The Annals of Statistics
1985, Vol. 13, No. 2, 564–582

DEMOGRAPHIC INCIDENCE RATES AND ESTIMATION OF INTENSITIES WITH INCOMPLETE INFORMATION

BY ØRNULF BORGAN AND HENRIK RAMLAU-HANSEN

University of Oslo and University of Copenhagen

Many population processes in demography, epidemiology and other fields can be represented by a time-continuous Markov chain model with a finite state space. If we have complete information on the life history of a cohort, the intensities of the Markov model may be estimated by the occurrence/exposure rates or by nonparametric techniques. In many situations, however, we have only incomplete information. In this paper we consider the special, but important, case where the occurrences and the total exposure are known, but not the distribution of the latter over the various separate statuses. Methods for handling such data, so-called demographic incidence rates, and methods for estimating the intensities from this kind of data, are known in the literature. However, their statistical properties are only vaguely known. The present paper gives a thorough presentation of the theory of these methods, and provide rigorous proofs of their statistical properties using stochastic process theory.

1. Introduction. In demography and fields with a similar methodology, a population process on the individual level can frequently be represented by a time-continuous Markov chain with a finite state space. The states of the Markov chain represent the demographic statuses and the jumps between the states correspond to the demographic events. The time parameter of the Markov process may be the age of an individual, the time since a specific event, or some such quantity. We will call it *seniority*.

Given this general framework, the population phenomena may be described by the transition intensities of the model. Consequently, it is of great interest to estimate these functions. In principle, this is an easy task if the individuals studied are watched continuously over some period of time. One may then use the classical methods based on occurrence/exposure rates (for a review, see Hoem, 1976), or for small populations or samples, nonparametric techniques developed by Aalen (1978) and others. Quite often, however, one does not obtain the sufficiently detailed data required to use any of these methods. This fact has inhibited the use of Markov chain models in demography and related fields. The development and study of statistical methods for situations with incomplete data is therefore of considerable interest.

There seems to be no general solution to the problem of estimating the intensities of partially observed Markov chains. In the present paper we consider a special, but important, type of incomplete observation. Namely, that transitions within a subset of states **K** are observed in detail, while counts of transitions out

Received February 1983; revised April 1984.

AMS 1980 *subject classifications*. Primary 62M05, 62P99.

Key words and phrases. Asymptotic theory, cohort analysis, counting processes, cumulative incidence rates, martingales, occurrence/exposure rates, time-continuous Markov chains.

of this subset are only observed aggregated over the states. In applications, **K** usually corresponds to the various demographic statuses for individuals who are alive, so that a transition out of this subset corresponds to a death. This type of data is often collected by Central Bureaus of Statistics. For instance, if marriages in a female birth cohort are studied, one may have detailed information about the marriages, but no information about the distribution of the women over the various marital statuses.

In accordance with our special observational plan, we will have to assume that the intensity for transition out of **K** does not depend on the state from which the transition is made. In most applications, this is an assumption of *nondifferential mortality*. This is, of course, quite a strong assumption. However, it should be fulfilled, at least approximately, in demographic studies of e.g. fertility and marriage formation and dissolution, in labor market studies, and in epidemiological studies of nonlethal diseases. Moreover, investigations by Finnäs (1980) indicate that the methods studied in this paper are fairly robust to deviations from the assumption of nondifferential risk of transition out of **K**.

For situations which fit into this framework, demographers have computed *incidence rates* for a long time. These are calculated as the number of occurrences of a specific event during a given period divided by the mean population alive during that period, for all statuses specified, taken together. Those not "at risk" for the event in question are not excluded from the population in the denominator, as is the case for occurrence/exposure rates. Such rates, e.g. the number of first marriages per 1000 women, are published regularly by Central Bureaus of Statistics. When incidence rates are cumulated over age for a closed cohort, one often gets an estimate for the prevalence of the event studied (Hoem, 1978).

For simple situations, like a Markov model for first marriages in a female birth cohort, it has been known for some time how the *cumulative incidence rates* can be used to compute estimates for the intensities themselves. Finnäs (1980) showed how this method can be generalized to any Markov chain model. He also gave a rather informal discussion of the statistical properties of the proposed estimators.

The main purpose of the present paper is to give a thorough presentation of the theory of incidence rates and the related estimators for the intensities, and to use stochastic process theory to provide rigorous proofs for their distributional properties. The plan of the paper is as follows. In Section 2 we introduce the Markov chain model and describe the observational plan. In Section 3 we show how the situation at hand may be formulated in a counting process framework, and review some useful results which emerge from this formulation. The cumulative incidence rate is introduced in Section 4, where we also discuss its statistical properties. In Sections 5 and 6 we discuss nonparametric estimation of the integrated intensities and estimation of piecewise constant intensities. Some approximation formulas are given in Section 7. In Section 8, the final section, a numerical example for a first marriage model is given.

We will use results from the theory of counting processes, martingales and stochastic integrals without further comment. The reviews given by Aalen (1978), Aalen and Johansen (1978) and Gill (1980) should cover our needs. An approach to the theory of counting processes attempting to minimize the dependence on general martingale theory is given by Jacobsen (1982).

2. Model and data. Assume that we study a closed cohort of n individuals, and that the phenomena of interest can be described by a time-continuous Markov chain model with time domain $[0, z]$ and with a finite state space $\mathbf{I}$. We assume that the transition probabilities $\{P_{ij}(x, y)\}$ are absolutely continuous in (x, y), and that the intensities, or forces of transition, defined as $\alpha_{ij}(x) = \lim_{y\downarrow x} P_{ij}(x, y)/(y - x)$, for $i \neq j$, exist and are integrable.

As mentioned in the introduction, $\mathbf{I}$ contains an absorbing subset $\mathbf{R}$ of states, i.e. $\sum_{k\in\mathbf{K}} P_{ik}(x, y) = 0$ for $0 \le x \le y \le z$ for all $i \in \mathbf{R}$ where $\mathbf{K} = \mathbf{I}\backslash\mathbf{R}$, such that transitions within $\mathbf{K}$ are observed in detail, while counts of transitions from $\mathbf{K}$ to $\mathbf{R}$ are only observed aggregated over the states. Moreover, we have to assume that $\sum_{j\in\mathbf{R}} \alpha_{ij}(\cdot) = \mu(\cdot)$ for all $i \in \mathbf{K}$, so that there is *nondifferential* risk of transition from $\mathbf{K}$ to $\mathbf{R}$.

For Markov chains with this structure, Hoem (1969) has proved that

$$P_{ij}(x, y) = \bar{P}_{ij}(x, y)p(x, y) \quad \text{for} \quad i, j \in \mathbf{K}, \tag{2.1}$$

where $p(x, y) = \exp(-\int_x^y \mu(s)\, ds)$ and the $\{\bar{P}_{ij}(x, y)\}$ are the transition probabilities of the *partial* Markov chain with state space $\mathbf{K}$ obtained by deleting $\mathbf{R}$, i.e. substituting 0 for α_{ij} for all (i, j) with $j \in \mathbf{R}$. The partial Markov chain is a convenient mathematical construction, but it does not necessarily have any real-life interpretation. However, in the present situation, with nondifferential risk of transition from $\mathbf{K}$ to $\mathbf{R}$, $\bar{P}_{ij}(x, y)$ coincides with the conditional probability that an individual in state i at seniority x will be in state j at seniority y, given that he is still in $\mathbf{K}$ at some later seniority w $(y \le w \le z)$ (Hoem, 1969).

