KM23(1987)
pag 85-115

ON THE ESTIMATION OF MULTIDIMENSIONAL DEMOGRAPHIC MODELS WITH POPULATION REGISTRATION DATA

Nico Keilman *
Richard Gill **

ABSTRACT

In this paper the estimation of multidimensional demographic models and their underlying stochastic processes is investigated in situations where population registration data are available. With this kind of aggregate data, estimation by traditional methods is not possible. We look at two versions of the multidimensional model: the constant intensities model and the linear integration model. Some logical inconsistencies in the derivation of the latter are discussed. In particular, we argue that the linear integration model is not compatible with a Markov process. A new estimation algorithm for the estimation of the constant intensities model with population registration data is proposed. Some preliminary results on the mathematical and statistical properties of this method are given. The algorithm is applied to Dutch nuptiality data.

Keywords: Multidimensional mathematical demography, multistate life table, Markov process, population registration data, aggregate data, constant intensities, linear integration, linear survival function.

* Netherlands Interuniversity Demographic Institute (NIDI), P.O. Box 11650, 2502 AR The Hague, tel. 070-469482.

** Center for Mathematics and Computer Science (CWI), P.O. Box 4079, 1009 AB Amsterdam, tel. 020-5924076.

We are very grateful to Frans Willekens, who originally posed the problem and with whom we have had many stimulating discussions. Thanks are also due to Tonny Nieuwstraten for expertly typing the manuscript, and to Willemien Kneppelhout for assistance with the English text.

1. INTRODUCTION

Multidimensional demographic models describe transitions which individuals experience during their lifetime, as they pass from one state to another: for example, from being alive to being dead, from living in one region to residing in another, from being single to being married. Multidimensional life tables or multidimensional projection models are often viewed as parts of an underlying time-continuous Markov-model, which can be estimated with the aid of observed data. Even when the Markov assumption is not made explicitly, life tables are often computed and projections are often carried out as if the process were Markovian. In order to facilitate estimation of the model, simplifying assumptions are usually made. Demographers often employ the so-called "linear integration hypothesis" (also known as the assumption of a linear survival function) which follows from the assumption of a uniform distribution of events over age or time intervals. This linear integration hypothesis states that the total amount of time a group of individuals experience the risk of making a transition from one state to another (e.g. from being divorced to being remarried) in a certain age (time) interval, may be found by computing the simple average of the numbers of persons at risk (divorced persons) at the beginning, and at the end of the interval. However, it has been argued that such "linear integration models" may produce implausible results, including negative transition probabilities. Examples may be found in Hoem and Funck Jensen (1982, pp. 156, 157, pp. 198-201) and in Nour and Suchindran (1984, p. 325). As an alternative to the linear integration approach, an assumption often proposed is that of constant transition intensities over age or time intervals. This approach leads to estimators for the transition intensities which have good properties under quite weak conditions. But not all types of data are suitable for this approach.

In this paper we investigate the estimation of multidimensional demographic models and their underlying Markov processes when population registration data are available. It is argued that such data, in which individual information is aggregated over time as well as over individuals, do not permit maximum likelihood estimation using the assumption of constant intensities. As an alternative, the linear integration approach is discussed in detail and reasons for implausible results are given. Since neither the assumption of constant intensities nor the linear integration hypothesis facilitates the estimation by traditional methods of a time-continuous Markov process when population registration data are available, a new estimation algorithm is proposed. Some preliminary results on the mathematical and statistical properties of this method are given. The method is applied to Dutch nuptiality data.

The outline of this paper is as follows. Section 2 reviews a simple Markov process with two states and only one transition possibility. We discuss the type of data currently used for the estimation of such simple Markov processes and the estimation of the constant intensity model and of the linear integration model. Section 2 serves as an introduction to section 3, in which the same issues are studied for a general multidimensional increment-decrement model. Next, section 4 discusses the new estimation algorithm for population registration data and section 5 contains a numerical illustration with Dutch nuptiality data. We reanalyse the marital status model, containing five states, which was introduced by Schoen and Nelson (1974), and recently criticized by Nour and Suchindran (1984). The main findings of this paper are given in section 6, where they are placed in a proper perspective.

