Z. Wahrscheinlichkeitstheorie verw. Gebiete
53, 97–116 (1980)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Nonparametric Estimation Based on Censored Observations of a Markov Renewal Process

R.D. Gill

Mathematical Centre, Kruislaan 413, Amsterdam, The Netherlands

Summary. Uniform consistency and weak convergence is proved of estimators of the transition probabilities of an arbitrary finite state space Markov renewal process, based on n independent and identically distributed "right censored" realizations of the process. The approach uses the theory of stochastic integrals and counting processes. It is shown how the results may be extended to the non-identically distributed case and to general censorship under suitable conditions.

1. Introduction

A Markov renewal process can often be used to describe the lengths of time spent in consecutive stages (not necessarily following a fixed order) of a disease, or in the functioning of a machine. In such a situation one might want to estimate the distribution functions (so called *transition probabilities*) which describe the probabilistic behaviour of this phenomenon. Lagakos et al. [9] (see this paper and its references for applications) proposed certain estimators on the basis of maximum likelihood considerations by maximizing the probability of n realizations of the process observed on fixed finite time intervals over all discrete transition probabilities; they also derived approximate variances and covariances of the resulting estimators by looking at the Fisher information, again assuming discrete distributions. The present paper gives rigorous derivations of consistency and weak convergence of these estimators as n, the number of (partial) realizations of the process, tends to infinity. We make *no* assumptions on the transition probabilities and also allow quite general types of *censoring*: that is to say, the mechanism through which only incomplete realizations of the Markov renewal process are available.

Let $J_0, J_1, \ldots$ be the consecutive states of a Markov renewal process and let $X_1, X_2, \ldots$ be the sojourn times in these states (we take our notation from Pyke [13] apart from two changes which we mention shortly). So $J_0, J_1, \ldots$ are r.v.'s taking values in the finite set of states $\{1, \ldots, m\}$ for some $m \in \mathbb{N}$ and $X_1, X_2, \ldots$

0044-3719/80/0053/0097/$04.00

are r.v.'s taking values in $(0, \infty]$ with the interpretation that the process starts off at time $S_0=0$ in state J_0, and at time $S_n=\sum_{i=1}^{n} X_i$, after a stay of length X_n in state J_{n-1}, jumps to state J_n (at least, if $S_n<\infty$). The joint distribution of J_0, X_1, J_1, X_2, $J_2, \ldots$ is built up recursively from an initial state distribution

$$\alpha_k=P(J_0=k) \qquad k \leqq m \tag{1}$$

and transition probabilities (which we want to estimate)

$$Q_{ij}(t)=P(X_n \leqq t, J_n=j \mid J_{n-1}=i), \qquad i,j \leqq m, \ t \in [0,\infty], \ n \in \mathbb{N}, \tag{2}$$

satisfying $Q_{ij}(0)=0$ by supposing that for each $n \in \mathbb{N}$

$$P(X_n \leqq t, J_n=j \mid J_0, X_1, J_1, \ldots, X_{n-1}, J_{n-1})=P(X_n \leqq t, J_n=j \mid J_{n-1}) \quad \text{a.s.} \tag{3}$$

We shall also be interested in estimating the (possibly defective) distribution functions H_i, defined by

$$H_i(t)=P(X_n \leqq t \mid J_{n-1}=i)=\sum_j Q_{ij}(t). \tag{4}$$

Rather than working with the r.v.'s $J_0, X_1, J_1, \ldots$ we step over to *counting processes* $\tilde{N}_{ij}$ (Aalen [1]) which register the transitions made from state i to state j up to time t:

$$\tilde{N}_{ij}(t)=\#\{n \geqq 1: S_n \leqq t, J_{n-1}=i, J_n=j\} \qquad i,j \leqq m, \ t \in [0,\infty). \tag{5}$$

Because the state space is finite, by Pyke [12, Lemma 4.1] these processes have sample paths which (almost surely) are finite for all t, zero at time zero, integer valued and right continuous with jumps of size $+1$ only, these jumps furthermore not occurring simultaneously in different processes. Define also

$$\begin{aligned}
\tilde{N}_{i.}(t)&=\sum_j \tilde{N}_{ij}(t) \\
\tilde{N}_{.j}(t)&=\sum_i \tilde{N}_{ij}(t) \qquad \text{(denoted } N_j(t) \text{ in Pyke [12])} \\
\tilde{N}(t)&=\sum_{i,j} \tilde{N}_{ij}(t) \qquad \text{(denoted } N(t) \text{ in Pyke [12])}
\end{aligned} \tag{6}$$

and define

$$\begin{aligned}
Z(t)&=J_{\tilde{N}(t)} \\
L(t)&=t-S_{\tilde{N}(t-)}.
\end{aligned} \tag{7}$$

Pyke [12] calls the process Z, the state occupied at time $t+$, a *semi-Markov process*, and the multivariate process $\{\tilde{N}_{.j}; j \leqq m\}$ a *Markov renewal process.* $L(t)$ is a left continuous version of what is called the *backward recurrence time:* it is the length of time which at time $t-$ has elapsed since the last jump of the Markov renewal process. It plays an important role in the sequel.

We shall later use the fact that $\tilde{N}(t)$ has finite moments of all orders. For, as m is finite, there exists a distribution function F such that $F(0)=0$ and

$Q_{ij}(t) \leqq F(t)$ for all i, j and t. So $\tilde{N}(t)$ is stochastically dominated by the evaluation at time t of a renewal process having recurrence time distribution F; and for this process the result is well known.

The rest of this paper is organized as follows. We shall conclude this section with some more notation and some facts from the theory of stochastic integrals and counting processes for use in Sect. 2. There we formally define our censoring model, by introducing a process K taking the values 0 and 1, with the interpretation that the Markov renewal process is "under observation" at time t if and only if $K(t)=1$. We assume that K is *predictable* which loosely speaking means that $K(t)$ does not depend on the development of the Markov renewal process in $[t, \infty)$. It is easy to construct counter examples where K does not have this property and the estimators considered here are inconsistent. Some further technical assumptions are made, such as a restriction to "right censorship". Then we derive some important equalities involving the first and second moments of processes counting sojourn times in the various states "observed" to be $\leqq t$ and to be followed by a transition to a particular state, or just "observed" to be $\geqq t$. In Sect. 3 we give definitions of the estimators of Lagakos et al. [9] based on n independent and identically distributed censored observations of the Markov renewal process. The estimators depend on the data through the processes counting observed sojourn times. We now prove three theorems: on consistency of the estimators with arbitrary Q_{ij}'s, on asymptotic normality with discrete Q_{ij}'s, and on weak convergence with arbitrary Q_{ij}'s. Theorem 2 with its easy proof (once Theorem 1 has been established) is included to illustrate Theorem 3, without the latter's technicalities (which are themselves largely postponed to Sect. 5 in the form of some lemmas). Section 4 discusses these results and shows how the assumptions of right censorship and identical distributions may be relaxed.

Notation. For a real valued stochastic process $Y=\{Y(t); t\in[0,\infty)\}$ whose sample paths have left hand limits, Y_- is the process defined by

$$Y_-(0)=0$$

and

$$Y_-(t)=Y(t-).$$

In dealing with an indexed family of processes, we write e.g. Y_{i-} for $(Y_i)_-$. We define Y_+ similarly under the obvious conditions. ΔY is the process $Y-Y_-$ and $\mathscr{E}Y$ the function $t\to\mathscr{E}Y(t)$. By $\int X\,dY$ we denote (with a single exception in Theorem 3) the process with sample paths $t\to \int_{[0,t]} X(s)\,dY(s)$, the latter being interpreted as a pathwise Lebesgue-Stieltjes integral, such an interpretation being possible in all the cases considered. We write $\int^- X\,dY$ for the process $(\int X\,dY)_-$ with sample paths $t\to \int_{[0,t)} X(s)\,dY(s)$. Some miscellaneous notations are χ_A for the indicator function of the set A and $\|\cdot\|_\tau$ for the supremum norm on $[0,\tau]$, also written $\|\cdot\|$ if there can be no confusion. Unless otherwise stated, time variables s, t, τ, etc. are always in $[0,\infty)$ and state variables i,j in $\{1,\dots,m\}$. The σ-algebra generated by a family of r.v.'s is denoted by $\sigma\{\cdot\}$, while to indicate that generated by a union of σ-algebras we use the symbol $\vee$. Maximum and minimum are denoted by $\vee$ and $\wedge$ respectively.

Stochastic Integrals. We collect together here a few results from Meyer [11]. Let $(\Omega, \mathscr{F}, P)$ be a complete probability space and let $\{\mathscr{F}_t; t\in[0,\infty)\}$ be an increasing right continuous family of sub σ-algebras of Ω such that $\mathscr{F}_0$ contains all P-null sets of $\mathscr{F}$. A process X is *predictable* if as a function of (t,ω) it is measurable with respect to the σ-algebra on $[0,\infty)\times\Omega$ generated by the left continuous adapted processes. A process X possesses a property *locally* if there exists a so called *localizing* sequence of stopping times $\{T_n\}$, $T_n\uparrow\infty$, such that for each n the process $t\to\chi_{\{T_n>0\}}X(t\wedge T_n)$ has this property.