Let us assume that the individuals start in a state in $\mathbf{K}$ at seniority 0 independently of each other and according to an initial distribution $(\pi_k;\ k \in \mathbf{K})$. Define $\bar{P}_i(x) = \sum_{k\in\mathbf{K}} \pi_k \bar{P}_{ki}(0, x)$, and let $P_i(x) = \bar{P}_i(x)p(x)$ for $i \in \mathbf{K}$, where $p(x) = p(0, x)$. Thus $P_i(x)$ is the probability of being in state i at seniority x, while the similar partial quantity $\bar{P}_i(x)$, in the present context, may be interpreted as the conditional probability of being in state i at seniority x, given that the individual is still in $\mathbf{K}$ at some later seniority w $(x \le w \le z)$. We assume that out of the n individuals, N_k start out in state $k \in \mathbf{K}$, so that $N_k/n \to_P \pi_k$ as $n \to \infty$.

Demographers often estimate quantities like the proportion of survivors in a female birth cohort which have experienced a birth of a certain order, or the proportion ever married at the various ages (cf. Hoem, 1978). More generally, one may want to estimate the expected number of transitions $B_{ij}(x)$ directly from a state $i \in \mathbf{K}$ to another state $j \in \mathbf{K}$ in the seniority interval $[0, x]$ for an individual who is still in $\mathbf{K}$ at seniority x. It is seen that

$$B_{ij}(x) = \int_0^x \alpha_{ij}(s)\bar{P}_i(s)\, ds. \tag{2.2}$$

Using the Kolmogorov differential equations one gets

$$\bar{P}_i(x) = B_{\cdot i}(x) - B_{i\cdot}(x) + \pi_i, \tag{2.3}$$

as in Finnäs (1980), where the dots here and in what follows signify summation over all $k \in \mathbf{K}\backslash\{i\}$, if another definition is not explicitly given. Formula (2.3) also follows directly from the fact that the indicator random variable for being in state i at seniority x, for an individual who is known to be in $\mathbf{K}$ at this seniority, equals the number of transitions into i, minus the number of transitions out of i, plus the indicator for being in i at seniority 0. It should be realized that the assumption of nondifferential risk of transition from $\mathbf{K}$ to $\mathbf{R}$ is essential for (2.2), and therefore also (2.3), to hold true. Moreover, since (2.2) and (2.3) are key relations in what follows, this assumption is absolutely necessary for the results presented in this paper.

Estimation of the $\{B_{ij}\}$ is discussed in Section 4 below. Although these are important quantities, it is the intensities $\{\alpha_{ij}\}$ which measure the instantaneous risk of transition between the various states. Therefore, it is the intensities which are the entities of main interest. In Sections 5 and 6 it is shown how the estimators for the $\{B_{ij}\}$ can be used to estimate the intensities themselves.

3. A counting process formulation. Denote by $K_{ij}(x)$ the number of transitions directly from state i to state j experienced by the cohort in the seniority interval $[0, x]$, and let $Y_i(x)$ be the number of individuals in state i "just before" seniority x, i.e. $Y_i(\cdot)$ is left-continuous. Moreover, let $\mathscr{F}_x$ be the σ-algebra generated by $(N_k;\, k \in \mathbf{K})$ and $(K_{ij}(s);\, 0 \le s \le x,\, i, j \in \mathbf{I},\, i \neq j)$. Then $(K_{ij}(x);\, 0 \le x \le z,\, i, j \in \mathbf{I},\, i \neq j)$ is a multivariate counting process where $K_{ij}(\cdot)$ has the intensity process $\alpha_{ij}(\cdot)Y_i(\cdot)$ relative to the increasing family of σ-algebras $(\mathscr{F}_x)$. By the theory of counting processes, this implies that $(M_{ij}(x);\, 0 \le x \le z,\, i, j \in \mathbf{I},\, i \neq j)$, given by

$$M_{ij}(x) = K_{ij}(x) - \int_0^x \alpha_{ij}(s)Y_i(s)\, ds, \tag{3.1}$$

are orthogonal square integrable martingales with respect to $(\mathscr{F}_x)$. The variance process $\langle M_{ij}\rangle(\cdot)$ of M_{ij} is

$$\langle M_{ij}\rangle(x) = \int_0^x \alpha_{ij}(s)Y_i(s)\, ds, \tag{3.2}$$

which means that $M_{ij}^2 - \langle M_{ij}\rangle$ is a square integrable martingale.

For the situation with complete information, i.e. when the information structure is $(\mathscr{F}_x)$, Aalen (1978) proposed the estimators

$$\hat{A}_{ij}(x) = \int_0^x J_i(s)[Y_i(s)]^{-1}\, dK_{ij}(s), \tag{3.3}$$

for the cumulative intensities $A_{ij}(x) = \int_0^x \alpha_{ij}(s)\, ds$, generalizing the estimator proposed by Nelson (1969). Here $J_i(x) = I(Y_i(x) > 0)$ is an indicator process, and 0/0 is interpreted as 0. Results on uniform consistency and asymptotic normality of these estimators can be found in Aalen (1978).

In the situation considered in this paper, however, one does not observe the $\{Y_i\}$ and all the $\{K_{ij}\}$, so we are not able to calculate the Nelson-Aalen estimators

(3.3). Our observational plan is to observe (N_k), the K_{ij} with $i, j \in \mathbf{K}$, and $D = \sum_{i\in\mathbf{K}} \sum_{j\in\mathbf{R}} K_{ij}$, the process counting the number of absorptions in $\mathbf{R}$. This means that the individual $\{Y_i;\ i \in \mathbf{K}\}$ are not observed, but $Y.(x) = \sum_{i\in\mathbf{K}} Y_i(x) = n - D(x-)$ is still observable. Hence, the observed "history" of the cohort in the seniority interval $[0, x]$ can be described by the σ-algebra $\mathscr{G}_x$ generated by $(N_k;\ k \in \mathbf{K})$ and $(D(s), K_{ij}(s);\ 0 \le s \le x,\ i, j \in \mathbf{K},\ i \ne j)$.

For this situation we could try to use the EM algorithm to estimate the cumulative intensities $A_{ij}(x)$. Note that analogously to (3.3), we may estimate $\int_0^x \mu(s)\, ds$ by $\int_0^x [Y.(s)]^{-1}\, dD(s)$. So in principle, given initial estimates of $A_{ij}(x)$ for all $i, j \in \mathbf{K}$, $i \ne j$, $\hat{Y}_i(x) = E(Y_i(x) \mid \mathscr{G}_z)$ could be computed if unknown parameters are replaced by estimates. This would give $\int_0^x [\hat{Y}_i(s)]^{-1}\, dK_{ij}(s)$ as a new estimate for $A_{ij}(x)$. Iteration would then give nonparametric estimators of the $\{A_{ij}(x)\}$. However, the expectation step in this iterative procedure is not feasible in practice (Borgan and Ramlau-Hansen, 1983, Section 6), so we will have to derive estimators for the integrated intensities from more ad hoc arguments. This will be done in Section 5.

4. Cumulative incidence rates. The purpose of this section is, within the counting process framework introduced above, to consider nonparametric estimation of the $\{B_{ij}\}$ defined by (2.2), and to prove uniform consistency and asymptotic normality of the estimators. Introduce the *cumulative incidence rate*

$$\hat{B}_{ij}(x) = \int_0^x J(s)[Y.(s)]^{-1}\, dK_{ij}(s), \tag{4.1}$$

where $J(x) = I(Y.(x) > 0)$, as an estimator for $B_{ij}(x)$ for $i, j \in \mathbf{K}$, $i \ne j$. This is an estimator of the same type as the Nelson-Aalen estimator (3.3).

