^{m}\tfrac{1}{2}\lambda_j^2\alpha_j(t) \\ +hn^{\frac{3}{2}}R^2(t)\sum_{j=1}^{m}\tfrac{1}{3}\theta_j(t,h)\lambda_j^2\alpha_j(t) \end{cases}$$

where $\theta_j(t,h)$; $|\theta_j(t,h)| \le 1$; is a random variable which depends on t and h only via $R(t)-\int_t^{t+h}\alpha_j(s)ds$. Hence $\theta_j(t) = \lim_{h\to 0}\theta_j(t,h)$ exists, and so by (7.7):

$$(7.11) \begin{cases} \varphi_n'(t) = -a(t)E_t[nW_{1,n}(t)R_n(t)] \\ +\sum_{j=1}^{m}\tfrac{1}{3}\lambda_j^2\alpha_j(t)E_t[n^{\frac{3}{2}}R_n^2(t)\theta_{j,n}(t)W_{1,n}(t)]+o(n^{-1}) \end{cases}$$

where:

$$(7.12) \quad a(t) = \sum_{j=1}^{m}\tfrac{1}{2}\lambda_j^2\alpha_j(t)$$

We have:

$$(7.13) \quad E_t[nW_{1,n}(t)R_n(t)] = r(t)\varphi_n(t)+E_t[W_{1,n}(t)(nR_n(t)-r(t))] + o(n^{-1})$$

By lemma 3.2 we have for any $\tau < \infty$:

$$E_t[W_{1,n}(t)(nR_n(t)-r(t))] \to 0 \text{ uniformly on } [0,\tau]$$

$$E_t[n^{\frac{3}{2}}R_n^2(t)\theta_{j,n}(t)W_{1,n}(t)] \to 0 \quad \text{uniformly on} \quad [0,\tau]$$

This fact together with (7.11), (7.13) and lemma 3.4 allows us to conclude that

$$\varphi_n(t) \to \exp[-\int_0^t a(s)r(s)ds]$$

The theorem follows by the definition (7.12) of $a(t)$.

Q.E.D.

We now define for $j = 1,\ldots,m$:

$$(7.14) \quad Y_{j,n}(t) = n^{\frac{1}{2}}[\hat{p}_{j,n}(t)-p_j(t)]$$

We will prove the following theorem:

<u>Theorem 7.2.</u> $Y_{j,n}(t)$; $j = 1,\ldots,m$; are asymptotically independent and normally distributed with means 0 and variances

$$p_j^2(t)\int_0^t \alpha_j(s)r(s)ds \qquad j = 1,\ldots,m$$

respectively.

<u>Proof:</u> By (2.3) we have:

$$\log \hat{p}_{j,n}(t) = \sum_{\nu=1}^{M_{j,n}(t)} \log\{1-[M_n(\nu;j)+1]^{-1}\} .$$

The following result is easily derived:

For $0 \leq x \leq \frac{1}{2}$ there exists a function $\xi(x)$, $0 \leq \xi(x) \leq 1$, such that $\log(1-x) = - x - \xi(x)x^2$.

We assume $M_n(\tau) > 0$ and $0 \leq t \leq \tau$. Then $[M_n(\nu;j)+1]^{-1} \leq \frac{1}{2}$ and hence we can write:

$$(7.15)\quad \log\hat{p}_{j,n}(t) = -\hat{\beta}_{j,n}(t)-\theta_{j,n}(t)\hat{\sigma}_{j,n}(t)$$

where $\theta_{j,n}(t)$ is a random variable with $0 \leq \theta_{j,n}(t) \leq 1$ and where $\hat{\sigma}_{j,n}(t)$ is given by (6.1).

From now on we drop n and t from the notation in order to achieve some simplification. By (7.14) and (7.15) we have:

$$Y_j = n^{\frac{1}{2}}(\exp(-\hat{\beta}_j-\theta_j\hat{\sigma}_j)-p_j)$$

and hence:

$$|Y_j-n^{\frac{1}{2}}[\exp(-\hat{\beta}_j)-p_j]| = n^{\frac{1}{2}}\exp(-\hat{\beta}_j)[1-\exp(-\theta_j\hat{\sigma}_j)]$$

$$\leq n^{\frac{1}{2}}\exp(-\theta_j\hat{\sigma}_j)[\exp(\theta_j\hat{\sigma}_j)-1] \leq n^{\frac{1}{2}}[\exp(\hat{\sigma}_j)-1] .$$

By the mean value theorem there exists a random variable V such that $0 \leq V \leq \hat{\sigma}_j$ and

$$(7.16)\quad |Y_j-n^{\frac{1}{2}}[\exp(-\hat{\beta}_j)-p_j]| \leq n^{\frac{1}{2}}e^{V}\hat{\sigma}_j$$

By (3.3) and theorem 6.1 the expression on the right side in (7.16) converges almost surely to 0 . Hence

$$(7.17)\quad Y_{j,n}(t) = n^{\frac{1}{2}}[\exp(-\hat{\beta}_{j,n}(t))-p_j(t)]+Z_{j,n}(t)$$

where by lemma 7.1:

$$(7.18)\quad \operatorname{plim}_{\tau} Z_{j,n}(t) = 0 \quad \text{when} \quad n \to \infty$$

$\operatorname{plim}_{\tau}$ denotes that we assume $M_n(\tau) > 0$.