Martingales are always supposed to have right continuous paths with left hand limits. Let M and N be local square integrable martingales and let H and K be predictable and locally bounded. There exists a unique predictable process $\langle M,N\rangle$ with paths of locally integrable variation such that $MN-\langle M,N\rangle$ is a local martingale, zero at time zero. A process $H\circ M$ is uniquely defined by requiring that

(i) it should be a local square integrable martingale, and

(ii) $\langle H\circ M, H\circ M\rangle=\int H^2\,d\langle M,M\rangle$.

Almost everywhere where the path-wise Lebesgue-Stieltjes integral $\int H\,dM$ is well defined, the paths of $H\circ M$ and $\int H\,dM$ coincide. Furthermore $\langle H\circ M, K\circ N\rangle=\int HK\,d\langle M,N\rangle$. We actually make most use of the simple corollaries of these results, that if the processes $\int H\,dM$ and $\int K\,dN$ exist, if all the localizing stopping times above can be taken as constants, and if $M(0)=N(0)=0$, then the following equalities (between functions on $[0,\infty)$) hold:

$$\mathscr{E}(\int H\,dM)=\mathscr{E}(\int K\,dN)=0, \tag{8}$$

$$\mathscr{E}(\int H\,dM\int K\,dN)=\mathscr{E}(\int HK\,d\langle M,N\rangle). \tag{9}$$

If the word "local" can be dropped above, then the same relationships hold on $[0,\infty]$.

Counting Processes. Let $\{N_i; i=1,\dots,k\}$ be an indexed family of processes with right continuous paths, which are zero at time zero, nondecreasing and integer valued, and have jumps of size $+1$ only, no two processes jumping at the same time. Suppose that for all t,

$$\mathscr{F}_t=\mathscr{F}_0\vee\sigma\{N_i(s); i=1,\dots,k, s\leqq t\}.$$

Define $T_0=0$ and $T_n=\inf\{t\colon \sum_i N_i(t)\geqq n\}$, $n=1,2,\dots$, i.e. T_n is the time of the n-th jump of $\{N_i;\ i=1,\dots,k\}$. For $n\geqq 1$ let J_n be the index of the particular process which if T_n is finite jumps at time T_n, i.e. $N_{J_n}(T_n)=N_{J_n}(T_n-)+1$ if $T_n<\infty$. Define $F_{ni}(t)$ to be a regular version of $P(T_n\leqq t$ and $J_n=i\,|\,\mathscr{F}_{T_{n-1}})$, $t\in[0,\infty)$, $i=1,\dots,k$, and $n=1,2,\dots$, and

$$A_i(t)=\sum_{n=1}^{\infty}\int_0^t \frac{dF_{ni}(s)}{1-\sum_{i'=1}^{k}F_{ni'}(s-)}. \tag{10}$$

Jacod [8] shows, for $i=1, \ldots, k$, that $M_i = N_i - A_i$ is a local martingale, zero at time zero, while by Gill [6, Propositions 12 and 13], or Liptser and Shiryaev [10, Lemma 18.12], the M_i's are local square integrable martingales with

$$\langle M_i, M_j \rangle = \begin{cases} \int (1-\Delta A_i)\, dA_i & i=j \\ -\int \Delta A_i\, dA_j & i \neq j. \end{cases} \tag{11}$$

The localizing stopping times may be taken as T_n, but if $\mathscr{E} N_i(t) < \infty$ for all i and t also as constants.

2. Random Right Censorship of a Markov Renewal Process

We suppose that it is not possible to observe complete realizations of J_0, X_1, $J_1, \ldots$ but rather that observations consist of the values taken by the processes Z_-, L and $\Delta \tilde{N}_{ij}$ $(i, j \leq m)$ on some random subset of the time axis. Letting K be the indicator process of this set, then we say that the (Markov renewal) process is under observation at time t if $K(t)=1$, when the state at $t-$, how long it has been occupied, and a possible transition at that time instant and the new state can be observed; otherwise $K(t)=0$. Let $(\Omega, \mathscr{F}, P)$ be a complete probability space on which the process K and the r.v.'s J_0, X_1, $J_1, \ldots$ are defined. Then the following assumptions about K are made, and with the exception of A3* are taken to be in force throughout the rest of the paper, unless explicitly stated otherwise:

A1. There exist r.v.'s T_n such that almost surely,

$$S_n \leq T_n \leq S_{n+1}\ \forall n \quad \text{and} \quad K(t) = \sum_n \chi_{(S_n, T_n]}(t)\ \forall t.$$

A2. There exists a sub σ-algebra $\mathscr{A}$ of $\mathscr{F}$ containing all P-null sets of $\mathscr{F}$, conditionally independent of $\sigma\{\tilde{N}_{ij}(s); i, j \leq m, s \in [0, \infty)\}$ given J_0, and such that for each n T_n (see A1) is a stopping time with respect to the family of σ-algebras $\{\mathscr{F}_t; t \in [0, \infty)\}$ defined by

$$\mathscr{F}_t = \mathscr{A} \vee \sigma\{J_0, \tilde{N}_{ij}(s); i, j \leq m, s \in [0, t]\}.$$

A3. $\mathscr{E}(\#\{n: T_n > S_n\}) < \infty$.

A3*. $\mathscr{E}((\#\{n: T_n > S_n\})^{7+\varepsilon}) < \infty$ for some $\varepsilon > 0$.

The first assumption limits us to what might be called "right-censorship". A2 implies that the process K is *predictable* and might be interpreted as stating that $K(t)$ does not depend on the development of the Markov renewal process in $[t, \infty)$ given the past at time t. According to A3 or A3* the number of at least partially observed sojourn times is certainly almost surely finite. A3 is sufficient for proving consistency of our estimators, and weak convergence when the X_i's take on values in $\mathbb{N}$; for general Q_{ij} we need A3* in proving weak convergence. It ought to be possible to weaken A3* by using perhaps a different method of proof: however in practical applications A3 and A3* will generally hold. For suppose that for some fixed finite k and t observation always stops after at most k transitions after the time instant t; i.e. $s > S_{\tilde{N}(t)+k}$

$\Rightarrow K(s)=0$. In this case $\#\{n: T_n > S_n\} \leqq \tilde{N}(t)+k$ which has finite moments of all orders.

We mention briefly some special cases. If $K(t)=\chi_{[0,T]}(t)$ for some r.v. T which is conditionally independent of $X_1, J_1, \ldots$ given J_0, then A1 and A2 hold. If $m=1$, and $K(t)=\chi_{[0,U_r]}(L(t))$ for $t \in (S_{r-1}, S_r]$, $r=1,\ldots,n$ and zero on (S_n, ∞) for a fixed $n \geqq 1$ and positive r.v.'s $U_1,\ldots,U_n$ independent of $X_1, X_2, \ldots$ then all the assumptions hold and we have in fact the usual model of n censored observations of $X_1,\ldots,X_n$ with censoring variables $U_1,\ldots,U_n$.

Next we introduce processes which count observed sojourn times: for $u \geqq 0$ we define

$$\begin{aligned} N_{ij}(u) &= \#\{n \geqq 1: J_{n-1}=i, J_n=j, X_n \leqq u, K(S_n)=1\} \\ &= \text{number of sojourn times in state } i \text{ observed to take on a} \\ &\quad\ \text{value } \leqq u \text{ and to be followed by a jump to state } j \\ &= \#\{t: \Delta \tilde{N}_{ij}(t)=1,\ Z(t-)=i,\ L(t) \leqq u,\ K(t)=1\}, \\ N_i(u) &= \sum_j N_{ij}(u). \end{aligned} \tag{12}$$

For $u>0$ we define

$$\begin{aligned} Y_i(u) &= \#\{n \geqq 1: J_{n-1}=i, X_n \geqq u, K(S_{n-1}+u)=1\} \\ &= \text{number of sojourn times in state } i \text{ observed to take on a value } \geqq u \\ &= \#\{t: Z(t-)=i, L(t)=u, K(t)=1\}, \\ Y_i(0) &= Y_i(0+), \\ Y(u) &= \sum_i Y_i(u) \qquad \text{(also for } u=0). \end{aligned} \tag{13}$$

Note that $Y(0)=\#\{n: T_n > S_n\}$.

By A1 and A3, $\mathscr{E} Y_i(u)$ and $\mathscr{E} N_{ij}(u) \leqq \mathscr{E} Y(0) < \infty$ for all i, j and u. The sample paths of the processes N_{ij} are almost surely zero at time zero, right continuous with left hand limits, nondecreasing and integer valued; while the sample paths of Y_i are nonincreasing, left continuous with right hand limits, and nonnegative integer valued. According to the interpretation of K, these processes are observable.

By the counting process results of Sect. 1 applied to $\{\tilde{N}_{ij}; i,j \leqq m\}$, if we define

$$\begin{aligned} A_{ij}(t) &= \int_0^t \chi_{\{i\}}(Z(s-)) \frac{dQ_{ij}(L(s))}{1-H_i(L(s)-)} \\ &= \sum_{r=1}^{\tilde{N}(t-)} \chi_{\{i\}}(J_{r-1}) \int_0^{X_r} \frac{dQ_{ij}(u)}{1-H_i(u-)} + \chi_{\{i\}}(J_{\tilde{N}(t-)}) \int_0^{L(t)} \frac{dQ_{ij}(u)}{1-H_i(u-)} \end{aligned} \tag{14}$$

and

$$M_{ij} = \tilde{N}_{ij} - A_{ij}, \tag{15}$$

it follows that with respect to $\{\mathscr{F}_t\}$ (defined in A2) the M_{ij}'s are local square integrable martingales, zero at time zero, with

$$\langle M_{ij}, M_{ij}\rangle = \int (1-\Delta A_{ij})\, dA_{ij}$$
$$\langle M_{ij}, M_{ij'}\rangle = -\int \Delta A_{ij}\, dA_{ij'}, \quad j \neq j'$$
$$\langle M_{ij}, M_{i'j'}\rangle = 0 \qquad i \neq i'. \tag{16}$$

Here the localizing stopping times may be taken as constants by finiteness of $\mathscr{E}\tilde{N}_{ij}(t)$ for all i, j and t.