To see that (4.1) is an approximately unbiased estimator for $B_{ij}(x)$, note that by (3.1)

$$\begin{aligned} \hat{B}_{ij}(x) = &\int_0^x \alpha_{ij}(s) Y_i(s)[Y.(s)]^{-1}\, ds \\ &+ \int_0^x J(s)[Y.(s)]^{-1}\, dM_{ij}(s). \end{aligned} \tag{4.2}$$

Here, the final term is a stochastic integral with respect to a square integrable martingale, and hence it is a zero mean martingale itself. Thus

$$E\hat{B}_{ij}(x) = \int_0^x \alpha_{ij}(s) E[Y_i(s)/Y.(s)]\, ds.$$

Since we have assumed a nondifferential risk of transition from $\mathbf{K}$ to $\mathbf{R}$, $(Y_i(s);\ i \in \mathbf{K})$ is multinomially distributed with parameters $(\bar{P}_i(s);\ i \in \mathbf{K})$, conditionally on $Y.(s) > 0$. Therefore,

$$\begin{aligned} E\hat{B}_{ij}(x) &= \int_0^x \alpha_{ij}(s)\bar{P}_i(s) P(Y.(s) > 0)\, ds \\ &= \int_0^x P(Y.(s) > 0)\, dB_{ij}(s), \end{aligned}$$

where $P(Y_{\cdot}(s) > 0) = 1 - (1 - p(s))^n$. Thus, $\hat{B}_{ij}(x)$ is almost an unbiased estimator for $B_{ij}(x)$ when n is large.

To discuss the asymptotic properties of (4.1), consider the sequence of counting processes we get by letting $n \to \infty$, and index all relevant quantities by n. We will prove that $\hat{B}_{ij}^{(n)}$ is an uniformly consistent estimator for B_{ij}.

THEOREM 4.1. *Let $\hat{B}_{ij}^{(n)}(x)$ and $B_{ij}(x)$ be given by* (4.1) *and* (2.2), *respectively. Then under the assumptions given in Section* 2

$$\sup_{x\in[0,z]} |\hat{B}_{ij}^{(n)}(x) - B_{ij}(x)| \to_P 0$$

as $n \to \infty$.

PROOF. By (4.2) it is sufficient to prove that

$$\sup_{s\in[0,z]} | Y_i^{(n)}(s)[Y_{\cdot}^{(n)}(s)]^{-1} - \bar{P}_i(s)| \to_P 0 \tag{4.3}$$

and

$$\sup_{x\in[0,z]} \left| \int_0^x J^{(n)}(s)[Y_{\cdot}^{(n)}(s)]^{-1}\, dM_{ij}^{(n)}(s) \right| \to_P 0 \tag{4.4}$$

as $n \to \infty$. By standard results (4.3) is fulfilled. When we use Lenglart's (1977) inequality (cf. Andersen and Gill, 1982, Appendix I) and (3.2), we have

$$P\left\{\sup_{x\in[0,z]} \left| \int_0^x J^{(n)}(s)[Y_{\cdot}^{(n)}(s)]^{-1}\, dM_{ij}^{(n)}(s) \right|^2 \geq \varepsilon\right\}$$

$$\leq \frac{\delta}{\varepsilon} + P\left\{\int_0^z J^{(n)}(s)[Y_{\cdot}^{(n)}(s)]^{-2}\, d\langle M_{ij}^{(n)}\rangle(s) > \delta\right\}$$

$$= \frac{\delta}{\varepsilon} + P\left\{\int_0^z \alpha_{ij}(s) Y_i^{(n)}(s)[Y_{\cdot}^{(n)}(s)]^{-2}\, ds > \delta\right\}$$

for all $\varepsilon, \delta > 0$. Since $\int_0^z \alpha_{ij}(s) Y_i^{(n)}(s)[Y_{\cdot}^{(n)}(s)]^{-2}\, ds \to_P 0$ as $n \to \infty$, (4.4) is also fulfilled. □

Let us then turn to the problem of proving an asymptotic distributional result for the cumulative incidence rates (4.1). The usual way of proving such results within the counting process framework is to apply some version of the martingale central limit theorem (Rebolledo, 1978, 1980). In our case, however, it seems difficult to proceed in this way, and we will therefore use a Skorohod construction as applied in Breslow and Crowley (1974, Theorem 4).

For this purpose, we need the asymptotic distribution of the number of transitions between the various states in the Markov chain. Let us denote the normalized number of transitions between state i and j, $i, j \in \mathbf{I}$, $i \neq j$, by

$$X_{ij}^{(n)}(t) = \sqrt{n}\{n^{-1}K_{ij}^{(n)}(t) - \nu_{ij}(t)\}, \tag{4.5}$$

where $\nu_{ij}(t) = n^{-1}EK_{ij}^{(n)}(t) = \int_0^t P_i(\sigma)\alpha_{ij}(\sigma)\, d\sigma$. Then by the central limit theorem for Markov chains given in the Appendix, we have that $X^{(n)} = (X_{ij}^{(n)};\ i, j \in \mathbf{I}, i \neq j)$ converges weakly to a mean zero Gaussian process $X = (X_{ij};\ i, j \in \mathbf{I}, i \neq j)$ with covariance structure given for $s \le t$ by

$$
\begin{aligned}
\operatorname{Cov}(X_{ij}(s), X_{kl}(t)) = & \int_0^s \int_0^\tau P_{jk}(\sigma, \tau)\alpha_{kl}(\tau)\, d\nu_{ij}(\sigma)\, d\tau \\
& + \int_s^t \int_0^s P_{jk}(\sigma, \tau)\alpha_{kl}(\tau)\, d\nu_{ij}(\sigma)\, d\tau \\
& + \int_0^s \int_0^\sigma P_{li}(\tau, \sigma)\alpha_{ij}(\sigma)\, d\nu_{kl}(\tau)\, d\sigma \\
& + \delta_{ik}\delta_{jl}\nu_{ij}(s) - \nu_{ij}(s)\nu_{kl}(t),
\end{aligned}
\tag{4.6}
$$

where δ_{ik} is a Kronecker delta.

To apply this result to prove weak convergence of the cumulative incidence rates given in (4.1), we introduce

$$
U^{(n)}(x) = \sqrt{n}\{n^{-1}Y_{\cdot}^{(n)}(x) - p(x)\}, \tag{4.7}
$$

and denote the corresponding limiting process by U. Then, by (4.6) and the assumption of nondifferential risk of transition from $\mathbf{K}$ to $\mathbf{R}$, it is seen that $(U^{(n)}, X_{ij}^{(n)};\ i, j \in \mathbf{K}, i \neq j)$ converges weakly to $(U, X_{ij};\ i, j \in \mathbf{K}, i \neq j)$, where

$$
\operatorname{Cov}(X_{ij}(x), U(y)) = \begin{cases} \displaystyle\int_0^x p(s, y)\, d\nu_{ij}(s) - p(y)\nu_{ij}(x) & \text{for } x \le y \\ \displaystyle\nu_{ij}(x) - \nu_{ij}(y) + \int_0^y p(s, y)\, d\nu_{ij}(s) - p(y)\nu_{ij}(x) & \text{for } y < x \end{cases} \tag{4.8}
$$

and

$$
\operatorname{Cov}(U(x), U(y)) = p(y)(1 - p(x)) \quad \text{for } x \le y. \tag{4.9}
$$

We are now able to state the following result.