By the Taylor formula we have:

$$(7.19)\quad \exp(-\hat{\beta}_j)-p_j = -p_j(\hat{\beta}_j-\beta_j)+\tfrac{1}{2}e^W(\hat{\beta}_j-\beta_j)^2$$

for a random variable W such that $|W-\beta_j| \le |\hat{\beta}_j-\beta_j|$. We have:

$$n^{\frac{1}{2}}e^W(\hat{\beta}_j-\beta_j)^2 = e^W[n^{\frac{1}{4}}(\hat{\beta}_j-\beta_j)]^2$$

which by theorem 3.1 converges in probability to 0 when $M_n(\tau) > 0$ is given. This fact together with (7.17) and (7.19) yields:

$$Y_{j,n}(t) = n^{\frac{1}{2}}p_j(t)[\hat{\beta}_{j,n}(t)-\beta_j(t)]+Z_{j,n}(t)+U_{j,n}(t)$$

where $U_{j,n}(t)$ converges in probability to 0 when $M_n(\tau) > 0$ is given. A limit theorem due to Cramèr (1946, Chapter 20.6) together with lemma 7.1 gives us that the $Y_{j,n}(t)$; $j = 1,\dots,m$; have the same simultaneous asymptotic distribution as

$$n^{\frac{1}{2}}p_j(t)[\hat{\beta}_{j,n}(t)-\beta_j(t)] \qquad j = 1,\dots,m .$$

The theorem now follows by theorem 7.1.

Q.E.D.

We now define:

$$V_{j,n}(t) = n^{-\frac{1}{2}}[\hat{\sigma}_{j,n}(t)]^{-1}[\hat{\beta}_{j,n}(t)-\beta_j(t)]$$

$$W_{j,n}(t) = n^{-\frac{1}{2}}[\hat{p}_{j,n}(t)]^{-2}[\hat{\sigma}_{j,n}(t)]^{-1}[\hat{p}_{j,n}(t)-p_j(t)] .$$

The following corollary to theorem 7.1 is an immediate consequence of (3.3) and part (ii) of theorem 6.1:

Corollary to theorem 7.1. $V_{j,n}(t)$; $j = 1,\dots,m$; are asymptotically independent and normally (0,1) distributed.

The following corollary to theorem 7.2 is an immediate consequence of (3.8) and part (ii) of theorem 6.2:

Corollary to theorem 7.2. $W_{j,n}(t)$; $j = 1,\dots,m$; are asymptotically independent and normally (0,1) distributed.

These two corollaries are obviously relevant for testing statistical hypotheses about $\beta_j(t)$ and $p_j(t)$.

8. Approximation of $\hat{\beta}_j(t)$ by a normal process with independent increments. A test of Kolmogorov-Smirnov type.

For a fixed j we define:

$$X_n(t) = n^{\frac{1}{2}}[\hat{\beta}_{j,n}(t)-\beta_j(t)] .$$

We will study the convergence of X_n as a stochastic process. First we will give a theorem on the convergence of the finite-dimensional distributions. Let $t_o,t_1,\dots,t_l$ for $l \geq 2$ be numbers such that $0 = t_o < t_1 < t_2 < \dots < t_l$. Then we have:

Theorem 8.1. $X_n(t_1),X_n(t_2)-X_n(t_1),\dots,X_n(t_l)-X_n(t_{l-1})$ are asymptotically independent and normally distributed.

Remark: The asymptotic means and variances follow from theorem 7.1.