The framework chosen for our analysis is that of a continuous-time stochastic process, usually, but not always, assumed to be a Markov process. The reason for this choice was that many issues which have been discussed in the literature on deterministic multidimensional demographic models may be studied in greater depth by constructing an underlying stochastic process. However, we present our findings mainly in terms familiar to demographers; for technical details we refer the reader to the references.

2. A SIMPLE MORTALITY MODEL

In many statistical studies three elements are involved: there is a model describing some part of reality, an observational plan describing how the data relate to the model and a set of statistical algorithms determining how inferences are made from the data. This section discusses each of these three elements in the context of a simple Markov process with only two states and one transition.

2.1. The model

Let x denote a continuous duration parameter describing age or time. Consider an individual and let the random variable $I_i(x)$ take the value 1 if this person is in state i at duration x, and take the value 0 otherwise. Let i = 1 correspond to the state "alive" and i = 2 to the state "dead". The collection of random variables $I_i(x)$ for all considered values of x and i forms a stochastic process representing an individual's life history. The probability of being in state j at duration y, given that the individual was in state i at duration $x \leq y$ is

$$P_{ij}(x,y) = \Pr\{ I_j(y) = 1 \mid I_i(x) = 1\}$$

and the corresponding intensity is

$$\mu_{ij}(x) = \lim_{y \downarrow x} P_{ij}(x,y) / (y-x) , \quad i \neq j .$$

(We assume that $P_{ij}(x,y)$ is a smooth enough function of y for this limit to exist). In this pure mortality model we see that $P_{11}(x,x+t)$ equals the familiar survival probability ${}_tp_x$, whereas $P_{12}(x,x+t)$ is the probability of dying ${}_tq_x$ and $P_{22}(x,x+t) = 1$. The state "dead" is an "absorbing state" and $P_{21}(x,y) = 0$. Finally, $\mu_{12}(x)$ is the death intensity or force of mortality for age x, sometimes denoted as μ_x.

The process defined above is a Markov process. For, in a Markov process the conditional probability of a transition from i to j in a duration interval (x,y), given the complete history of the individual up to duration x, depends only on his state at duration x. For instance, for $z < x < y$

$$\Pr\{ I_j(y) = 1 \mid I_i(x) = 1, I_k(z) = 1\} = \Pr\{ I_j(y) = 1 \mid I_i(x) = 1 \} .$$

It can be readily verified that in this simple model the Markov property holds. For the more complex multidimensional increment-decrement model in section 3 it will be assumed that the process is Markovian.

We introduce the initial distribution

$$l_i(0) = \Pr\{ I_i(0) = 1 \}$$

and define

$$l_i(x) \quad = \Pr\{ I_i(x) = 1 \} .$$

In a life table $l_1(0) = l_0 = 1$ (usually multiplied by some factor like 10^5) is called the radix, whereas $l_1(x) = l_x$ is the probability of surviving to age x.

When $P_{ij}(x,y)$ is a so-called absolutely continuous function of y, the probability of surviving over the duration interval (x,x+t) is

$${}_tp_x \quad = \exp\{ - \int_0^t \mu(x+s)ds \}$$

with the obvious relation ${}_{t+s}p_x = {}_tp_x \cdot {}_sp_{x+1}$.

A useful variable is the model occurrence/exposure rate $m_{ij}(x,x+h)$, which in general is defined as

> the expected number of jumps from state i to state j in a given duration interval (x,x+h), divided by the expected total time of exposure in state i during the interval. (1)

In a Markov process one has in particular

$$m_{ij}(x,x+h) = \int_0^h l_i(x+s)\mu_{ij}(x+s)ds \Big/ \int_0^h l_i(x+s)ds \quad , \; i \neq j .$$

In our simple model, $m_{12}(x,x+h)$ coïncides with the familiar mortality rate ${}_hm_x$.