Now by A3, it follows that for all i and j

$$\mathscr{E}\int_0^\infty K(s)\, d\tilde{N}_{ij}(s) \leqq \mathscr{E}\, Y(0) < \infty. \tag{17}$$

But because K is predictable, nonnegative and bounded, by the stochastic integral results of Sect. 1,

$$\mathscr{E}\int_0^t K(s)\, d\tilde{N}_{ij}(s) = \mathscr{E}\int_0^t K(s)\, dA_{ij}(s) \qquad \text{for all } t < \infty. \tag{18}$$

So letting $t \to \infty$,

$$\mathscr{E}\int_0^\infty K(s)\, dA_{ij}(s) < \infty. \tag{19}$$

Next, since

$$\begin{aligned}\mathscr{E}\left(\int_0^t K(s)\, dM_{ij}(s)\right)^2 &= \mathscr{E}\int_0^t K(s)^2\, d\langle M_{ij}, M_{ij}\rangle(s)\\ &= \mathscr{E}\int_0^t K(s)(1-\Delta A_{ij}(s))\, dA_{ij}(s),\end{aligned} \tag{20}$$

by letting $t \uparrow \infty$, we find

$$\lim_{t\uparrow\infty} \mathscr{E}\left(\int_0^t K(s)\, dM_{ij}(s)\right)^2 < \infty. \tag{21}$$

But since $\int K\, dM_{ij}$ is a martingale w.r.t. $\{\mathscr{F}_t\}$ this shows that $\int K\, dM_{ij}$ is in fact a (zero mean) square integrable martingale, clearly with

$$\langle \int K\, dM_{ij}, \int K\, dM_{i'j'}\rangle = \int K\, d\langle M_{ij}, M_{i'j'}\rangle. \tag{22}$$

Next, as in Gill [6, Lemma 3 and Proposition 4] (where similar results are obtained in a rather special case), we find that for any bounded measurable functions f and f' on $[0, \infty)$

$$\int_0^\infty f(u)\, dN_{ij}(u) = \int_0^\infty f(L(t))\, d\tilde{N}_{ij}(t), \tag{23}$$

$$\int_0^\infty f(u)\, Y_i(u) \frac{dQ_{ij}(u)}{1-H_i(u-)} = \int_0^\infty f(L(t))\, K(t)\, dA_{ij}(t), \tag{24}$$

and so defining

$$\begin{aligned} Z_{ij}(v) &= N_{ij}(v) - \int_0^v Y_i(u) \frac{dQ_{ij}(u)}{1-H_i(u-)}\\ Z_i(v) &= \sum_j Z_{ij}(v),\end{aligned} \tag{25}$$

we see that

$$\int_0^\infty f(L(s))\,K(s)\,dM_{ij}(s)=\int_0^\infty f(u)\,dZ_{ij}(u), \tag{26}$$

where $f(L)$ is a bounded predictable process (L is left continuous with right hand limits and adapted). So by the theory of stochastic integrals, taking expectations in (26),

$$0=\mathscr{E}\int_0^\infty f(u)\,dZ_{ij}(u)=\int_0^\infty f(u)\,d\mathscr{E}N_{ij}(u)-\int_0^\infty f(u)\,\mathscr{E}\,Y_i(u)\frac{dQ_{ij}(u)}{1-H_i(u-)} \tag{27}$$

and similarly (by (9) and (26))

$$\begin{aligned}
&\mathscr{E}\left(\left(\int_0^\infty f(u)\,dZ_{ij}(u)\right)\left(\int_0^\infty f'(u)\,dZ_{i'j'}(u)\right)\right)\\
&=\mathscr{E}\int_0^\infty f(L(s))\,f'(L(s))\,K(s)\,d\langle M_{ij}, M_{i'j'}(s)\rangle\\
&=\begin{cases}\displaystyle\int_0^\infty f(u)\,f'(u)\,\mathscr{E}\,Y_i(u)\left(1-\frac{\Delta Q_{ij}(u)}{1-H_i(u-)}\right)\frac{dQ_{ij}(u)}{1-H_i(u-)}, & i=i',\ j=j'\\[2ex]
\displaystyle -\int_0^\infty f(u)\,f'(u)\,\mathscr{E}\,Y_i(u)\frac{\Delta Q_{ij}(u)}{1-H_i(u-)}\,\frac{dQ_{ij'}(u)}{1-H_i(u-)}, & i=i',\ j\neq j'\\[2ex]
0, & i\neq i'\end{cases}
\end{aligned} \tag{28}$$

(where the last step follows by similar equalities to (24)).

3. Estimation of Transition Probabilities Q_{ij}

Suppose we are given n independent identically distributed observations of N_{ij} and Y_i. Let N_{ij}^n, Y_i^n, N_i^n, Z_{ij}^n, Z_i^n denote the sums of the n realizations of N_{ij}, Y_i, ... (see definitions (12), (13) and (25)). We build up estimators of Q_{ij} in two steps, first estimating H_i by $\hat{H}_i^n$, defined by

$$\hat{H}_i^n(t)=1-\prod_{s\leq t}\left(1-\frac{\Delta N_i^n(s)}{Y_i^n(s)}\right), \tag{29}$$

where by convention (which we adopt from now on) $0/0=0$. An equivalent implicit definition is

$$\hat{H}_i^n(t)=\int_0^t (1-\hat{H}_i^n(s-))\frac{dN_i^n(s)}{Y_i^n(s)}. \tag{30}$$

Now we can define $\hat{Q}_{ij}^n$ by

$$\hat{Q}_{ij}^n(t)=\int_0^t (1-\hat{H}_i^n(s-))\frac{dN_{ij}^n(s)}{Y_i^n(s)}. \tag{31}$$

Adding over j we see that $\hat{H}_i^n = \sum_j \hat{Q}_{ij}^n$ (cf. (4)). These definitions can be motivated by the following facts. Putting in (27) $f = \chi_{[0,t]}$ we find

$$\mathscr{E} N_{ij}(t) = \int_0^t \mathscr{E} Y_i(s)(1 - H_i(s-))^{-1} \, dQ_{ij}(s)$$

which implies, adding over j, that

$$\mathscr{E} N_i(t) = \int_0^t \mathscr{E} Y_i(s)(1 - H_i(s-))^{-1} \, dH_i(s).$$

So if t satisfies $\mathscr{E} Y_i(t) > 0$, which implies that $\mathscr{E} Y_i(s) > 0$ and $1 - H_i(s-) > 0$ on $[0, t]$, we find

$$H_i(t) = \int_0^t (1 - H_i(s-)) \frac{d\mathscr{E} N_i(s)}{\mathscr{E} Y_i(s)} \tag{32}$$

(cf. (30)). We shall see in Lemma 1 that such an expression defines H_i uniquely. Similarly, also supposing $\mathscr{E} Y_i(t) > 0$, we find

$$Q_{ij}(t) = \int_0^t (1 - H_i(s-)) \frac{d\mathscr{E} N_{ij}(s)}{\mathscr{E} Y_i(s)} \tag{33}$$

(cf. (31)). Of course we can't expect to be able to estimate Q_{ij} outside of $\{t \colon \mathscr{E} Y_i(t) > 0\}$.

As a referee remarked, one might also be interested in the analogue of the well known "Nelson plot" (see e.g. Aalen [1, Sect. 6.1]), namely $\int_0^t Y_i^n(s)^{-1} \, dN_{ij}^n(s)$, which estimates the function $\int_0^t (1 - H_i(s-))^{-1} \, dQ_{ij}(s)$. Its properties can be derived in exactly the same way as those of $\hat{Q}_{ij}$ and $\hat{H}_i$ (in fact many of the arguments become easier). However we shall not go into the details here.

We first prove consistency, for which we need two lemmas.

Lemma 1. *Let A and B be right continuous nondecreasing functions on $[0, \infty)$, zero at time zero; suppose $\Delta A \leqq 1$ and $\Delta B < 1$ on $[0, \infty)$. Then the unique solution Z of*

$$Z(t) = 1 - \int_0^t \frac{Z(s-)}{1 - \Delta B(s)} (dA(s) - dB(s)) \tag{34}$$

which is locally bounded (i.e. bounded on $[0, t]$ for each $t < \infty$) and right continuous with left hand limits is given by

$$Z(t) = \frac{\prod_{s \leqq t} (1 - \Delta A(s)) \cdot \exp(-A_c(t))}{\prod_{s \leqq t} (1 - \Delta B(s)) \cdot \exp(-B_c(t))},$$

where $A_c(t) = A(t) - \sum_{s \leqq t} \Delta A(s)$, etc.