Proof: Let $\lambda_1,\dots,\lambda_l$ be arbitrary given numbers. We define for $k = 1,\dots,l$:

$$Y_n(k) = \sum_{\nu=1}^{k} \lambda_\nu[X_n(t_\nu)-X_n(t_{\nu-1})] .$$

We also define:

$$\varphi_n(t_1,\dots,t_k) = E \exp(iY_n(k))$$

We state the following induction hypothesis:

$$(8.1) \qquad \varphi_n(t_1,\dots,t_k) \to \exp[-\sum_{\nu=1}^{k} \frac{\lambda_\nu^2}{2} \int_{t_{\nu-1}}^{t_\nu} \alpha_j(s)r(s)ds]$$

By theorem 7.1 this is true for $k = 1$. We will prove that if it is true for k, then it must be true when k is replaced by $k+1$. We have:

$$\varphi_n(t_1,\dots,t_k) = E \exp[iY_n(k)+i\lambda_{k+1}(X_n(t_{k+1})-X_n(t_k))] .$$

If we put

$$U_n = E[\exp(iY_n(k))|M_n(t_k)]$$

$$V_n = E\{\exp[i\lambda_{k+1}(X_n(t_{k+1})-X_n(t_k))]|M_n(t_k)\}$$

we can write:

$$(8.2) \qquad \varphi_n(t_1,\dots,t_{k+1}) = E(U_nV_n)$$

we can put V_n in the form:

$$(8.3) \quad V_n=E\{\exp[i\lambda_{k+1}(\tfrac{1}{n}M_n(t_k))^{-\frac{1}{2}}(\tfrac{1}{n}M_n(t_k))^{\frac{1}{2}}(X_n(t_{k+1})-X_n(t_k))]| M_n(t_k)\}$$

By the strong law of large numbers we known that if $n \to \infty$, then almost surely $M_n(t_k) \to \infty$ in such a way that

$$(8.4) \qquad [\tfrac{1}{n}M_n(t_k)]^{-\frac{1}{2}} \to p(t_k)^{-\frac{1}{2}}$$

From theorem 7.1 it then follows that the conditional distribution given $M_n(t_k)$ of

$$[\tfrac{1}{n}M_n(t_k)]^{\frac{1}{2}}[X_n(t_{k+1})-X_n(t_k)]$$

$$= [M_n(t_k)]^{\frac{1}{2}}[\sum_{\nu=M_{j,n}(t_k)+1}^{M_{j,n}(t_{k+1})} (M(\nu;j)+1)^{-1} \int_{t_k}^{t_{k+1}} \alpha_j(s)ds]$$

almost surely converges to a normal distribution with expectation 0 and variance

$$(8.5) \qquad \int_{t_k}^{t_{k+1}} \alpha_j(s)\exp(\int_{t_k}^{s} \delta(u)du)ds$$

Since the convergence of a sequence of characteristic functions is uniform on every finite interval, it follows by (8.3), (8.4) and (8.5) that

$$(8.6) \qquad V_n \to \exp[-\tfrac{1}{2}\lambda_{k+1}^2 \int_{t_k}^{t_{k+1}} \alpha_j(s)r(s)ds]$$

almost surely. We denote the limits in (8.1) and (8.6) by a and b respectively. We have:

$$E(U_nV_n)-ab = E[U_n(V_n-b)]+b(EU_n-a)$$

so that

$$|E(U_nV_n)-ab| \leq E|V_n-b|+|EU_n-a| \; .$$

By the definition of a and b and since $|U_n| \leq 1$ and $|V_n \leq 1|$, the right side of the above inequality converges to 0 . Hence (8.1) is proved with k replaced by $k+1$.

Q.E.D.

The theorem shows that the finite-dimensional distributions of X_n converge to the finite-dimensional distributions of a normal process with independent increments (NPII). For the purpose of getting a test of Kolmogorov-Smirnov type we need to prove that $\sup_{0\leq t\leq 1} |X_n(t)|$ converges in distribution to the corresponding supremum of the NPII. By Billingsley (1968) this is the case if we manage to prove a certain "tightness" property of X_n.

From now on we consider the time interval $[0,1]$. Then, by theorem 15.6 in Billingsley (1968), it is enough to prove the following lemma:

Lemma 8.1. Assume $0 \leq t_1 \leq t \leq t_2$. Then there exists a constant k independent of t, t_1, t_2 and n such that

$$E\{ |X_n(t)-X_n(t_1)|^{\frac{4}{3}} |X_n(t_2)-X_n(t)|^{\frac{4}{3}}\} \leq k[\beta(t_2)-\beta(t_1)]^{\frac{4}{3}}$$

For the proof we need the following lemma:

Lemma 8.2.

There exists a constant k independent of n and $t \in [0,1]$ such that

$$\operatorname{var} X_n(t) \leq k\int_0^t \alpha_j(s)r(s)ds$$

The lemma is an immediate consequence of (3.3).