Several other quantities may be derived from μ_x, ${}_tp_x$, l_x and/or ${}_hm_x$. These need not to be given here, since they appear in standard text-books.

2.2. Observational plan

The data needed to estimate model parameters may be collected in many different ways. The observational plan describes a number of different aspects: have the data been obtained for individuals or only for groups of persons? Was the information collected over discrete time intervals or in a continous registration? Were the data used cohort data, period data or period-cohort data?

Have all the members of the target population been included or has only a sample been surveyed? Have the data been collected retrospectively or prospectively? These questions demonstrate that the observational plan acts as a kind of filter between what is potentially observable and what is actually observed (Hoem and Funck Jensen, 1980, p. 168).

The observational plan influences the statistical analysis in many subtle ways which need to be investigated for each separate case. The purpose of this paper is to investigate the estimation of time-continuous, finite state-space Markov processes using population registration data. Such data are found in countries where a population registration system exists. The observational plan for data of this type may be characterised as follows:

- data are obtained for groups of persons, not for individuals;
- the information is collected over discrete time intervals;
- the data consist of stocks (the population distributed over the states) at the end-points of the time intervals and flows (events) within these intervals. In particular, we have occurrences, but no exposures;
- when speaking in terms of the Lexis diagram, the data are of the period-cohort type. They apply to the behaviour of the members of a particular cohort during a certain period. The lifelines of these persons cover two successive age intervals;
- the data pertain to a complete (sub-)population;
- they are collected in a prospective manner, i.e. there is no selection of individuals according to the events of interest. For instance, persons dying during the particular interval under consideration should not be omitted from the study.

Such an observational plan is just one possibility out of a multitude of data collection strategies. Ledent and Rees (1986) discuss the construction of (deterministic) multiregional life tables using various data types. They compare life tables derived from so-called migration or movement data with those based on what they call migrant or transition data. The former data are derived from a registration of all changes in states (also called moves, direct transitions, events of jumps) in a given period, as in the case of a population registration system. The latter data are taken from population censuses, for instance, when respondents are asked to report their current state as well as the state they occupied at some earlier time ("transitions"). Lee, Judge and Zellner (1970) have investigated the situation in which a sequence of stock vectors (the distribution of a group of individuals over the states of the Markov chain) are available. Similar types of aggregate data are being handled by Kalbfleisch, Lawless and Vollmer (1983), by Van der Plas (1983) and by Kalbfleisch and Lawless (1984). Other observational plans are discussed by Hoem and Funck Jensen (1982, pp. 219-236) in the context of Markov processes. Their review includes transition data, retrospectively collected data, data involving a few recent events and data which involve occurrences with sufficient subspecifications by relevant type of risk, but in which the distribution of corresponding exposures over the same categories of risks is unobtainable ("incidence data"). Finally, we mention observational plans in which only a segment of the life history of each individual is collected. The treatment of censored data of this kind may be found in any text on life-history analysis.

2.3. Estimation procedures

This subsection discusses the statistical procedures which allow estimation of the model parameters of the simple mortality model in 2.1 using population registration data. The statistical properties of these estimators will be treated too.

Before estimation can be carried out, a few additional model assumptions have to be made. First, we assume that the model applies to a homogeneous population in which all individuals are subjected to the same intensity of mortality (Assumption 1). Second, one individual's behaviour is supposed to be independent of that of any other person in the group (Assumption 2). For Assumption 3 there are two common options. The first option is that we assume that the intensity of mortality μ_x is constant on a duration interval $(x,x+h]$. An alternative choice assumes that the probability of surviving l_t declines linearly for $x<t\leq x+h$. Equivalent to this second option is the assumption of a uniform distribution of events (deaths) over $(x,x+h]$. Estimation procedures for these two options will now be discussed in turn. We estimate these models in the simple case of one duration interval. Extensions to situations in which a number of successive intervals are studied (as in a life table), or in which several duration intervals are analysed simultaneously (as in a projection model) are straightforward.