THEOREM 4.2. *Under the assumptions in Section 2, the multivariate process* $(\sqrt{n}(\hat{B}_{ij}^{(n)} - B_{ij});\ i, j \in \mathbf{K}, i \neq j)$ *converges weakly to a zero mean Gaussian process* $(Z_{ij};\ i, j \in \mathbf{K}, i \neq j)$, *where*

$$
\begin{aligned}
Z_{ij}(x) = & -\int_0^x U(s)[p(s)]^{-2}\, d\nu_{ij}(s) + X_{ij}(x)[p(x)]^{-1} \\
& - \int_0^x X_{ij}(s)\mu(s)[p(s)]^{-1}\, ds.
\end{aligned}
\tag{4.10}
$$

For $x \leq y$, the covariance structure of the limiting process is given by

$$\begin{aligned}
\operatorname{Cov}(Z_{ij}(x), Z_{kl}(y)) = & \int_0^x \int_0^u P_{jk}(r, u)\alpha_{kl}(u)[p(r)p(u)]^{-1}\, d\nu_{ij}(r)\, du \\
& + \int_0^x \int_0^r P_{li}(u, r)\alpha_{ij}(r)[p(r)p(u)]^{-1}\, d\nu_{kl}(u)\, dr \\
& + \int_x^y \int_0^x P_{jk}(r, u)\alpha_{kl}(u)[p(r)p(u)]^{-1}\, d\nu_{ij}(r)\, du \\
& + \delta_{ik}\delta_{jl} \int_0^x [p(u)]^{-2}\, d\nu_{ij}(u) \qquad (4.11) \\
& - \int_0^x \int_0^u [p(r)]^{-2}[p(u)]^{-1}\, d\nu_{ij}(r)\, d\nu_{kl}(u) \\
& - \int_0^x \int_0^u [p(r)]^{-2}[p(u)]^{-1}\, d\nu_{kl}(r)\, d\nu_{ij}(u) \\
& - \left(\int_0^x [p(r)]^{-2}\, d\nu_{ij}(r)\right)\left(\int_x^y [p(u)]^{-1}\, d\nu_{kl}(u)\right).
\end{aligned}$$

PROOF. The convergence of $(\sqrt{n}(\hat{B}_{ij}^{(n)} - B_{ij}))$ to (Z_{ij}), given by (4.10), follows by a Skorohod construction just as in the proof of Theorem 4 in Breslow and Crowley (1974). (Be aware of the misprint pointed out by Gill, 1983, page 50.) The covariance structure (4.11) follows by straightforward computations given in some detail in Borgan and Ramlau-Hansen (1983, Appendix B). □

In general it seems as if we need data at the individual level to estimate the covariances given by (4.11). However, when $(i, j) = (k, l)$ and only one transition from i to j is possible for each individual, the first three terms of (4.11) vanish, so that $\operatorname{Cov}(Z_{ij}(x), Z_{ij}(y))$ may be estimated by substituting $n^{-1}Y_{\cdot}^{(n)}(\cdot)$ and $n^{-1}K_{ij}^{(n)}(\cdot)$ for $p(\cdot)$ and $\nu_{ij}(\cdot)$, respectively. An example of such a situation is given in Section 8.

5. Estimation of cumulative intensities. The purpose of this section is to develop nonparametric estimators of the integrated intensities $A_{ij}(x) = \int_0^x \alpha_{ij}(s)\, ds$, within our observational plan, and to outline how their asymptotic distributional properties may be derived. As discussed in Section 3, the estimators have to be derived by ad hoc arguments.

By (2.3), a natural "estimator" for the unobserved number $Y_i(x)$ at risk is

$$\tilde{Y}_i(x) = Y_{\cdot}(x)[\hat{B}_{\cdot i}(x-) - \hat{B}_{i\cdot}(x-) + N_i/n]. \qquad (5.1)$$

Therefore, by analogy with (3.3), we propose

$$\tilde{A}_{ij}(x) = \int_0^x \tilde{J}_i(s)[\tilde{Y}_i(s)]^{-1}\, dK_{ij}(s) \tag{5.2}$$

as an estimator for $A_{ij}(x)$, where $\tilde{J}_i(x) = I(\tilde{Y}_i(x) > 0)$.

One should realize that the "estimated number at risk" $\tilde{Y}_i(x)$ can be negative. Consider for example a first marriage model, where $\mathbf{K} = \{0, 1\}$, $\mathbf{R} = \{2\}$, and no transition is possible from state 1 to 0. Let $N_0 = N_1 = 1$, and assume that at τ_1 we have a transition $1 \to 2$, at $\tau_2 > \tau_1$ a transition $0 \to 1$, and at $\tau_3 > \tau_2$ a transition $1 \to 2$. Then $\tilde{Y}_0(x) = Y_.(x)(\frac{1}{2} - \hat{B}_{01}(x-))$, and we find $\tilde{Y}_0(\tau_1) = 2(\frac{1}{2} - 0) = 1$, $\tilde{Y}_0(\tau_2) = 1(\frac{1}{2} - 0) = \frac{1}{2}$, and $\tilde{Y}_0(\tau_3) = 1(\frac{1}{2} - 1) = -\frac{1}{2}$. The way we have defined $\tilde{A}_{ij}$, the seniorities with negative $\tilde{Y}_i$ do not contribute to the estimator, so we get a nondecreasing estimator for the integrated intensity, as we should. But the possible negativity of the "estimated number at risk" does suggest that the estimators (5.2) may behave badly in small samples. For large samples, however, they behave reasonably, as it is seen from the following result.

THEOREM 5.1. *Assume that $P_i(\cdot)$ is bounded away from zero. Then*

$$\sup_{x\in[0,z]} | \tilde{A}_{ij}^{(n)}(x) - A_{ij}(x) | \to_P 0 \quad \textit{as} \quad n \to \infty.$$

PROOF. According to Theorem 4.1 and (2.3),

$$\sup_{x\in[0,z]} | \tilde{Y}_i^{(n)}(x)/n - P_i(x) | \to_P 0 \quad \text{as} \quad n \to \infty, \tag{5.3}$$

and the remaining part of the proof follows as in Theorem 4.1 by applying Lenglart's (1977) inequality. □

It is possible to derive the asymptotic distributional properties of the $\{\tilde{A}_{ij}\}$ by an argument similar to the one used in Theorem 4.2. We will not do this here, however, for the estimators proposed in our next section will usually be preferred in the applications we have in mind.

6. Estimation of piecewise constant intensities. In demography and other fields, it is quite common to assume that the intensities are piecewise constant (e.g. Hoem, 1976; Hoem and Jensen, 1982). This may be appropriate when the focus is concentrated on the global behavior of the intensities, so that we deliberately want to smooth "nuisance" variation of the intensities (Hoem, 1972, Section 10). The assumption may also be made to simplify computations, especially for large samples, or it may be forced on us in cases when only grouped data is available (cf. Section 7 below). Therefore, in this section we let $0 = a_1 < a_2 < \cdots < a_{R+1} = z$ be a partitioning of the seniority interval $[0, z]$ into subintervals $(a_r, a_{r+1}]$, $r = 1, 2, \cdots, R$, and assume that the intensities are constant on each of the subintervals, i.e. $\alpha_{ij}(x) = \alpha_{ijr}$ for $x \in (a_r, a_{r+1}]$.

If the cohort had been observed completely, we would have estimated the $\{\alpha_{ijr}\}$ by the occurrence/exposure rates (cf. Hoem, 1976)

$$\hat{\alpha}_{ijr}^{(n)} = F_{ijr}^{(n)}/L_{ir}^{(n)} \tag{6.1}$$

where $F_{ijr}^{(n)} = K_{ij}^{(n)}(a_{r+1}) - K_{ij}^{(n)}(a_r)$ and $L_{ir}^{(n)} = \int_{a_r}^{a_{r+1}} Y_i^{(n)}(u)\,du$. For the situation considered in this paper, such detailed data are not available. However, we may "estimate" the exposure $L_{ir}^{(n)}$ by

$$\tilde{L}_{ir}^{(n)} = \int_{a_r}^{a_{r+1}} \tilde{Y}_i^{(n)}(u)\,du, \tag{6.2}$$

where $\tilde{Y}_i^{(n)}$ is given by (5.1). We therefore suggest the estimators

$$\tilde{\alpha}_{ijr}^{(n)} = F_{ijr}^{(n)}/\tilde{L}_{ir}^{(n)}, \tag{6.3}$$

for the $\{\alpha_{ijr}\}$.

By (5.3) it is seen that the estimators (6.3) are consistent. We can also prove the following result.