Proof. The proof given by Liptser and Shiryaev [10, Lemma 18.8] for the case $B=0$ goes through exactly, replacing their a with $(1-\Delta B)^{-1}$, A with $A-B$, $Z(0)$ with 1, and α with $\int(1-\Delta B)^{-1}(dA+dB)$. □

Corollary. *For all t such that* $1-H_i(t)>0$

$$1-H_i(t)=\prod_{s\leq t}\left(1-\frac{\Delta H_i(s)}{1-H_i(s-)}\right)\exp\left(-\int_0^t\frac{dH_{ic}(s)}{1-H_i(s-)}\right) \tag{36}$$

and

$$\frac{1-\hat{H}_i^n(t)}{1-H_i(t)}=1-\int_0^t\frac{1-\hat{H}_i^n(s-)}{1-H_i(s-)}\left(1-\frac{\Delta H_i(s)}{1-H_i(s-)}\right)^{-1}\left(\frac{dN_i^n(s)}{Y_i^n(s)}-\frac{dH_i(s)}{1-H_i(s-)}\right). \tag{37}$$

Proof. Since (34) holds with $B\equiv 0$, $A=\int(1-H_{i-})^{-1}dH_i$ and $Z=1-H_i$ we can substitute these quantities in (35) which gives (36). Thus by (22) and (29), (35) holds with $A=\int(Y_i^n)^{-1}dN_i^n$, $B=\int(1-H_{i-})^{-1}dH_i$, and $Z=(1-\hat{H}_i^n)/(1-H_i)$; substituting them in (34) gives (37). □

Note that if $Y_i^n(t)>0$, if $1-H_i(t-)>0$ and if $1-H_i(t)=0$ then almost surely $\hat{H}_i^n(t)=H_i(t)=1$; also note that $1-\Delta H_i/(1-H_{i-})=(1-H_i)/(1-H_{i-})$. So we can rewrite (37) as

$$\hat{H}_i^n-H_i=(1-H_i)\int\frac{1-\hat{H}_{i-}^n}{1-H_i}\frac{n}{Y_i^n}\frac{dZ_i^n}{n} \tag{38}$$

on $\{t\colon Y_i^n(t)>0$ and $1-H_i(t-)>0\}$, where Z_i was defined in (25) and the convention $0/0=0$ may have to be invoked. Next we write, also on $\{t\colon Y_i^n(t)>0$ and $1-H_i(t-)>0\}$

$$\begin{aligned}\hat{Q}_{ij}^n-Q_{ij}&=\int(1-\hat{H}_{i-}^n)\left(\frac{dN_{ij}^n}{Y_i^n}-\frac{dQ_{ij}}{1-H_{i-}}\right)+\int\left(\frac{1-\hat{H}_{i-}^n}{1-H_{i-}}-1\right)dQ_{ij}\\&=\int(1-\hat{H}_{i-}^n)\frac{n}{Y_i^n}\frac{dZ_{ij}^n}{n}-\int\left(\int^-\frac{1-\hat{H}_{i-}^n}{1-H_i}\frac{n}{Y_i^n}\frac{dZ_i^n}{n}\right)dQ_{ij}\\&=\int(1-\hat{H}_{i-}^n)\frac{n}{Y_i^n}\frac{dZ_{ij}^n}{n}-Q_{ij}\int^-\frac{1-\hat{H}_{i-}^n}{1-H_i}\frac{n}{Y_i^n}\frac{dZ_i^n}{n}\\&\quad+\int^-Q_{ij}\frac{1-\hat{H}_{i-}^n}{1-H_i}\frac{n}{Y_i^n}\frac{dZ_i^n}{n},\end{aligned} \tag{39}$$

where we have used (31), (37) and integration by parts. (Adding over j and integrating by parts gives (38) again of course!) □

The next lemma and its corollary give conditions under which expressions such as the right hand sides of (38) and (39) converge uniformly in t to zero, in probability: this will be the consistency result of Theorem 1.

Lemma 2. *Suppose H_n and Z_n are processes on $[0,\tau]$ whose sample paths are almost surely right continuous with left hand limits, of bounded variation, and satisfying $Z_n(0)=0$; $n=1,2,\ldots$. Suppose $\|Z_n\|=\sup_{t\in[0,\tau]}|Z_n(t)|\to_P 0$. If*

$\int_{(0,\tau)} |dH_n(t)|$ *is bounded in probability as* $n\to\infty$, *then* $\|\int H_{n-}\, dZ_n\| \to_P 0$ *and* $\|\int^- H_n dZ_n\| \to_P 0$; *if* $\int_{(0,\tau]} |dH_n(t)|$ *is bounded in probability, then* $\|\int H_n dZ_n\| \to_P 0$.

Proof. We can write

$$(\int H_{n-}\, dZ_n)(t) = \int_{(0,t)} (Z_n(t) - Z_n(s))\, dH_n(s)$$

and

$$(\int H_n dZ_n)(t) = \int_{(0,t]} (Z_n(t) - Z_n(s-))\, dH_n(s).$$

So $\|\int H_{n-}\, dZ_n\|$ and $\|\int^- H_n dZ_n\| \leqq 2\|Z_n\| \int_{(0,\tau)} |dH_n(t)| \to_P 0$ as $n\to\infty$ and similarly $\|\int H_n dZ_n\| \to_P 0$. ∎

Corollary. *Suppose Lemma* 2 *allows us to conclude that* $\|\int H_n^{(i)} dZ_n\| \to_P 0$, $i = 1, \ldots, r$, *and* $\|\int H_{n-}^{(i)}\, dZ_n\| \to_P 0$, $i = r+1, \ldots, s$. *Then*

$$\left\|\int \prod_{i=1}^{r} H_n^{(i)} \cdot \prod_{i=r+1}^{s} H_{n-}^{(i)}\, dZ_n\right\| \to_P 0,$$

as $n\to\infty$.

Proof. Apply Lemma 2 first to $H_n^{(1)}$ and Z_n, then to $H_n^{(2)}$ and $\int H_n^{(1)} dZ_n$, etc. ∎

Theorem 1. *Let* $\tau_i = \sup\{t\colon\ \mathscr{E}\, Y_i(t) > 0\}$. *Then as* $n\to\infty$

$$\sup_{t\in[0,\tau_i]} |\hat{H}_i^n(t) - H_i(t)| \to_P 0$$

and

$$\sup_{t\in[0,\tau_i]} |\hat{Q}_{ij}^n(t) - Q_{ij}(t)| \to_P 0,$$

unless $\mathscr{E}\, Y_i(\tau_i) = 0$ *and* $\Delta H_i(\tau_i)$ *or* $\Delta Q_{ij}(\tau_i) > 0$, *in which case* $[0, \tau_i]$ *must be replaced by* $[0, \tau_i)$ *in the corresponding supremum.*

Proof. By the weak law of large numbers and monotonicity arguments it is easy to show that $\|n^{-1} N_{ij}^n - \mathscr{E} N_{ij}\|_\infty \to_P 0$ and

$$\|\int n^{-1} Y_i^n (1 - H_{i-})^{-1} dQ_{ij} - \int \mathscr{E}\, Y_i (1 - H_{i-})^{-1} dQ_{ij}\|_\infty \to_P 0.$$

So $\|n^{-1} Z_{ij}^n\|_\infty \to_P 0$ (using (27) with $f = \chi_{[0,v]}$). Suppose $\tau \leqq \tau_i$ is such that $\mathscr{E}(Y_i(\tau)) > 0$. Then $\int_{(0,\tau)} |d(n/Y_{i+}^n(t)| \leqq n/Y_i^n(\tau)$, bounded in probability as $n\to\infty$, and so by the corollary to Lemma 2 and (39) $\|\hat{Q}_{ij}^n - Q_{ij}\|_\tau \to_P 0$ as $n\to\infty$; adding over j shows $\|\hat{H}_i^n - H_i\|_\tau \to_P 0$. If $\mathscr{E}\, Y_i(\tau_i) > 0$ we are ready. Otherwise monotonicity arguments show that the required results hold with $[0, \tau_i)$ in place of $[0, \tau_i]$; if $\Delta H_i(\tau_i) = 0$ or $\Delta Q_{ij}(\tau_i) = 0$ then adding τ_i to the range of the supremum in the corresponding term changes nothing. ∎

Actually almost sure convergence in the appropriate set up is also easy to derive using the strong law of large numbers and Glivenko-Cantelli type arguments, and a suitable modification of Lemma 2.

Now we turn to proving weak convergence, giving first a result for the case that Q_{ij}'s give weight only to the positive integers. Again we use the representations (38) and (39), and also need Theorem 1.