<u>Proof of lemma 8.1.</u> We have:

$$(8.7)\quad \begin{cases} f_n(t,t_1,t_2)=E\{|X_n(t)-X_n(t_1)|^{\frac{4}{3}}|X_n(t_2)-X_n(t)|^{\frac{4}{3}}\} \\ =E\{E[|X_n(t)-X_n(t_1)|^{\frac{4}{3}}|M_n(t)]E[|X_n(t_2)-X_n(t)|^{\frac{4}{3}}|M_n(t)]\} \end{cases}$$

By the Hölder inequality and lemma 8.2 we have for $M_n(t) > 0$:

$$(8.8)\quad \begin{cases} E[|X_n(t_2)-X_n(t)|^{\frac{4}{3}}|M_n(t)]\leq\{E[(X_n(t_2)-X_n(t))^2|M_n(t)]\}^{\frac{2}{3}} \\ \leq\{k_1nR_n(t)\int_t^{t_2}\alpha_j(s)r(s)ds\}^{\frac{2}{3}}\leq k_2\{nR_n(t)[\beta(t_2)-\beta(t_1)]\}^{\frac{2}{3}} \end{cases}$$

where k_1 and k_2 can be chosen independent of n,t,t_1 and t_2. By the Hölder inequality, lemma 3.2 (i) and (ii) and lemma 8.2 we have (using notation (1.12)):

$$E_t\{[nR_n(t)]^{\frac{2}{3}}|X_n(t)-X_n(t_1)|^{\frac{4}{3}}\}\leq\{E_t[n^2R_n^2(t)]\}^{\frac{1}{3}}\{E_t[X_n(t)-X_n(t_1)]^2\}^{\frac{2}{3}}$$

$$\leq k_3[Pr(M(t)>0)]^{-1}\{E[X_n(t)-X_n(t_1)]^2\}^{\frac{2}{3}}$$

$$\leq k_3[Pr(M(1)>0]^{-1}\{E[E(X_n(t)-X_n(t_1))^2|M_n(t_1)]\}^{\frac{2}{3}}$$

$$\leq k_4\{E_{t_1}[k_5nR_n(t_1)\int_{t_1}^{t}\alpha_j(s)r(s)ds]\}^{\frac{2}{3}}$$

$$\leq k_6\{E_{t_1}[nR_n(t_1)]\}^{\frac{2}{3}}[\beta(t_2)-\beta(t_1)]^{\frac{2}{3}}\leq k_7[\beta(t_2)-\beta(t_1)]^{\frac{2}{3}}$$

where the k's can be chosen independent of n,t,t_1 and t_2. Together with (8.7) and (8.8) this gives us lemma 8.1.

Q.E.D.

Let $X(t)$ be a normal process with independent increments and such that

$$EX(t) = 0 \ , \ \operatorname{var} X(t) = \int_0^t \alpha_j(s)r(s)ds \ .$$

Let X^* be this process with t restricted to $[0,1]$. Let X_n^* be our X_n-process with t restricted to $[0,1]$. By Billingsley (1968, theorem 15.6) we have proved the following theorem:

<u>Theorem 8.2.</u> X_n^* converges in distribution to X^*.

The following corollary follows by corollary 1 to theorem 5.1 in Billingsley (1968):

<u>Corollary 1.</u> $\sup\limits_{0\leq t\leq 1} |X_n(t)|$ converges in distribution to $\sup\limits_{0\leq t\leq 1} |X(t)|$.

We define:

$$\hat{\rho}_n = [n\hat{\sigma}_{j,n}(1)]^{-\frac{1}{2}} \ , \ f(t) = \int_0^t \alpha_j(s)r(s)ds \ , \ \rho = [f(1)]^{-\frac{1}{2}}$$

$$Y_n(t) = \hat{\rho}_n X_n(t) \ , \ Y(t) = \rho X(t) \ .$$

We will prove the following corollary to theorem 8.2.

<u>Corollary 2.</u> $\sup\limits_{0\leq t\leq 1} |Y_n(t)|$ converges in distribution to $\sup\limits_{0\leq t\leq 1} |Y(t)|$.

<u>Proof:</u> We have:

$$\sup_{0\leq t\leq 1} |Y_n(t)| = (\hat{\rho}_n-\rho) \sup_{0\leq t\leq 1} |X_n(t)| + \sup_{0\leq t\leq 1} |Y(t)| \ .$$

By theorem 6.1 and (3.3) $\hat{\rho}_n$ converges almost surely to ρ. Hence corollary 2 follows by a limit theorem due to Cramèr (1946, Chapter 20.6).

Q.E.D.

Obviously $\sup_{0\le t\le 1} |Y(t)|$ has the same distribution as the corresponding supremum in a Wiener process. By a well-known result (see e.g. Feller (1971), p. 343)

$$\Pr\left(\sup_{0\le t\le 1} |Y(t)|\le y\right) = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{2k+1}\exp[-\pi^2(2k+1)^2/8y^2] .$$

Together with corollary 2 this gives us a test of Kolmogorov-Smirnov type for the hypothesis:

$$\alpha_j(t) = \alpha_j^{(o)}(t) \quad \text{for all} \quad t \in [0,1]$$

for a specified function $\alpha_j^{(o)}(t)$.