THEOREM 6.1. *Consider a fixed pair (i, j) and assume that $\int_{a_r}^{a_{r+1}} P_i(u)\,du > 0$ for $r = 1, \cdots, R$. Then,*

$$\{\sqrt{n}(\tilde{\alpha}_{ijr}^{(n)} - \alpha_{ijr}),\ r = 1, \cdots, R\} \to_D N_R(0, \Sigma),$$

where $\Sigma = (\sigma_{qr})$ is given by

$$\sigma_{rr} = \frac{\alpha_{ijr}}{\int_{a_r}^{a_{r+1}} P_i(\sigma)\,d\sigma} + \frac{2\alpha_{ijr}^2}{(\int_{a_r}^{a_{r+1}} P_i(\sigma)\,d\sigma)^2} \times \int_{a_r}^{a_{r+1}} \int_{a_r}^{r} \int_0^u \frac{p(u)p(r)}{p(\sigma)} \bar{P}_i(\sigma)(1 - \bar{P}_i(\sigma))\mu(\sigma)\,d\sigma\,du\,dr \tag{6.4a}$$

and

$$\sigma_{qr} = \frac{\alpha_{ijq}\alpha_{ijr}}{\int_{a_q}^{a_{q+1}} P_i(\sigma)\,d\sigma \int_{a_r}^{a_{r+1}} P_i(\sigma)\,d\sigma} \times \int_{a_r}^{a_{r+1}} \int_{a_q}^{a_{q+1}} \int_0^u \frac{p(u)p(r)}{p(\sigma)} \bar{P}_i(\sigma)(1 - \bar{P}_i(\sigma))\mu(\sigma)\,d\sigma\,du\,dr \tag{6.4b}$$

for $q < r$, and $N_R(0, \Sigma)$ denotes the R-dimensional multivariate normal distribution.

PROOF. Let $X_{ij}^{(n)}$ be given as in Section 4 by (4.5) and introduce

$$V_i^{(n)}(t) = \sqrt{n}\left(\int_0^t \tilde{Y}_i^{(n)}(u)/n\,du - \int_0^t P_i(u)\,du\right).$$

Then, by a Taylor series expansion, it follows that $\{\sqrt{n}(\tilde{\alpha}_{ijr}^{(n)} - \alpha_{ijr}),\ r = 1, \cdots, R\}$ has the same asymptotic distribution as the vector with rth component

$$\left[\int_{a_r}^{a_{r+1}} P_i(u)\,du\right]^{-1} (X_{ij}^{(n)}(a_{r+1}) - X_{ij}^{(n)}(a_r)) - \alpha_{ijr}\left[\int_{a_r}^{a_{r+1}} P_i(u)\,du\right]^{-1} (V_i^{(n)}(a_{r+1}) - V_i^{(n)}(a_r)). \tag{6.5}$$

Now we may write

$$V_i^{(n)}(t) = \int_0^t \bar{P}_i(s)\sqrt{n}(Y_{\cdot}^{(n)}(s)/n - p(s))\,ds$$
$$+ \int_0^t p(s)\sqrt{n}(\hat{B}_{\cdot i}^{(n)}(s-) - B_{\cdot i}(s))\,ds$$
$$- \int_0^t p(s)\sqrt{n}(\hat{B}_{i\cdot}^{(n)}(s-) - B_{i\cdot}(s))\,ds + \sqrt{n}\left(\frac{N_i^{(n)}}{n} - \pi_i\right)\int_0^t p(s)\,ds$$
$$+ \int_0^t \left(\hat{B}_{\cdot i}^{(n)}(s-) - \hat{B}_{i\cdot}^{(n)}(s-) + \frac{N_i^{(n)}}{n} - \bar{P}_i(s)\right)\sqrt{n}(Y_{\cdot}^{(n)}(s)/n - p(s))\,ds,$$

and it follows by a Skorohod construction that the sequence of processes $(X_{ij}^{(n)}, V_i^{(n)})$ converges weakly to a limiting Gaussian process (X_{ij}, V_i), where X_{ij} is defined in Section 4 and

$$V_i(t) = \int_0^t \bar{P}_i(s)U(s)\,ds + \int_0^t p(s)\{Z_{\cdot i}(s) - Z_{i\cdot}(s) + M_i\}\,ds.$$

Here U and Z_{ij} are given by (4.7) and (4.10), respectively, and M_i is a normal $N(0, \pi_i(1 - \pi_i))$ random variable, independent of X_{ij}, U, Z_{ij}, describing the number of individuals which start out in state $i \in \mathbf{K}$ at seniority 0. Substituting X_{ij} and V_i for $X_{ij}^{(n)}$ and $V_i^{(n)}$ in (6.5), we get random variables with the same distribution as the asymptotic distribution of $\{\sqrt{n}(\tilde{\alpha}_{ijr}^{(n)} - \alpha_{ijr}), r = 1, \cdots, R\}$. The expressions for the variances and covariances follow by some straightforward, but very tedious, calculations using (2.1), (2.3), (4.6), (4.8), (4.9) and (4.11). Details are given in Borgan and Ramlau-Hansen (1983, Appendix C). □

The first term in σ_{rr} is exactly the asymptotic variance of the occurrence/exposure rate (6.1). Therefore, the efficiency of our estimation method may be compared easily with the situation where complete information is available and the occurrence/exposure rates are used. An example of such efficiency calculations is given in Section 8. Note also that the asymptotic variances and covariances may be estimated consistently by substituting $\tilde{Y}_i^{(n)}(x)/n$ for $P_i(x)$, $\hat{B}_{\cdot i}^{(n)}(x-) - \hat{B}_{i\cdot}^{(n)}(x-) + N_i^{(n)}/n$ for $\bar{P}_i(x)$, $Y_{\cdot}^{(n)}(x)/n$ for $p(x)$, and $\tilde{\alpha}_{ijr}^{(n)}$ for α_{ijr} in (6.4a) and (6.4b). Moreover, by (6.4b), σ_{qr} is always positive, so that the estimators $(\tilde{\alpha}_{ijr}^{(n)}, r = 1, \cdots, R)$ are all positively correlated.

7. Approximation formulas. When handling large populations, which is often the case in demographic applications, the cumulative incidence rates (4.1) will be computationally demanding. This will also be the case for the estimators (6.3), which are based on (4.1). It will therefore be useful to have simple approximation formulas for (4.1) and (6.3).

To derive such approximations, let us assume that the partitioning $0 = a_1 < a_2 < \cdots < a_{R+1} = z$ of $[0, z]$ is so fine that $Y_{\cdot}(\cdot)$ does not vary much over each subinterval $(a_r, a_{r+1}]$, $r = 1, 2, \cdots, R$. Let $L_{\cdot r} = \sum_{i \in \mathrm{K}} L_{ir}$, where L_{ir} is defined just below (6.1). (We omit the index n in this section.) Then for all $s \in (a_r, a_{r+1}]$, $Y_{\cdot}(s)$ is close to its average value over this subinterval $L_{\cdot r}/(a_{r+1} - a_r)$. Hence, for $x = a_{r+1}$, we get the following approximation formula for (4.1)

$$\hat{B}_{ij}(a_{r+1}) \approx \textstyle\sum_{m=1}^{r} (F_{ijm}/L_{\cdot m})(a_{m+1} - a_m), \tag{7.1}$$

where F_{ijm} is defined just below (6.1). Using the same approximation for $Y_{\cdot}(\cdot)$, and approximating $K_{\cdot i}(x)$ and $K_{i\cdot}(x)$ by interpolating linearly between their values for $x = a_r$ and $x = a_{r+1}$, we find (cf. (5.1) and (6.2))

$$\tilde{L}_{ir} \approx L_{\cdot r}[N_i/n + \hat{B}_{\cdot i}(a_r) - \hat{B}_{i\cdot}(a_r)] + (F_{\cdot ir} - F_{i\cdot r})(a_{r+1} - a_r)/2.$$

By this we arrive at the following approximation formula for (6.3)

$$\tilde{\alpha}_{ijr} \approx F_{ijr}/\{L_{\cdot r}[N_i/n + \hat{\hat{B}}_{\cdot i}(a_r) - \hat{\hat{B}}_{i\cdot}(a_r)] + (F_{\cdot ir} - F_{i\cdot r})(a_{r+1} - a_r)/2\}, \tag{7.2}$$

where for the sake of brevity we have written $\hat{\hat{B}}_{ij}$ for the right-hand side of (7.1).