Theorem 2. *Suppose* $X_1, X_2, \ldots$ *take values in* $\mathbb{N}$. *Let* $\tau_i = \sup\{t \in \mathbb{N}: \mathscr{E} Y_i(t) > 0\}$ $(\tau_i \in \mathbb{N} \cup \{\infty\})$. *Then* $\{n^{\frac{1}{2}}(\hat{Q}^n_{ij}(t) - Q_{ij}(t)), n^{\frac{1}{2}}(\hat{H}^n_i(t) - H_i(t)); t \in \mathbb{N}, t \leqq \tau_i, i, j \leqq m\}$ *is distributed asymptotically as*

$$\left\{\sum_{s=1}^{t} \frac{1-H_i(s-)}{\mathscr{E} Y_i(s)} U_{ij}(s) - \sum_{s=1}^{t-1} \frac{(Q_{ij}(t)-Q_{ij}(s))}{\mathscr{E} Y_i(s)} \frac{1-H_i(s-)}{1-H_i(s)} U_i(s),\right.$$

$$\left.(1-H_i(t)) \sum_{s=1}^{t} \frac{1-H_i(s-)}{1-H_i(s)} \frac{1}{\mathscr{E} Y_i(s)} U_i(s);\ t \in \mathbb{N},\ t \leqq \tau_i,\ i, j \leqq m\right\},$$

where the $U_{ij}(s)$*'s are multivariate normally distributed r.v.'s with expectations zero;* $U_i(s) = \sum_j U_{ij}(s)$; *and*

$$\operatorname{var}(U_{ij}(s)) = \mathscr{E} Y_i(s) \left(1 - \frac{\Delta Q_{ij}(s)}{1-H_i(s-)}\right) \frac{\Delta Q_{ij}(s)}{1-H_i(s-)},$$

$$\operatorname{cov}(U_{ij}(s), U_{ij'}(s)) = -\mathscr{E} Y_i(s) \frac{\Delta Q_{ij}(s)}{1-H_i(s-)} \frac{\Delta Q_{ij'}(s)}{1-H_i(s-)} \qquad j \neq j',$$

$$\operatorname{cov}(U_{ij}(s), U_{i'j'}(s')) = 0 \qquad i \neq i' \text{ or } s \neq s'.$$

Proof. Multiply (38) and (39) by $n^{\frac{1}{2}}$ and rewrite the integrals in the right hand sides as sums over $s \leqq t$ or $s < t$ as appropriate, with $n^{-\frac{1}{2}} dZ^n_{ij}(s)$ replaced with $n^{-\frac{1}{2}} \Delta Z^n_{ij}(s)$. By Theorem 1 and convergence in probability of $n^{-1} Y^n_i(s)$ the coefficients of $n^{-\frac{1}{2}} \Delta Z^n_{ij}(s)$ all converge in probability, while by the central limit theorem and (27) and (28) with $f = \chi_{\{s\}}$, $f' = \chi_{\{s'\}}$,

$$\{n^{-\frac{1}{2}} \Delta Z^n_{ij}(s);\ i, j \leqq m,\ s \in \mathbb{N}\} \to_{\mathscr{D}} \{U_{ij}(s);\ i, j \leqq m,\ s \in \mathbb{N}\}$$

and the theorem is proved. □

Apart from the fact that the final theorem on weak convergence with arbitrary Q_{ij} uses A3*, it includes Theorem 2 as a special case. The method of proof is essentially the same, though many more technical details are encountered.

Theorem 3. *Suppose* A3* *holds, and choose* τ_i, $i \leqq m$, *such that* $\mathscr{E} Y_i(\tau_i) > 0$. *Then considered as a random element of* $\prod_i (D[0, \tau_i])^{m+1}$ *(see Billingsley [3])* $\{n^{\frac{1}{2}}(\hat{Q}_{ij} - Q_{ij}), n^{\frac{1}{2}}(\hat{H}_i - H_i); i, j \leqq m\}$ *is asymptotically distributed as*

$$\left\{\int \frac{1-H_{i-}}{\mathscr{E} Y_i} dW_{ij} - Q_{ij} \int^{-} \frac{1-H_{i-}}{1-H_i} \frac{1}{\mathscr{E} Y_i} dW_i + \int^{-} Q_{ij} \frac{1-H_{i-}}{1-H_i} \frac{1}{\mathscr{E} Y_i} dW_i,\right.$$

$$\left.(1-H_i) \int \frac{1-H_{i-}}{1-H_i} \frac{1}{\mathscr{E} Y_i} dW_i;\ i, j \leqq m\right\},$$

where the W_{ij}'s are jointly zero mean Gaussian processes with independent multivariate increments, the sets $\{W_{ij};\ j \leq m\}$, $i=1, \ldots, m$, being independent of one another; $W_i = \sum_j W_{ij}$; and

$$\operatorname{var}(W_{ij}(t)) = \int_0^t \mathscr{E}\, Y_i(s) \left(1 - \frac{\Delta Q_{ij}(s)}{1-H_i(s-)}\right) \frac{dQ_{ij}(s)}{1-H_i(s-)},$$

$$\operatorname{cov}(W_{ij}(t), W_{ij'}(t)) = -\int_0^t \mathscr{E}\, Y_i(s) \frac{\Delta Q_{ij}(s)}{1-H_i(s-)} \frac{dQ_{ij'}(s)}{1-H_i(s-)}.$$

The integrals with respect to W_{ij} and W_i are stochastic integrals in the sense of Meyer [11] (the W_{ij}'s are square integrable martingales with respect to the natural family of σ-algebras) or can equivalently be defined by formal integration by parts, the resulting expressions having a pathwise definition.

Proof. We only sketch the proof here; details are given as three lemmas in Sect. 5. The proof is again based on the representations (38) and (39). Multiplying these equations by $n^{\frac{1}{2}}$, we note that the expressions on the right hand sides consist of integrals, where the integrands are products of fixed functions and the processes $1-\hat{H}^n_{i-}$ and n/Y^n_i, which, in probability, converge uniformly on $[0, \tau_i]$ to $1-H_{i-}$ and $(\mathscr{E}\, Y_i)^{-1}$. The integrals are taken with respect to $n^{-\frac{1}{2}} Z^n_{ij}$ and $n^{-\frac{1}{2}} Z^n_i = \sum_j n^{-\frac{1}{2}} Z^n_{ij}$. Now these processes have finite dimensional distributions which converge to those of W_{ij} and W_i by using the central limit theorem and (27) and (28) with $f = \chi_{[0,t]}$ and $f' = \chi_{[0,t']}$. In Lemma 3 in Sect. 5 we prove tightness of $n^{-\frac{1}{2}} Z^n_{ij}$ in $D[0, \tau_i]$, here A3* is used. Actually we need to prove a little more, because in the next step of the proof we want to apply the Skorohod-Dudley theorem (see e.g. Pyke [13]) and consider processes $(\hat{H}^n_i - H_i)'$, $(n^{-1} Y^n_i - \mathscr{E}\, Y_i)'$, and $(n^{-\frac{1}{2}} Z^n_{ij})'$ $(i, j \leq m,\ n=1, 2, \ldots)$ defined on a new probability space with the same joint distribution for each n as their unprimed equivalents, and converging almost surely in the supremum norm to 0, 0, and W'_{ij} respectively (the W'_{ij}'s also having the same joint distribution as the W_{ij}'s). The construction is possible with the supremum norm distance rather than the Skorohod distance, if the sample paths of each W_{ij} can be taken to be continuous with probability one; i.e. if the Q_{ij}'s are continuous. However, we can get round this problem by inserting a time interval at each jump point t of Q_{ij}, joining up $n^{-\frac{1}{2}} Z^n_{ij}(t-)$ to $n^{-\frac{1}{2}} Z^n_{ij}(t)$ with a straight line across this interval, and proving joint weak convergence of these new processes on the resulting extended time interval to the corresponding objects obtained from the W_{ij}'s, which can be taken to be continuous. Of course we shall need weak convergence of the finite dimensional distributions of $n^{-\frac{1}{2}} Z^n_{ij-}$ and $n^{-\frac{1}{2}} Z^n_{ij}$; which again follows from the central limit theorem and (27) and (28). In Lemma 4 we show that the above programme can indeed be carried out. Now we are at liberty to apply the Skorohod-Dudley theorem (after which we remove the extra intervals again). Finally the corollary to Lemma 5 shows that the suprema over $[0, \tau]$ of the absolute difference between the primed versions of $n^{\frac{1}{2}}$ times (38) and (39) and their "obvious" limits (given in the statement of the theorem above) converge almost surely to zero, where some care is needed because these obvious limits cannot be defined directly as pathwise integrals if the Q_{ij}'s have continuous components. This problem is also resolved in Lemma 5. □

4. Remarks and Generalizations

Though the limiting covariances of $n^{\frac{1}{2}}(\hat{Q}^n_{ij}-Q_{ij})$ can be consistently estimated, it seems difficult to use them to construct confidence bands. At least this is not the case for $n^{\frac{1}{2}}(\hat{H}^n_i-H_i)$; e.g. Gill [6 or 7]. An advantage of the Nelson plot mentioned earlier (immediately before Lemma 1) is that confidence bands for them are easy to construct, because the corresponding asymptotic process will have independent increments; i.e. we will obtain a deterministically time transformed Brownian motion. Also Theorem 3 does not give us in general a limiting distribution of the estimators of the transition probabilities $Q_{ij}(\infty)$ of the Markov chain $J_0, J_1, \ldots$ associated with the Markov renewal process, supposing that estimation should be possible (Theorem 1 gives conditions for consistent estimation).

Theorem 3 specialized to the case $m=1$ and assuming a rather special form for the censoring process K gives us a weak convergence theorem for the so-called product limit estimator. This result generalizes that obtained by Breslow and Crowley [4], for which continuous distributions of both censoring variables and the variables of interest were assumed. In Gill [7] we derive this result in a more direct fashion and in more generality.