9. Comparison with the "occurrence-exposure" rates.

In the first part of this section we will assume that all forces of transition are constant on the time interval $[0,1]$, i.e.:

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

where the α_i are positive numbers. We put:

$$\delta = \sum_{i=1}^{m} \alpha_i \ , \quad \delta_A = \sum_{i\in A} \alpha_i \ .$$

In this case the expression (3.13) for $\operatorname{var} \hat{P}_i(1;A)$ may be explicitely computed. Doing this we get:

$$\lim_{n\to\infty} n \operatorname{var} \hat{P}_i(1;A) = \frac{\alpha\delta_A(\delta-2\alpha)e^{\delta-2\delta_A} - \alpha\delta(\delta_A-\delta) - \alpha^2 e^{-2\delta_A}(\delta-2\delta_A)}{\delta\delta_A(\delta-2\delta_A)}$$

The maximum likelihood estimators, or the socalled "occurrence-exposure" rates (O.C.) for the α_i (see e.g. Sverdrup (1967)) are given by:

$$\alpha_i^* = \frac{M_i(1)}{T(1)}$$

where $T(t)$ is the sum of the times each process has been in state 0 until time t. By (1.2) we have:

$$P_i(1;A) = \frac{\alpha_i}{\delta_A}(1-e^{-\delta_A})$$

Hence it is natural to estimate $P_i(1;A)$ in the following way:

$$P_i^*(1;A) = \frac{\alpha_i^*}{\delta_A^*}(1-e^{-\delta_A^*})$$

where $\delta_A^* = \sum_{i\in A} \alpha_i^*$. By wellknown results (Sverdrup (1967)) the asymptotic variance of this estimator is by some computations found to be:

$$\operatorname{asvar} P_i^*(1;A) = \frac{1}{n}\,\frac{\alpha^2\delta_A^2\delta e^{-2\delta_A} + \alpha\delta(\delta_A-\delta)(1-e^{-\delta_A})^2}{(1-e^{-\delta})\delta_A^3}$$

Hence we have the following expression for the quotient between $\operatorname{asvar} \hat{P}_i(1;A)$ and $\operatorname{asvar} P_i^*(1;A)$:

$$\text{(9.1)} \qquad \frac{(1-e^{-\delta})\delta_A^2[\delta_A e^{\delta-2\delta_A}(\delta-2\alpha) - \delta(\delta_A-\alpha) - \alpha e^{-2\delta_A}(\delta-2\delta_A)]}{\delta(\delta-2\delta_A)[\alpha\delta_A^2\delta e^{-2\delta_A} + \delta(\delta_A-\alpha)(1-e^{-\delta_A})^2]}$$

We denote this expression by $f(\alpha,\delta_A,\delta)$.

The O.C. are known to have asymptotically least possible variance among a large class of estimators (Sverdrup (1965)). Hence it is interesting to study $f(\alpha, \delta_A, \delta)$.

I have not yet performed an extensive study of $f(\alpha, \delta_A, \delta)$, but I have computed it for certain special cases. The results are given below:

(i) $\alpha = \delta_A = \delta$. Then we have:

$$(9.2) \quad g(\delta) = f(\delta,\delta,\delta) = \delta^{-2} e^{\delta} (1-e^{-\delta})^2$$

This expression occurs in Sverdrup (1967, p. 61) in a study of the simple special case $m = 1$. From Sverdrup we take the following table:

δ	0.1	0.5	1.0	2.0
$g(\delta)$	1.0008	1.024	1.086	1.386

(ii) $\delta_A = \frac{2}{3}\delta$, $\alpha = \frac{1}{3}\delta$. We then have:

$$h(\delta) = f(\tfrac{1}{3}\delta, \tfrac{2}{3}\delta, \delta) = \frac{4(1-e^{-\delta})(3e^{\frac{4}{3}\delta} - 2e^{\delta} - 1)}{4\delta^2 + 9(e^{\frac{2}{3}\delta} - 1)^2}$$

The following table may be computed:

δ	0.1	0.5	1.0	2.0
$h(\delta)$	1.0005	1.009	1.035	1.097

(iii) $\delta_A = \delta = 1$. We then have:

$$f(\tfrac{1}{4}, 1, 1) = 1.009$$

$f(\frac{1}{2},1,1) = 1.02$

These examples should indicate that our nonparametric estimators are asymptotically almost equally good as the O.C. even if the forces of transition are constant.