The right-hand sides of (7.1) and (7.2) are nothing but the "classical" cumulative incidence rates (cf. Hoem, 1978) and Finnäs' (1980) estimators for the intensities. So this strongly suggests that these estimators are only applicable for situations where they are good approximations to our (4.1) and (6.3). It should be realized that we need only information on the occurrences $\{F_{ijr}\}$, the total exposures $\{L_{\cdot r}\}$, and the initial distribution $\{N_i\}$ in order to compute the approximation formulas (7.1) and (7.2). Therefore, these formulas may be used in situations where data availability does not allow exact computation of the original estimators (4.1) and (6.3).

8. An example. A first marriage model. In order to illustrate the use of the demographic incidence rates (4.1) and the estimators (6.3) based on these, we have studied a first marriage model. This simple Markov model is illustrated in Fig. 1. All women start out in state 0. Once a woman gets married she moves to state 1, where she remains until death. At death she moves on to state 2. A woman who dies before she gets married (for the first time), moves directly from state 0 to state 2.

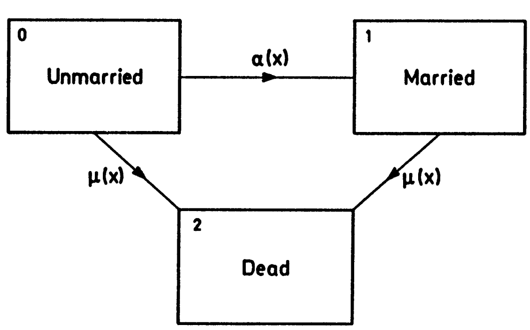

FIG. 1. *A first marriage model for a female birth cohort.*

For the female birth cohort of 32542 women born in Denmark in 1940, data were available so that we could compute the cumulative first marriage incidence rates (4.1), and estimate the first marriage intensity by the incidence rate method (6.3), as well as by the occurrence/exposure rates (6.1). (The data were also used by Finnäs, 1980, for illustrative purposes.) In the actual computations, the approximation formulas of Section 7 were used.

Let us assume that the intensities of the model in Fig. 1 are constant over single year age intervals, and let us denote the first marriage intensity and the force of mortality in the age interval $(r, r + 1]$ by α_r and μ_r, respectively. Then the cumulative incidence rates $\{\hat{B}_r = \hat{B}_{01}(r)\}$ are shown in Fig. 2, whereas the occurrence/exposure rates $\{\hat{\alpha}_r\}$ and the estimates based on the incidence rates $\{\tilde{\alpha}_r\}$ are shown in Fig. 3. The rates are also given in Table 1. As mentioned in Section 2, $\hat{B}_r$ is an estimate of the probability that a woman still alive at age r has experienced her first marriage before this age, and as such it represents a measure of the "risk" of getting married. This risk is also illustrated by the two sets of estimates for the intensities shown in Fig. 3. It is remarkable that the difference between the two sets of estimates is so small, and it suggests that the incidence method (6.3) is quite efficient.

We have also estimated the asymptotic variances and covariances given by (4.11) and (6.4), respectively. The former of these are estimated as indicated at the end of Section 4, and by applying an approximation argument similar to those of Section 7. The estimated standard deviations of the $\{\hat{B}_r\}$ are given in Table 1. The estimated correlation coefficients between $\hat{B}_r$ and $\hat{B}_s$ are high when r and s are close to each other, but diminish when the distance between r and s increases. The increments are all negatively correlated, e.g. the correlation coefficient between $\hat{B}_{24}$ and $\hat{B}_{25} - \hat{B}_{24}$ is -0.425 whereas the correlation between $\hat{B}_{24}$ and $\hat{B}_{39} - \hat{B}_{24}$ is -0.839.

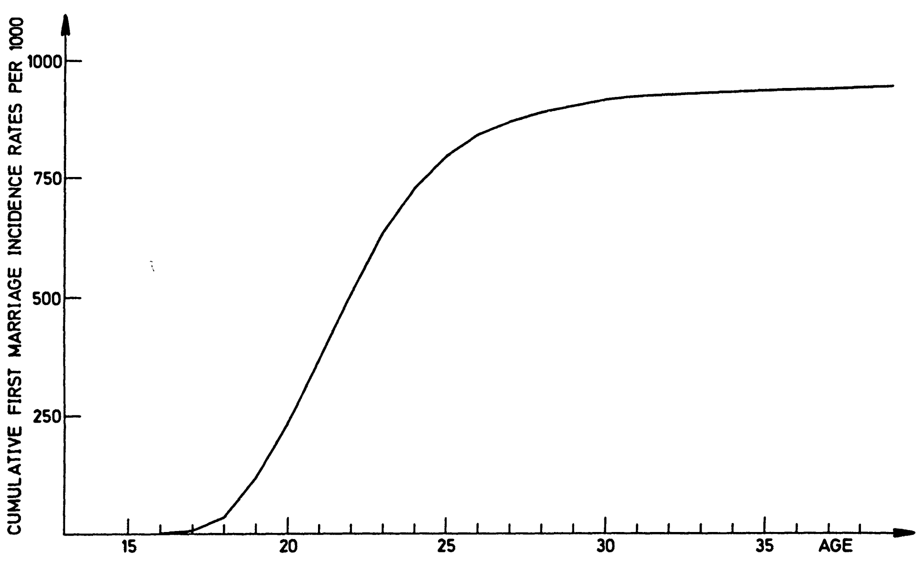

FIG. 2. *Cumulative first marriage incidence rates for women born in* 1940 *in Denmark.*

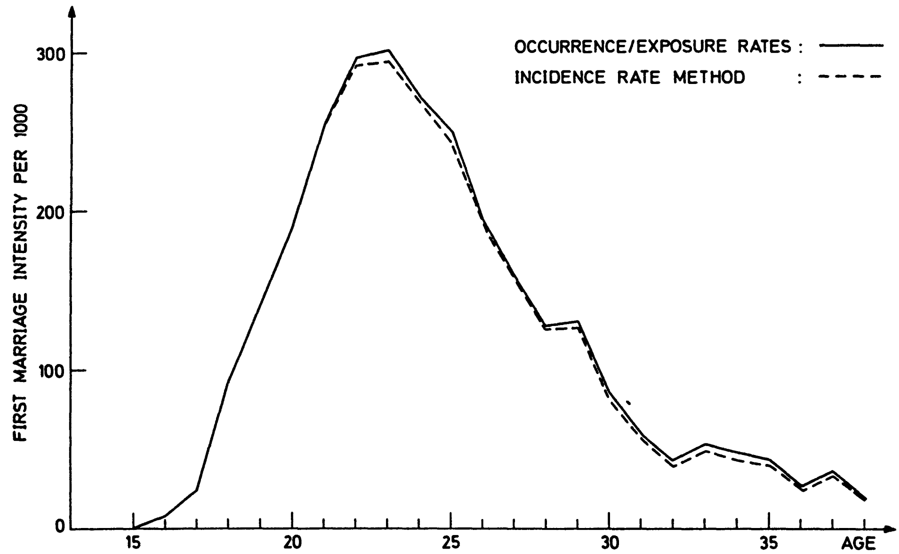

FIG. 3. *Estimated first marriage intensity for women born in* 1940 *in Denmark.*

To estimate the covariances in (6.4), the integrals in (6.4) have been expressed as functions of the $\{\alpha_r\}$ and the $\{\mu_r\}$. This is easily done, by using the relations $\bar{P}_0(\sigma) = \bar{P}_0(r)\exp(-\alpha_r(\sigma - r))$, $p(\sigma) = p(r)\exp(-\mu_r(\sigma - r))$, and $P_0(\sigma) = \bar{P}_0(\sigma)p(\sigma)$ for $\sigma \in (r, r + 1]$. Here, $\bar{P}_0(r) = \exp(-\sum_{s=0}^{r-1} \alpha_s)$ and $p(r) = \exp(-\sum_{s=0}^{r-1} \mu_s)$. In Table 1 we have given the estimated standard deviations and the asymptotic efficiency of $\tilde{\alpha}_r$ relative to $\hat{\alpha}_r$. The values have been computed by substituting $\tilde{\alpha}_r$ for α_r and

$$\hat{\mu}_r = 2(Y_{\cdot}^{(n)}(r) - Y_{\cdot}^{(n)}(r + 1))/(Y_{\cdot}^{(n)}(r) + Y_{\cdot}^{(n)}(r + 1))$$

for μ_r in (6.4a). The estimated standard deviations of the two methods differ very little, which implies an efficiency of the incidence rate method of 99.5 per cent or more for all ages.