We next discuss possible generalizations of our results. Firstly, can we drop the restriction A1 to right censorship? In the discrete case this gives no problems: we must assume that $\mathscr{E}\, Y_i(s)<\infty$ for all i and $s=1, 2, \ldots$ and realize that we can only estimate $\Delta Q_{ij}(s)/(1-H_i(s-))$ for s such that $\mathscr{E}\, Y_i(s)>0$ (where Y_i is still defined by (13)). However in the general case we have made strong use of many of the properties of the processes Y_i as the following list of "corrections" shows: replace A1 with the *assumption* that the sample paths of Y_i are left continuous with right hand limits; in A2 instead of the stopping time condition assume that K is predictable with respect to the given σ-algebras; and in A3 and A3* replace $\#\{n\colon T_n>S_n\}$ with $\#\{n\colon K(s)=1$ for some $s\in(S_n, S_{n+1}]\}$. Theorem 1 then remains true if we replace τ_i with $\tau_i=\sup\{t\colon \mathscr{E}\int_{(0,t)}|d Y_{i+}(s)|<\infty$ and $\mathscr{E}\min_{s\in(0,t]} Y_i(s)>0\}$ and modify the condition "unless $\mathscr{E}\, Y_i(\tau_i)=0$" accordingly. Theorem 3 remains valid if we choose τ_i such that $\mathscr{E}\int_{(0,\tau_i)}|d Y_{i+}(t)|<\infty$, $\mathscr{E}\min_{s\in(0,\tau_i]} Y_i(s)>0$, $\|n^{-1} Y_i^n-\mathscr{E}\, Y_i\|_{\tau_i}\to_P 0$ and $\int_{(0,\tau_i)}|d\mathscr{E}\, Y_{i+}(t)|<\infty$, where actually the last two properties are a consequence of the first one.

Alternatively, what happens when we drop the assumption of identically distributed observations of $\{N_{ij}, Y_i\}$? Consider a triangular array $\{N^{kn}_{ij}, Y^{kn}_i; i,j\leqq m\}$, $k=1, \ldots, n$; $n=1, 2, \ldots$, for each (k, n) defined as in Sects. 1 and 2 with fixed Q_{ij}'s, but possibly differing initial distributions and "censoring distributions", and independent over k for each n. Define

$$N_{ij}^{n}=\sum_{k=1}^{n} N_{ij}^{kn}$$

and

$$Y_i^{n}=\sum_{k=1}^{n} Y_i^{kn}$$

and proceed as in Sect. 3. It is not too difficult to see that if we replace A3 with the assumption that for some fixed C, $\mathscr{E}\, Y^{kn}(0)<C<\infty$ for all k and n and similarly modify A3*, and *define* $\mathscr{E}\, Y_i$ to be $\lim_{n\to\infty} n^{-1}\mathscr{E}\, Y_i^n$, which we assume to exist and to be left continuous with right hand limits, then Theorems 1 to 3 still hold. Important for applications to medical trials is the fact that even if for some states i the transition probabilities vary with n and k, for the other states all our results go through.

Finally, instead of looking at n independent observations, suppose that a single Markov renewal process is, with censoring, observed over an expanding sequence of time intervals. As a specific example, consider the model of Sects. 1 and 2 where we drop assumptions A1 and A3 but do suppose that the X_i's take values in $\mathbb{N}$. In A2 we assume that K is predictable instead of making the stopping times assumption. Defining

$$Y_i^n(u)=\#\{t\leqq n\colon Z(t-)=i,\ L(t)=u,\ K(t)=1\}$$

and

$$\Delta N_{ij}^n(u)=\#\{t\leqq n\colon \Delta \tilde{N}_{ij}(t)=1,\ Z(t-)=i,\ L(t)=u,\ K(t)=1\}$$

(cf. (12) and (13)) it is possible, as in Bather [2, Lemma 1], to apply Chow [5, Theorem 5] to show that on the set of ω for which $Y_i^n(u)\to\infty$ as $n\to\infty$ (for fixed i and u), $\Delta N_{ij}^n(u)/Y_i^n(u)\to\Delta Q_{ij}(u)/(1-H_i(u-))$ a.s. It is not yet clear to the author what can be done for general Q_{ij}'s, nor indeed in this special case how weak convergence can be proved.

It should be pointed out that we could not apply the general theory of Aalen [1] to derive our results, despite the strong similarity of models. This is because of the occurrence of the process L in formula (14), whose saw-tooth paths prevent the transformation (26) from M_{ij} to Z_{ij} (defined in (15) and (25)) from preserving the martingale property of M_{ij}. However, it preserves enough of it (the properties of first and second moments) for our asymptotic results.

5. Technical Lemmas Needed for Proof of Theorem 3

Lemma 3. *For each i, j and τ such that $\mathscr{E}\, Y_i(\tau)>0$, $n^{-\frac{1}{2}}Z_{ij}^n$ is tight in $D[0,\tau]$ as $n\to\infty$ if* A3* *holds.*

Proof. Let I_1 and I_2 be the intervals $(t_1, t]$ and $(t, t_2]$ for some time instants $0\leqq t_1<t<t_2\leqq\tau$. Write $\Delta_k X$ for $\int_{I_k} dX$, $k=1$ or 2, for a process or function X of

bounded variation. We show that we can find $C>0$ and $\alpha\in(\frac{1}{2},1)$ such that

$$n^{-2}\mathscr{E}(\Delta_1 Z^n_{ij}\cdot\Delta_2 Z^n_{ij})^2\leqq C(\Delta_1 Q_{ij}\cdot\Delta_2 Q_{ij})^\alpha, \tag{40}$$

which proves tightness in view of Billingsley [3, Theorem 15.6 and the remarks on p. 133]. Now $(\Delta_1 Z^n_{ij}, \Delta_2 Z^n_{ij})$ is a sum of n independent random variables with zero means (by (27) with $f=\chi_{I_k}$) each distributed as $(\Delta_1 Z_{ij}, \Delta_2 Z_{ij})$. So

$$\begin{aligned} n^{-2}\mathscr{E}(\Delta_1 Z^n_{ij}\cdot\Delta_2 Z^n_{ij})^2 = n^{-1}&\mathscr{E}(\Delta_1 Z_{ij}\cdot\Delta_2 Z_{ij})^2\\ &+((n-1)/n)\,\mathscr{E}(\Delta_1 Z_{ij})^2\,\mathscr{E}(\Delta_2 Z_{ij})^2\\ &+2((n-1)/n)(\mathscr{E}(\Delta_1 Z_{ij}\cdot\Delta_2 Z_{ij}))^2. \end{aligned}\tag{41}$$

Replacing f with χ_{I_1} and f' with χ_{I_2} in (28) shows that

$$\mathscr{E}(\Delta_1 Z_{ij}\cdot\Delta_2 Z_{ij})=0, \tag{42}$$

while replacing f and f' by χ_{I_k} shows

$$\begin{aligned}\mathscr{E}(\Delta_k Z_{ij})^2 &= \int_{I_k}\mathscr{E}\,Y_i\left(1-\frac{\Delta Q_{ij}}{1-H_{i-}}\right)\frac{dQ_{ij}}{1-H_{i-}}\\ &\leqq C\Delta_k Q_{ij} \qquad \text{because } (1-H_{i-})^{-1} \text{ is bounded on } [0,\tau]\\ &\leqq C(\Delta_k Q_{ij})^\alpha \qquad \text{for any } \alpha\in(0,1). \end{aligned}\tag{43}$$

Note that in the sequel the constant C may be different on each appearance; however, it can always be chosen not to depend on I_1 and I_2 though it will often depend on α. Substituting (42) and (43) back in (41), we see that to establish (40) it remains to suitably bound $\mathscr{E}(\Delta_1 Z_{ij}\cdot\Delta_2 Z_{ij})^2$. Now since $\Delta_k Z_{ij} =\Delta_k N_{ij}-\int_{I_k} Y_i(1-H_{i-})^{-1}\,dQ_{ij}$,

$$\begin{aligned}|\Delta_1 Z_{ij}\cdot\Delta_2 Z_{ij}| &\leqq \Delta_1 N_{ij}\cdot\Delta_2 N_{ij}+\Big(\int_{I_1} Y_i(1-H_{i-})^{-1}dQ_{ij}\Big)\Big(\int_{I_2} Y_i(1-H_{i-})^{-1}dQ_{ij}\Big)\\ &\quad+\Delta_1 N_{ij}\Big(\int_{I_2} Y_i(1-H_{i-})^{-1}dQ_{ij}\Big)+\Delta_2 N_{ij}\Big(\int_{I_1} Y_i(1-H_{i-})^{-1}dQ_{ij}\Big)\\ &\leqq C\,Y(0)^2, \end{aligned}\tag{44}$$