To get a more complete picture of the situation one should also study the robustness of the O.C. when the forces of transition are not constant. In this case the O.C. are usually regarded as estimators of the average of the forces of transition over the time interval in question, i.e.:

$$(9.3) \qquad \frac{1}{t}\int_0^t \alpha_i(s)ds \qquad\qquad i = 1,\ldots,m$$

It is easily shown that the corresponding O.C. converges almost surely to

$$(9.4) \qquad (1-p_i(t))\left(\int_0^t p_i(s)ds\right)^{-1}$$

while (9.3) might be written:

$$(9.5) \quad -\frac{1}{t}\log(p_i(t))$$

(9.4) and (9.5) are of course not equal in general. Hence the O.C. will generally not be consistent when they are intended to estimate (9.3). In contrast to this we have shown in section 7 that $\frac{1}{t}\hat{\beta}_i(t)$ is a strongly consistent estimator of (9.3).

To make a closer comparison of (9.4) and (9.5) we look at the simple special case when

$$(9.6) \qquad \alpha_i(t) = \frac{1}{1+t}$$

Then (9.4) takes the form

$$(9.7) \quad \frac{t}{(1+t)\log(1+t)}$$

while (9.5) takes the form

$$(9.8) \quad \frac{1}{t}\log(1+t)$$

We now make the transormation $u = \beta_i(t) = \log(1+t)$ in (9.7) and (9.8) and then take the quotient of the resulting expressions ((9.7) as denominator). We then get:

$$(9.9) \quad u^{-2}e^{u}(1-e^{-u})^{2}$$

This turns out to be exactly the function g defined in (9.2). The table below (9.2) gives values of (9.9) for different values of u. One can say that (9.9) measures the degree of inconsistency of the O.C. when it is regarded as an estimator of (9.3) and when the assumption (9.6) is made. Even if the assumption (9.6) represents a relatively large deviation from constancy, the degree of inconsistency of the O.C. is seen to be small for $\delta \leq 1$. In practice we will probably nearly always have $\delta \leq 1$ because this means that the probability of leaving state 0 is less than $1-e^{-1} = 0.632$.

We might perhaps say that the results in this section seem to support a conclusion of the following kind: For the nonparametric estimators to be considerably better than the O.C., the deviation of the forces of transition from constancy will have to be quite large. On the other hand, even if the forces of transition are constant, the nonparametric estimators are almost equally good as the O.C.

Of course we must take the reservation that we have not yet made any thorough investigation of the relationship between our nonparametric estimators and the O.C.

10. Appendix.

Proof of Lemma 3.2: For $x \geq 1$ we have:

$$(10.1) \quad \frac{1}{x} = \frac{1}{x+1} + \frac{1}{x(x+1)} , \quad \frac{1}{x(x+1)} \leq \frac{3}{(x+1)(x+2)}$$

$$(10.2) \quad \begin{cases} \dfrac{1}{x^2} = \dfrac{1}{(x+1)(x+2)} + \dfrac{3x+2}{x^2(x+1)(x+2)} \\[2ex] \dfrac{3x+2}{x^2(x+1)(x+2)} \leq \dfrac{24}{(x+1)(x+2)(x+3)} \end{cases}$$

We put $a = [1-(1-p)^n]^{-1}$, $q = 1-p$, $q_o = 1-p_o$. It is easily seen that

$$(10.3) \quad a \leq \frac{1}{p_o}$$

Direct computations give:

$$(10.4) \quad E^*(\frac{1}{X+1}) = \frac{a}{p(n+1)}[1-q^{n+1}-(n+1)pq^n]$$

$$(10.5) \quad E^*(\frac{1}{(X+1)(X+2)}) = \frac{a}{p^2(n+1)(n+2)}[1-q^{n+2}-(n+2)pq^{n+1} -\tfrac{1}{2}(n+1)(n+2)p^2q^n]$$

$$(10.6) \quad E^*(\frac{1}{(X+1)(X+2)(X+3)}) \leq \frac{a}{p^3(n+1)(n+2)(n+3)} \leq \frac{1}{p_o{}^4 n^3}$$

Proof of (i): By (10.1), (10.3), (10.4) and (10.5) we have:

$$|E^*(\frac{n}{X})- \frac{1}{p}| \leq |E^*(\frac{n}{X+1})- \frac{1}{p}|+3E^*(\frac{n}{(X+1)(X+2)}) \leq \frac{a}{p(n+1)}|-1+q^n-n^2pq^n|+\frac{3a}{np^2}$$

$$\leq \frac{1+q_o^n+n^2q_o^n}{np_o^2} + \frac{3}{np_o^3}$$

so that:

$$n|E^*(\frac{n}{X})- \frac{1}{p}| \leq \frac{1+q_o^n+n^2q_o^n}{p_o^2} + \frac{3}{p_o^3}$$

where the right side is obviously bounded with respect to n since $p_o > 0$ and $q_o < 1$. Hence the first part of (i) is proved. The second part is easily proved in the same way by using (10.2), (10.3), (10.5) and (10.6).