The estimated correlation coefficients between $\tilde{\alpha}_q$ and $\tilde{\alpha}_r$ do not exceed 0.005 for any two age intervals, so the $\{\tilde{\alpha}_r\}$ are nearly asymptotically independent. This indicates that the slightly lower estimates obtained by the incidence rate method are not due to the positive correlation between the estimates, but, as suggested by Finnäs (1980), are due to the fact that out-migration for this birth cohort mostly takes place among the unmarried women. Thus, the assumption about nondifferential "mortality" (or more correctly, the total effect of mortality and migration) is not completely satisfied.

The very high efficiencies obtained in this example are partly explained by the low mortality for ages below 40 years in the cohort of Danish women born in 1940. The average yearly "mortality" rate is 1.7 per 1000. Since the two estimation methods coincide when there is no mortality, it is not surprising that we get such high efficiencies in our case. To study the effect of the mortality further, we have also calculated the asymptotic relative efficiencies for situations where the $\{\alpha_r\}$

TABLE 1
Cumulative first marriage incidence rates and estimated first marriage intensities, with standard deviations, for women born in Denmark in 1940

Age	(1) Cumulative incidence rates $\hat{B}_r$ per 1000	(2) Standard deviation on $\hat{B}_r$ per 1000	(3) Occ/exp rates $\hat{\alpha}_r$ per 1000	(4) Incidence rate method $\tilde{\alpha}_r$ per 1000	(5) Standard deviation on $\hat{\alpha}_r$ per 1000	(6) Standard deviation on $\tilde{\alpha}_r$ per 1000	(7) Efficiency $\{(5)/(6)\}^2$
15	—	—	0.4	0.4	0.111	0.111	1.000
16	0.4	0.106	6.7	6.7	0.455	0.455	1.000
17	7.0	0.463	25.6	25.6	0.900	0.900	1.000
18	32.1	0.981	92.6	92.3	1.766	1.766	1.000
19	117.5	1.799	138.6	138.4	2.294	2.294	1.000
20	231.7	2.360	191.4	190.6	2.925	2.925	1.000
21	365.4	2.695	255.2	254.1	3.777	3.778	0.999
22	508.5	2.803	297.4	292.4	4.648	4.651	0.999
23	633.8	2.705	301.9	293.9	5.399	5.405	0.998
24	727.7	2.504	271.4	266.9	5.929	5.939	0.997
25	791.8	2.288	250.2	241.8	6.417	6.430	0.996
26	836.7	2.088	194.5	190.9	6.357	6.371	0.996
27	865.2	1.934	160.2	157.6	6.306	6.320	0.996
28	884.9	1.812	128.0	125.6	6.051	6.063	0.996
29	898.5	1.718	130.6	126.8	6.482	6.499	0.995
30	910.6	1.627	85.7	81.9	5.494	5.504	0.996
31	917.6	1.571	60.2	56.7	4.733	4.739	0.997
32	922.2	1.533	42.7	39.3	4.034	4.038	0.998
33	925.2	1.507	54.1	49.4	4.622	4.628	0.997
34	928.8	1.475	47.5	43.0	4.415	4.420	0.998
35	931.8	1.448	44.2	40.2	4.362	4.367	0.998
36	934.5	1.423	26.6	24.2	3.441	3.443	0.999
37	936.0	1.408	36.5	33.1	4.082	4.087	0.998
38	938.1	1.388	19.7	17.8	3.033	3.034	0.999
39	939.2	1.377	—	—	—	—	—

remain unchanged, but the mortality is increased by a factor 2, 4, 6, 8, or 10 for all ages. In all these cases the lowest efficiency was obtained for age 29, where it attained the values 0.990, 0.981, 0.972, 0.964, and 0.957, respectively. So the relative efficiencies will exceed 95 per cent, even when we increase the mortality by 1000 per cent. This suggests that in this particular example, the level of the mortality does not influence the efficiency of the incidence rate method (6.3) very much.

To explore further how the efficiencies depend on the values of the $\{\alpha_r\}$ and $\{\mu_r\}$, we have made some additional numerical computations. Since some cohorts are watched over their entire life span, and not only over a limited period as in the previous example, we have calculated relative efficiencies for a period of 70 years. For simplicity we assume in all these examples that $\alpha_r = \alpha$ and $\mu_r = \mu$ for all r, for some α and μ. The resulting efficiencies for $\alpha = 0.05, 0.10, 0.15$ and $\mu = 0.001, 0.01$ are shown in Fig. 4.

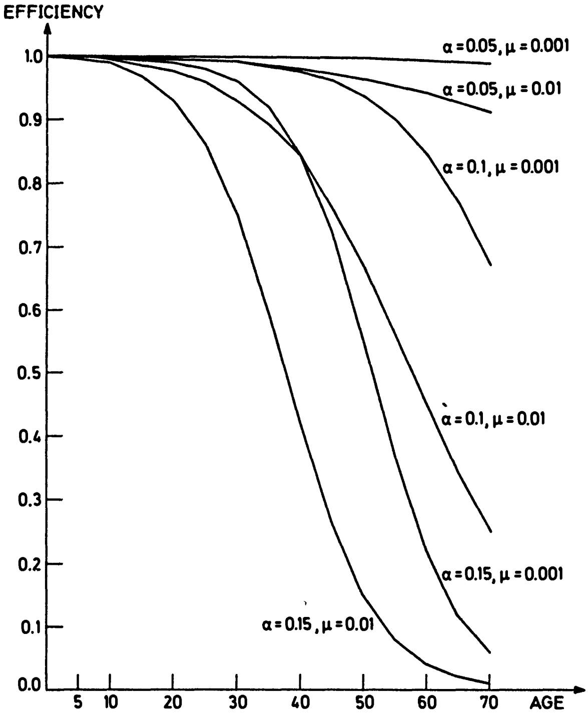

FIG. 4. *Asymptotic relative efficiencies of the incidence rate method for the first marriage model assuming constant intensities throughout all ages.*

The efficiencies decrease with increasing age for all the six cases considered, and the higher the values of α and μ are, the faster is the decrease. The value of α seems to be of great importance for the efficiency. The efficiencies are above 90 per cent for ages below 20 years. For higher ages the efficiencies may be low. Consider for example the case where $\alpha = 0.15$, $\mu = 0.001$. Here the efficiencies decrease sharply after age 30 years. The reason seems to be that at age 30 years, there will only be 1 per cent of the original cohort left in state 1. Therefore (5.1) will be a poor "estimator" for the true number at risk, which again makes $\tilde{\alpha}_r$ an unreliable estimator for α_r.

To summarize, the efficiencies for the incidence rate method (6.3) for the model in Fig. 1 seem to be rather high for younger ages, unless the parameters $\{\alpha_r\}$ take very large values. The efficiencies may become smaller in the higher age groups, where the number of individuals at risk is small. The level of mortality is also important, but its does not seem to influence the efficiencies as much as the values of the $\{\alpha_r\}$ do.