while expanding (42) we find that for $\alpha_1=1-\beta_1\in(0,1)$

$$\begin{aligned}&\mathscr{E}(\Delta_1 N_{ij}\cdot\Delta_2 N_{ij})+\mathscr{E}\Big(\Big(\int_{I_1} Y_i(1-H_{i-})^{-1}\,dQ_{ij}\Big)\cdot\Big(\int_{I_2} Y_i(1-H_{i-})^{-1}\,dQ_{ij}\Big)\Big)\\ &\quad=\mathscr{E}\Big(\Delta_1 N_{ij}\cdot\Big(\int_{I_2} Y_i(1-H_{i-})^{-1}\,dQ_{ij}\Big)\Big)+\mathscr{E}\Big(\Delta_2 N_{ij}\cdot\Big(\int_{I_1} Y_i(1-H_{i-})^{-1}\,dQ_{ij}\Big)\Big)\\ &\quad\leqq C\cdot\mathscr{E}(\Delta_1 N_{ij}\cdot Y(0))\cdot\Delta_2 Q_{ij}+C\cdot\mathscr{E}(\Delta_2 N_{ij}\cdot Y(0))\cdot\Delta_1 Q_{ij}\\ &\quad\leqq C\cdot[\mathscr{E}((\Delta_1 N_{ij})^{\alpha_1}\cdot Y(0)^{1+\beta_1})\cdot(\Delta_2 Q_{ij})^{\alpha_1}\\ &\qquad+\mathscr{E}((\Delta_2 N_{ij})^{\alpha_1}\cdot Y(0)^{1+\beta_1})\cdot(\Delta_1 Q_{ij})^{\alpha_1}]\\ &\quad\leqq C[(\mathscr{E}(\Delta_1 N_{ij}))^{\alpha_1}(\mathscr{E}\,Y(0)^{1+1/\beta_1})^{\beta_1}(\Delta_2 Q_{ij})^{\alpha_1}\\ &\qquad+(\mathscr{E}(\Delta_2 N_{ij}))^{\alpha_1}(\mathscr{E}\,Y(0)^{1+1/\beta_1})^{\beta_1}(\Delta_1 Q_{ij})^{\alpha_1}]\\ &\quad\leqq C(\Delta_1 Q_{ij}\cdot\Delta_2 Q_{ij})^{\alpha_1} \end{aligned}\tag{45}$$

if

$$\mathscr{E}(Y(0))^{1+1/\beta_1}) < \infty, \tag{46}$$

where we have used Hölder's inequality and the fact that

$$\mathscr{E}\Delta_k N_{ij} = \int_{I_k} \mathscr{E} Y_i (1-H_{i-})^{-1} dQ_{ij}.$$

Therefore for $\alpha_2 = 1-\beta_2 \in (0,1)$,

$$\begin{aligned}\mathscr{E}(\Delta_1 Z_{ij}\cdot \Delta_2 Z_{ij})^2 &= \mathscr{E}(|\Delta_1 Z_{ij}\cdot \Delta_2 Z_{ij}|^{\alpha_2}\cdot|\Delta_1 Z_{ij}\cdot \Delta_2 Z_{ij}|^{1+\beta_2})\\ &\leqq C\cdot\mathscr{E}(|\Delta_1 Z_{ij}\cdot \Delta_2 Z_{ij}|^{\alpha_2} Y(0)^{2+2\beta_2}) \quad \text{(by (44))}\\ &\leqq C(\mathscr{E}|\Delta_1 Z_{ij}\cdot \Delta_2 Z_{ij}|)^{\alpha_2}(\mathscr{E}\, Y(0)^{2+2/\beta_2})^{\beta_2}\end{aligned} \tag{47}$$

(by Hölder's inequality)

$$\leqq C(\Delta_1 Q_{ij}\cdot \Delta_2 Q_{ij})^{\alpha_1\alpha_2}$$

by (44) and (45) if (46) holds and if

$$\mathscr{E}(Y(0)^{2+2/\beta_2}) < \infty. \tag{48}$$

Now by A3* we can choose $\beta_1 < \frac{1}{6}$ such that (46) holds and $\beta_2 < \frac{2}{5}$ such that (48) holds. For such a choice $\alpha_1\alpha_2 = (1-\beta_1)(1-\beta_2) > \frac{5}{6}\cdot\frac{3}{5} = \frac{1}{2}$ and therefore, by (47), for sufficiently small $\alpha > \frac{1}{2}$ we can find a $C < \infty$ such that

$$\mathscr{E}(\Delta_1 Z_{ij}\cdot \Delta_2 Z_{ij})^2 \leqq C(\Delta_1 Q_{ij}\cdot \Delta_2 Q_{ij})^{\alpha} \quad \text{for all } t_1, t, t_2, \tag{49}$$

which completes the proof. □

We have now shown that for τ_i satisfying $\mathscr{E}Y_i(\tau_i) > 0$,

$$\{n^{-\frac{1}{2}} Z_{ij}^n; i,j \leqq m\} \to_{\mathscr{D}} \{W_{ij}; i,j \leqq m\} \quad \text{on } \prod_{i=1}^{m} D[0,\tau_i]^m$$

where the limit has been defined in the statement of Theorem 3, while jointly $\hat{H}_i^n - H_i \to_{\mathscr{D}} 0$ and $\hat{Q}_{ij}^n - Q_{ij} \to_{\mathscr{D}} 0$ each on $D[0,\tau_i]$. The next lemma on the Skorohod-Dudley construction is only stated and proved for a single process $n^{-\frac{1}{2}} Z_{ij}^n$, but the required simultaneous result can obviously be proved in the same way.

Lemma 4. *A Skorohod-Dudley construction is possible for* $n^{-\frac{1}{2}} Z_{ij}^n$, $n = 1, 2, \ldots$ *with respect to the supremum norm on* $D[0,\tau_i]$ *where* τ_i *satisfies* $\mathscr{E}Y_i(\tau_i) > 0$.

Proof. Write $W_n = n^{-\frac{1}{2}} Z_{ij}^n$, $n = 1, 2, \ldots$; $W = W_{ij}$; and $\tau = \tau_i$. Let $t_1, t_2, \ldots$ be an enumeration of the jump points of Q_{ij} in $[0,\tau]$ and let $\delta_k = \Delta Q_{ij}(t_k) \geqq 0$ for all k, $\sum_k \delta_k \leqq 1$. Define $v(t) = t + \sum_{t_k \leqq t} \delta_k$ and $\delta(t) = v(t) - v(t-) = \delta_k$ if $t = t_k$ for some $k = 1, 2, \ldots$ and $= 0$ otherwise. By the comments above, $W_n \to_{\mathscr{D}} W$ as $n \to \infty$ in $D[0,\tau]$ where W is a zero mean Gaussian process with independent increments and variance function $\operatorname{var}(W(t)) = \int_0^t f(s)\, dQ_{ij}(s) = A(t)$ say, for some bounded non-negative measurable function f on $[0,\tau]$. Let W^* be the zero mean

Gaussian process on $[0, v(\tau)]$ with independent increments and almost surely continuous sample paths, such that $\operatorname{var}(W^*(u)) = A(t-) + f(t)(u - v(t-))$ where t is the unique solution of $v(t-) \leqq u \leqq v(t)$. (W^* exists with these properties because $\operatorname{var}(W^*(u))$ is a continuous nondecreasing function of u.)

Now define, for $u \in [0, v(\tau)]$, $n = 1, 2, \ldots$ and $m = 1, 2, \ldots, \infty$,

$$W^*_{n,m}(u) = \begin{cases} W_n(t) & \text{if } u = v(t) \text{ for some } t \in [0, \tau] \\ W_n(t_k -) & \text{if } v(t_k -) \leqq u < v(t_k) \text{ for some } k > m \\ W_n(t_k -) + \dfrac{u - v(t_k -)}{\delta_k} \cdot (W_n(t_k) - W_n(t_k -)) & \\ & \text{if } v(t_k -) \leqq u < v(t_k) \text{ for some } k \leqq m \end{cases}$$

and define, also for $u \in [0, v(\tau)]$ and $m = 1, 2, \ldots, \infty$,

$$W^*_{\infty,m}(u) = \begin{cases} W^*(v(t)) & \text{if } u = v(t) \text{ for some } t \in [0, \tau] \\ W^*(v(t_k -) & \text{if } v(t_k -) \leqq u < v(t_k) \text{ for some } k > m \\ W^*(v(t_k -)) + \dfrac{u - v(t_k -)}{\delta_k} (W^*(v(t_k)) - W^*(v(t_k -)) & \\ & \text{if } v(t_k -) \leqq u < v(t_k) \text{ for some } k \leqq m. \end{cases}$$

In words, for $n = 1, 2, \ldots$, $W^*_{n,m}$ is obtained from W_n by inserting time intervals of length δ_k at t_k and joining $W_n(t_k -)$ to $W_n(t_k)$ by a straight line across this interval if $k \leqq m$, but continuing a horizontal line from $W_n(t_k -)$ if $k > m$; while $W^*_{\infty,m}$ is obtained from W^* in a similar way except that W^* has already been defined on the extended time interval. We see from this construction that $W^*_{\infty,m}$ is a random element of $D[0, v(\tau)]$ for $m < \infty$ while $W^*_{\infty.\infty}$ is a random element of $C[0, v(\tau)]$.

Now fixing $m < \infty$ for the moment, we can prove that $W^*_{n,m} \to_{\mathscr{D}} W^*_{\infty,m}$ as $n \to \infty$ in $D[0, v(\tau)]$. For the convergence of the finite dimensional distributions is again straightforward. Tightness is proved by proving tightness on each of the $2m+1$ intervals $[0, v(s_1 -)]$, $[v(s_1 -), v(s_1)]$, $\ldots$, $[v(s_m), v(\tau)]$ where $s_1, \ldots, s_m$ is $t_1, \ldots, t_m$ put into increasing order. Tightness on an interval of the form $[v(s_k -), v(s_k)]$ follows from convergence in distribution of $(W_n(s_k -), W_n(s_k))$ as $n \to \infty$, while tightness on $[v(s_{k-1}), v(s_k -)]$ follows from tightness of W_n (redefined in the points s_k as $W_n(s_k -)$) on $[s_{k-1}, s_k]$ and the observation that the modulus of continuity $w'_x(\delta)$ is smaller for $W^*_{n,m}$ on $[v(s_{k-1}), v(s_k -)]$ than it is for W_n on $[s_{k-1}, s_k]$.