Prof of (ii): By Hölders inequality we have:

$$E^*|\frac{n}{X} - \frac{1}{p}| \leq [E^*(\frac{n}{X} - \frac{1}{p})^2]^{\frac{1}{2}} = [E^*(\frac{n^2}{X^2} - \frac{1}{p^2}) - \frac{2}{p}E^*(\frac{n}{X} - \frac{1}{p})]^{\frac{1}{2}}$$

Hence (ii) follows from (i).

Proof of (iii): This is a direct consequence of the first part of (i).

Q.E.D.

Proof of Lemma 3.4: We define for $n = 1,2,\dots$:

$$\varphi_n(x) = f_n(x)\exp\left[\int_0^x g_n(y)dy\right] - \int_0^x h(y)\exp\left[\int_0^y g_n(t)dt\right]dy$$

We have:

$$\varphi_n'(x) = \exp\left[\int_0^x g_n(y)dy\right]\left[f_n'(x)+g_n(x)f_n(x)-h(x)\right]$$

By the assumptions of the lemma we have when $n \to \infty$:

$$\varphi_n'(x) \to 0 \quad \text{uniformly on} \quad [0,a]$$

We also have $\varphi_n(0) = c$ for all n . By theorem 7.17 in Rudin (1964) we conclude that $\varphi_n(x) \to c$ uniformly on $[0,a]$. Hence the conclusion of the lemma.

Q.E.D.

Proof of Lemma 5.1. Assume $0 \leq x \leq 1$. Obviously we have:

$$(10.7) \quad \begin{cases} \left|\int_0^x g_n(y)dF_n(y) - \int_0^x g_o(y)dF_o(y)\right| \\ \leq \int_0^1 |g_n(y)-g_o(y)|dF_n(y) + \left|\int_0^x g_o(y)dF_n(y) - \int_0^x g_o(y)dF_o(y)\right| \end{cases}$$

Let $\epsilon > 0$ be given. Since $g_n(y)-g_o(y)$ converges uniformly to 0 on $[0,1]$, there exists a N_1 such that

$$(10.8) \quad n \geq N_1 \Rightarrow \int_0^1 |g_n(y)-g_o(y)|dF_n(y) \leq \epsilon$$

Since $g_o(y)$ is uniformly continuous on each of a finite set of subintervals of $[0,1]$, there exists a left-continuous step-function $s(y)$; $0 \leq s(y) \leq a$; with "jumping points" $x_1 < x_2 < \ldots < x_k$ such that

$$|g_o(y)-s(y)| \leq \epsilon \qquad \forall y \in [0,1] .$$

Let $m(x)$ denote the number of $x_i \leq x$, $h(x)$ the greatest $x_i \leq x$ and put $x_o = 0$. Then:

$$|\int_0^x g_o(y)dF_n(y)-\int_0^x g_o(y)dF_o(y)| \leq 2\epsilon+|\int_0^x s(y)dF_n(y)-\int_0^x s(y)dF_o(y)|$$

$$= 2\epsilon+|\sum_{i=0}^{m(x)-1} s(x_i)[F_n(x_{i+1})-F_n(x_i)]+s(x)[F_n(x)-F_n(h(x))]$$

$$- \sum_{i=0}^{m(x)-1} s(x_i)[F_o(x_{i+1})-F_o(x_i)]-s(x)[F_o(x)-F_o(h(x))]|$$

$$\leq 2\epsilon+a\sum_{i=0}^{k-1}[|F_n(x_{i+1})-F_o(x_{i+1})|+|F_n(x_i)-F_o(x_i)|]+a|F_n(x)-F_o(x)|$$

$$+ a|F_n(h(x))-F_o(h(x))|$$

Since $F_n(y)$ converges to $F_o(y)$ uniformly on $[0,1]$, there exists a N_2 such that for all $y \in [0,1]$:

$$n \geq N_2 \Rightarrow |F_n(y)-F_o(y)| \leq \frac{\epsilon}{2k+2}$$

Hence:

$$n \geq N_2 \Rightarrow |\int_0^x g_o(y)dF_n(y)-\int_0^x g_o(y)dF_o(y)| \leq 3\epsilon$$

for all $x \in [0,1]$. Together with (10.7) and (10.8) this proves the lemma.

Q.E.D.