Acknowledgments. We are grateful to Jan M. Hoem for helpful comments and Martin Jacobsen for clarifying discussions. Comments from an anonymous referee greatly helped to improve the presentation of this paper. The present proof of the theorem in the Appendix is also due to his suggestion. Ørnulf Borgan was supported by The Norwegian Research Council for Science and the Humanities, the Association of Norwegian Insurance Companies and Johan and Mimi Wessmann's legacy. Part of his work was done during a visit to the Department of Statistics at Stanford University in 1982.

APPENDIX

Central limit theorem for the number of transitions between the states in a time-continuous Markov chain.

In this appendix we prove the central limit theorem for the number of transitions between the states in a time-continuous Markov chain, which was applied in Section 4. The Markov chain does *not* need to have the special structure considered in the main body of this paper.

THEOREM. *Consider n independent copies of the same time-continuous Markov chain on* $[0, z]$ *with a finite state space* $\mathbf{I}$ *and with transition intensities* $\{\alpha_{ij}(s)\}$ *and probabilities* $\{P_{ij}(s, t)\}$, *respectively. Assume that* $\int_0^z \alpha_{ij}(s)\,ds$ *is finite for all* $i, j \in \mathbf{I}$, $i \neq j$, *and denote the normalized number of transitions from i to j in* $[0, t]$ *by* $X_{ij}^{(n)}(t)$ *as in* (4.5). *Then* $X^{(n)} = (X_{ij}^{(n)};\ i, j \in \mathbf{I},\ i \neq j)$ *converges weakly to a mean zero Gaussian process* $X = (X_{ij};\ i, j \in \mathbf{I},\ i \neq j)$ *with covariance structure given by* (4.6).

The convergence takes place in the space $D^m[0, z]$ of m-dimensional $D[0, z]$ functions equipped with the Skorohod product topology (Billingsley, 1968), where m is the number of component processes.

PROOF. The theorem is a central limit theorem for i.i.d. versions of the multivariate Markov chain $W(t) = (K_{ij}(t), N_i;\ i, j \in \mathbf{I}, i \neq j)$ where $K_{ij}(t)$ and N_i now denote the values for a single individual. According to Hoem and Aalen (1978, formula (12)), $\operatorname{Cov}(K_{ij}(s), K_{kl}(t))$ equals (4.6). Therefore, by Hahn (1978, Theorem 3), in order to show tightness, it is sufficient to demonstrate the existence of nondecreasing continuous functions F_{ij} such that

$$\text{(A.1)} \qquad \operatorname{esssup}_{w(s)} E_s(K_{ij}(t) - K_{ij}(s))^2 \leq (F_{ij}(t) - F_{ij}(s))^\beta, \quad s \leq t,$$

holds for some $\beta > 1/2$, where E_s denotes the conditional expectation given $W(s) = w(s)$ or equivalently given $W(u) = w(u)$, $u \leq s$. Hahn's Theorem 3 concerns a one-dimensional process, but since a multivariate process is tight if each component is tight, (A.1) is the multivariate generalization of her tightness condition (i). To verify (A.1), fix i, j and s and note that by (3.1)

$$K_{ij}(t) - K_{ij}(s) = \int_s^t Y_i(u)\alpha_{ij}(u)\,du + M_{ij}(t) - M_{ij}(s).$$

Since $Y_i(u)$ equals zero or one (for a single individual) $\int_s^t Y_i(u)\alpha_{ij}(u)\,du \leq$

$\int_s^t \alpha_{ij}(u)\,du$. Moreover, $M_{ij}(t) - M_{ij}(s)$ and

$$(M_{ij}(t) - M_{ij}(s))^2 - \int_s^t Y_i(u)\alpha_{ij}(u)\,du$$

are zero mean martingales in $t \geq s$ by the counting process theory of Section 3. Thus

$$\begin{aligned} E_s(K_{ij}(t) - K_{ij}(s))^2 &= E_s\left(\int_s^t Y_i(u)\alpha_{ij}(u)\,du + M_{ij}(t) - M_{ij}(s)\right)^2 \\ &\leq 2E_s\left(\int_s^t Y_i(u)\alpha_{ij}(u)\,du\right)^2 + 2E_s(M_{ij}(t) - M_{ij}(s))^2 \\ &= 2E_s\left(\int_s^t Y_i(u)\alpha_{ij}(u)\,du\right)^2 + 2E_s\left(\int_s^t Y_i(u)\alpha_{ij}(u)\,du\right) \\ &\leq c\int_s^t \alpha_{ij}(u)\,du \end{aligned}$$

for some constant c since $\int_0^z \alpha_{ij}(u)\,du < \infty$. This gives (A.1) and the theorem has been proved. □

As pointed out to us by Richard Gill, the theorem is also a consequence of Kurtz (1983, Theorems 2.1 and 2.2).

REFERENCES

AALEN, O. O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726.

AALEN, O. O. and JOHANSEN, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scand. J. Statist.* **5** 141–150.

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10** 1100–1120.

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

BORGAN, Ø. and RAMLAU-HANSEN, H. (1983). Demographic incidence rates and estimation of intensities with incomplete information. Statistical Research Report no 2/83. Inst. of Math., Univ. of Oslo.

BRESLOW, N. and CROWLEY, J. (1974). A large sample study of the life table and product-limit estimates under random censorship. *Ann. Statist.* **2** 437–453.

FINNÄS, F. (1980). A method to estimate demographic intensities via cumulative incidence rates. *Theor. Popul. Biol.* **17** 365–379.

GILL, R. D. (1980). *Censoring and Stochastic Integrals.* Mathematical Centre Tracts **124**. Mathematisch Centrum, Amsterdam.

GILL, R. D. (1983). Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.* **11** 49–58.

HAHN, M. (1978). Central limit theorems in $D[0, 1]$. *Z. Wahrsch. verw. Gebiete* **44** 89–101.

HOEM, J. M. (1969). Purged and partial Markov chains. *Scand. Aktuar. Tidskr.* **52** 147–155.

HOEM, J. M. (1972). On the statistical theory of analytic graduation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 569–600.

HOEM, J. M. (1976). The statistical theory of demographic rates. A review of current developments (with discussion). *Scand. J. Statist.* **3** 169–185.

HOEM, J. M. (1978). Demographic incidence rates. *Theor. Popul. Biol.* **14** 329–337.

HOEM, J. M. and AALEN, O. O. (1978). Actuarial values of payment streams. *Scand. Actuarial J.* 38–47.

HOEM, J. M. and JENSEN, U. F. (1982). Multistate life table methodology: A probabilist critique. *Multidimensional Mathematical Demography* 154–264. Ed. Kenneth C. Land and Andrei Rogers. Academic Press, New York.

JACOBSEN, M. (1982). Statistical analysis of counting processes. *Lecture Notes in Statistics* **12**. Springer-Verlag, New York.

KURTZ, T. G. (1983). Gaussian approximations for Markov chains and counting processes. *Proc. 44th session ISI, Bull. I.S.I.*, **1** *Book 1*, 361–376.

LENGLART, E. (1977). Relation de domination entre deux processus. *Ann. Inst. Henri Poincaré* **13** 171–179.

NELSON, W. (1969). Hazard plotting for incomplete failure data. *J. Qual. Technol.* **1** 27–52.

REBOLLEDO, R. (1978). Sur les applications de la théorie des martingales a l'étude statistique d'une famille de processus ponctuelles. *Springer Lecture Notes in Math.* **636** 27–70.

REBOLLEDO, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. verw. Gebiete* **51** 269–286.

INSTITUTE OF MATHEMATICS
UNIVERSITY OF OSLO
POB 1053
BLINDERN
N-0316 OSLO 3
NORWAY

LABORATORY OF ACTUARIAL MATHEMATICS
UNIVERSITY OF COPENHAGEN
UNIVERSITETSPARKEN 5
DK-2100 COPENHAGEN Ø
DENMARK