Next we show that $\mathscr{E} \| W^*_{n,m} - W^*_{n,\infty} \|^2_{v(\tau)} \to 0$ as $m \to \infty$ uniformly in $n = 1, 2, \ldots, \infty$. For any n

$$\| W^*_{n,m} - W^*_{n,\infty} \|^2_{v(\tau)} = \sup_{k > m} (W^*_{n,\infty}(v(t_k)) - W^*_{n,\infty}(v(t_k -)))^2$$
$$\leqq \sum_{k > m} (W^*_{n,\infty}(v(t_k)) - W^*_{n,\infty}(v(t_k -)))^2,$$

which implies that

$$\mathscr{E} \| W^*_{n,m} - W^*_{n,\infty} \|^2_{v(\tau)} \leqq \sum_{k > m} f(t_k) \delta_k \to 0$$

as $m\to\infty$ uniformly in $n=1,2,\ldots,\infty$. Since the Skorohod d_0-distance on $D[0,\nu(\tau)]$ is smaller than the supremum norm distance, we have now shown that in $D[0,\nu(\tau)]$

$$W^*_{n,m}\to_{\mathscr{D}} W^*_{\infty,m} \qquad \text{as } n\to\infty \text{ for } m=1,2,\ldots$$

$$W^*_{\infty,m}\to_{\mathscr{D}} W^*_{\infty,\infty} \qquad \text{as } m\to\infty$$

and

$$\lim_{m\to\infty}\limsup_{n\to\infty} P(d_0(W^*_{n,m}, W^*_{n,\infty})\geqq\varepsilon)=0 \qquad \text{for all } \varepsilon>0.$$

So by Billingsley [3, Theorem 4.2]

$$W^*_{n,\infty}\to_{\mathscr{D}} W^*_{\infty,\infty} \qquad \text{as } n\to\infty.$$

Since $W^*_{\infty,\infty}$ has almost surely continuous sample paths we can apply the Skorohod-Dudley theorem; and going back to the interval $[0,\tau]$ we have finally constructed W'_n, $n=1,2,\ldots$ and W' having the same marginal distributions as W_n *and* W but now defined on a single probability space and satisfying $\|W'_n-W'\|_\tau\to 0$ almost surely as $n\to\infty$. □

Lemma 5. *Let* H_n, *and* Z_n, $n=1,2,\ldots$ *and* Z *be random elements of* $D[0,\tau]$ *defined on a single probability space* $(\Omega,\mathscr{F},P)$ *and such that with probability* 1, H_n *and* Z_n *have paths of bounded variation and* $Z_n(0)=0$ *for each* $n=1,2,\ldots,(*)$ $\limsup_{n\to\infty}\int_{(0,\tau]}|dH_n(s)|<\infty$, $\|Z_n-Z\|\to 0$, *and* $\|H_n-h\|\to 0$ *where* h *is a fixed function of bounded variation on* $[0,\tau]$. *Then* $\|\int H_{n-}\,dZ_n-\int h_-\,dZ\|\to 0$ *and* $\|\int H_n\,dZ_n-\int h\,dZ\|\to 0$ *almost surely, where* $\int h\,dZ$ *is defined as* $hZ-\int Z_-\,dh$ *and* $\int h_-\,dZ$ *as* $hZ-\int Z\,dh$ *(because* Z *does not necessarily have paths of bounded variation). In fact to conclude* $\|\int H_{n-}\,dZ_n-\int h_-\,dZ\|\to 0$ *or* $\|\int^- H_n\,dZ_n-\int^- h\,dZ\|\to 0$ *almost surely we can weaken* $(*)$ *to* $\limsup_{n\to\infty}\int_{(0,\tau)}|dH_n(s)|<\infty$.

If furthermore stochastic integrals $h_-\circ Z$ *and* $h\circ Z$ (Meyer [11, definition 18]) *can be defined, then these coincide with* $\int h_-\,dZ$ *and* $\int h\,dZ$, *respectively.*

Proof. Fix an $\omega\in\Omega$ not in the exceptional event of probability zero specified above, and denote by H_n, Z_n and Z the functions on $[0,\tau]$: $H_n(\omega)$, $Z_n(\omega)$ and $Z(\omega)$. Choose an $\varepsilon>0$. Then there exists a Z^* $(=Z^*(\omega,\varepsilon))$ such that $Z^*(0)=0$, Z^* is of bounded variation, and $\|Z-Z^*\|<\varepsilon$. Next we write

$$\begin{aligned}\|\textstyle\int H_{n-}\,dZ_n-\int h_-\,dZ\| &\leqq \|\textstyle\int H_{n-}\,dZ_n-\int H_{n-}\,dZ^*\|\\ &\quad+\|\textstyle\int H_{n-}\,dZ^*-\int h_-\,dZ^*\|+\|\int h_-\,dZ^*-\int h_-\,dZ\|\\ &=A_n+B_n+C \qquad \text{(say)}.\end{aligned}$$

Now using the formula $\int_{(0,t]} X(s-)\,dY(s)=\int_{(0,t)}(Y(t)-Y(s))\,dX(s)$ for functions X, Y in $D[0,\tau]$ of bounded variation and with $Y(0)=0$, and the fact that $\limsup_{n\to\infty}\int_{(0,\tau)}|dH_n(s)|<\infty$ it is easy to show that

$$\limsup_{n\to\infty} A_n=o(1) \qquad \text{as } \varepsilon\downarrow 0$$

and that

$$C=o(1) \qquad \text{as } \varepsilon\downarrow 0.$$

Finally $B_n \leqq \|H_n - h\| \int_{[0,\tau]} |dZ^*(s)| \to 0$ as $n \to \infty$, so combining these relationships we have the required result. Similar arguments establish the other assertions of almost sure convergence in norm.

Now we look at the second part of the lemma. By Meyer [11, IV n° 23], if $h_- \circ Z$ and $h \circ Z$ can be defined, $hZ = h_- \circ Z + Z \circ h$; by [11, IV n° 29], $Z \circ h = \int Z\,dh$. So $hZ - \int Z\,dh = h \circ Z$. Again by [11, IV n° 38] $hZ = h \circ Z + Z \circ h$ where according to [11, IV n° 29] $Z_- \circ h = \int Z_-\,dh$. So $hZ - \int Z_-\,dh = h \circ Z$. □

Corollary. *Suppose Lemma 5 allows us to conclude that almost surely*

$$\left\|\int H_n^{(i)}\,dZ_n - \int h^{(i)}\,dZ\right\| \to 0, \qquad i = 1, \ldots, r$$

and

$$\|H_{n-}^{(i)}\,dZ_n - \int h_-^{(i)}\,dZ\| \to 0, \qquad i = r+1, \ldots, s.$$

Then $\left\|\int \prod_{i=1}^{r} H_n^{(i)} \cdot \prod_{i=r+1}^{s} H_{n-}^{(i)}\,dZ_n - \int \prod_{i=1}^{r} h^{(i)} \prod_{i=r+1}^{s} h_-^{(i)}\,dZ\right\| \to 0$ *almost surely.*

Proof. Apply Lemma 5 first to $H_n^{(1)}$ and Z_n, then to $H_n^{(2)}$ and $\int H_n^{(1)}\,dZ_n$, etc. □

Acknowledgements. I would like to thank Prof. Dr. J. Oosterhoff and a referee for their useful comments.

References

1. Aalen, O.O.: Nonparametric inference for a family of counting processes. Ann. Statist. **6**, 701–726 (1978)
2. Bather, J.A.: On the sequential construction of an optimal age replacement policy. Proc. 41st Session ISI, New Delhi (1977)
3. Billingsley, P.: Weak convergence of probability measures. New York: Wiley 1968
4. Breslow, N., Crowley, J.: A large sample study of the life table and product limit estimates under random censorship. Ann. Statist. **2**, 437–453 (1974)
5. Chow, Y.S.: Local convergence of martingales and the law of large numbers. Ann. Math. Statist. **36**, 552–558 (1965)
6. Gill, R.D.: Testing with replacement and the product limit estimator. Report SW 57, Mathematical Centre. Amsterdam (1978)
7. Gill, R.D.: Censoring and Stochastic Integrals. To appear in the series M.C. Tracts, Mathematical Centre, Amsterdam (1980)
8. Jacod, J.: Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. Z. Wahrscheinlichkeitstheorie verw. Gebiete **31**, 235–253 (1975)
9. Lagakos, S.W., Sommer, C.J., Zelen, M.: Semi Markov models for censored data. Biometrika **65**, 311–317 (1978)
10. Litser, R.S., Shiryaev, A.N.: Statistics of random processes. II: Applications. Berlin-Heidelberg-New York: Springer 1978
11. Meyer, P.A.: Un cours sur les intégrales stochastiques. Lecture Notes in Mathematics **511**, 245–400. Berlin-Heidelberg-New York: Springer 1976
12. Pyke, R.: Markov renewal processes: definitions and preliminary properties. Ann. Math. Statist. **32**, 1231–1242 (1961)
13. Pyke, R.: Applications of almost surely convergent constructions of weakly convergent processes, Lecture Notes in Mathematics **89**, 187–200. Berlin-Heidelberg-New York: Springer 1969

Received December 22, 1978; in revised form February 28, 1980