Proof of Lemma 5.2: Choose ϵ such that $0 < \epsilon < \frac{1}{2}a$. Then there exists a N such that

$$n \geq N \Rightarrow |f_n(x)-f(x)| \leq \epsilon \qquad \text{for all} \quad x \in [0,1] .$$

We then have for $n \geq N$ and all $x \in [0,1]$:

$$|f_n(x)^{-1}-f(x)^{-1}| = \frac{|f_n(x)-f(x)|}{f_n(x)f(x)} \leq \frac{\epsilon}{a(a-\epsilon)} \leq 2\epsilon a^{-2} .$$

Hence the first statement in the lemma is proved. The second is proved in the same way. For the proof of the third we use the well known and easily established fact that if $|x| \leq \frac{1}{2}$ then $\log(1+x) = x+\theta(x)x^2$ where $|\theta(x)| \leq 1$. Hence for n large enough we have for all $x \in [0,1]$:

$$|n\log[1-(nf_n(x)+1)^{-1}]+f(x)| \leq |f(x)-n[nf_n(x)+1]^{-1}| + n[nf_n(x)+1]^{-2}$$

$$\leq \frac{\epsilon+\frac{1}{n}}{a(a-\epsilon)} + \frac{n}{(an+1)^2} \leq \frac{3\epsilon}{a^2} .$$

Q.E.D.

11. Further work. Generalizations.

It should be quite clear that the theory developed so far is in many ways quite incomplete. The approximations and inequalities in the moment formulae in section 3 may surely be improved. In section 5 it ought to be possible to prove strong convergence of $\hat{P}_{i,n}(t;A)$ uniformly with respect to $t \geq 0$. This would be a further generalization of the Glivenko-Cantelli theorem. The estimators in section 6 should of course be more thoroughly studied. It should also be possible to prove asymptotic normality of $\hat{P}_{i,n}(t;A)$ in general, and likewise to prove results of the kind in section 8 for these estimators. The possibility of non-asymptotic Kolmogorov-Smirnov type testing should also be studied. Lastly the comparison with the occurrence exposure rates and the study of the expression (9.1) should be done in much grater detail. The author will continue the work in all these areas.

One will also see that the simplicity of the multiple decrement model is in no way essential for the theory. It is probably quite possible to generalize the results to Markov chains of a much more general kind. This will also be the object of further work by the author.

We have used an especially simple observational scheme in this paper, in that we have assumed that all individual processes are observed during the same given time interval. It will be seen that this is not either essential for the theory.

12. Acknowledgement.

I would like to thank Jan M. Hoem who supervised my work with an earlier version of a part of this theory (consisting mainly of theorem 5.1). Mr. Hoem gave valuable comments and corrected several mistakes.

13. References.

AALEN, O. (1972). Estimering av risikorater for prevensjons-midlet "spiralen". (In norwegian.) Graduate thesis in statistics. Institute of Mathematics, University of Oslo.

BARLOW, R.E. (1968). Some recent developments in reliability theory. Amsterdam Mathematisch Centrum, Mathematical Centre Tracts. 27 49-65.

BARLOW, R.E. and van ZWET, W.R. (1969). Asymptotic properties of isotonic estimators for the generalized failure rate function. Parts I and II. University of California, Berkeley, Operations Research Center Reports, No. ORC 69-5 and 69-10.

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

CRAMÉR, H. (1946). Mathematical Methods of Statistics. Princeton University Press.

CHUNG, K.L. (1968). A Course in Probability theory. Harcourt, Brace and Word, Inc. New York.

FELLER, W. (1957). An Introduction to Probability Theory and Its Applications, vol. I. Second Edn. Wiley, New York.

FELLER, W. (1971). An Introduction to Probability Theory and Its Applications, vol. II. Second Edn. Wiley, New York.

GRENANDER, U. (1956). On the theory of mortality measurement. Skand. Aktuarietidskrift. 39 70-96 and 125-153.

HOEM, J.M. (1969). Purged and Partial Markov Chains. Skand. Aktuarietidskrift. 52 147-155.

HOEM, J.M. (1972). On the Statistical Theory of Analytic Graduation. Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability. University of California Press.

KAPLAN, E.L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. American Statistical Association Journal. 53.

LEHMANN, E.L. (1959). _Testing Statistical Hypotheses_. Wiley, New York.

RUDIN, W. (1964). _Principles of Mathematical Analysis_. McGraw-Hill, New York.

SIMONSEN, W. (1966). _Forsikringsmatematik_. Hefte I. Københavns Universitets Fond til Tilvejebringelse af Læremidler.

SVERDRUP, E. (1965). Estimates and Test Procedures in connection with Stochastic Models for Deaths, Recoveries and Transfers between different States of Health. _Skand. Aktuarietidskrift_. _48_ 184-211.

SVERDRUP, E. (1966). The Present State of the Decision Theory and the Neyman-Pearson Theory. _Review of the International Statistical Institute_. _34_.

SVERDRUP, E. (1967). _Statistiske metoder ved dødelighetsundersøkelser_. Annen utgave. Mimeographed. Institute of Mathematics, University of Oslo.

ZWINGGI, E. (1945). _Versicherungsmathematik_. Verlag Birkhäuser, Basel.