A maximum likelihood estimator for μ_x in the case of a constant intensity can be derived as follows. Imagine we observe Y individuals and D of them die during an interval (0,h] (we drop the duration index x for the sake of convenience). Suppose that individual no. k dies during (0,h]. We then set a corresponding indicator function D_k equal to one. If he survives at time h, we write $D_k = 0$. The probability that we observe $D_k = 1$, assuming $\mu_t = \mu$ for $0<t\leq h$, equals

$$\Pr\{D_k = 1\} = P_{12}(0,h) = {}_hq = 1-{}_hp = 1-\exp(-\mu h)$$

whereas

$$\Pr\{D_k = 0\} = {}_hp = \exp(-\mu h) .$$

Therefore, the likelihood function based on all Y individuals may be written as

$$L = \{(1-\exp(-\mu h))\}^{D} \{\exp(-\mu h)\}^{Y-D}, \text{ with } D = \sum_{k=1}^{Y} D_k .$$

Hence, the estimator $\hat{\mu}$ which maximizes L is

$$\hat{\mu} = -\{\ln(1-D/Y)\} / h .$$

From mathematical statistics we know that for large Y, the distribution of $\hat{\mu}$ is approximately normal with mean μ and variance

$$V(\hat{\mu}_y) = -1/E\{\frac{\partial^2 \ln L}{\partial \mu^2}\}.$$

Therefore, $V(\hat{\mu}_y) = 1 / E\{{}_hp \cdot h^2 D/{}_hq^2\} = {}_hq / ({}_hp\, h^2 Y)$.

It should be noted that routine calculation of the ML-estimator $\hat{\mu}$ with data which cannot arise from the model results in undefined parameter values. For instance, in extreme situations, a population registration system could produce values D/Y exceeding one. This is the case when a large number of immigrants with high mortality risks leads to an excessive number of deaths. Of course, under these circumstances a simple closed model (a model which explicitly excludes the possibility of moving into or out of the system) presents an unrealistic picture of reality and it would be necessary to take account of immigration in the model. For a population which is closed for migration, $\hat{\mu}$

leads to proper transition probabilities. It is just possible to obtain $D/Y = 1$ and $\hat{\mu} = \infty$ with data from the model. However, if μ is finite, the probability of this eventuality tends to zero exponentially fast, as Y tends to infinity.

We now derive ML-estimators for the model parameters under the linear integration hypothesis. Assume that the survival probability may be written as $l(t) = 1 + bt$ $(0 \leq t \leq h,\ b \leq 0)$. Then we have, for the intensity, a hyberbolic function of time

$$\mu(t) = -b / (1+bt)$$

and for the probability of surviving at duration h

$$_hp = \exp\{ - \int_0^h \mu(s)ds \} = 1 + bh \; .$$

The likelihood function for Y individuals is

$$L = (1+bh)^{Y-D} \; (-bh)^D \; .$$

Therefore, a ML-estimator for b is $\hat{b} = -D/Yh$ and hence $_h\hat{p} = 1 - D/Y$ and $_h\hat{q} = D/Y$.

Since D has a binomial distribution with mean $_hqY$, we see that $\hat{b}$, $_h\hat{p}$ and $_h\hat{q}$ are unbiased estimators and that the variance of $\hat{b}$ equals

$$V(\hat{b}) = {}_hp \cdot {}_hq / (h^2Y) \; .$$

These results show that the ML-estimators $_h\hat{p}$ and $_h\hat{q}$ under the linear integration hypothesis are the same as those found when a constant intensity is assumed. This is caused by the fact that the likelihood function, expressed in terms of $_hp$ and $_hq$, is identical in the two models. In fact, the two models are identical in terms of the class of probability distributions they imply for observed quantities.

Hence in both models, when the extreme (and, according to the model, impossible) condition of $D/Y > 1$ is encountered, a negative survival probability will result. Therefore, when dealing with population registration data, it is not the choice of option for Assumption 3 that may produce anomalous estimates for this simple model, but rather the fact that immigration is omitted from the model.

The linear integration model produces ML estimates of ${}_hp$ and ${}_hq$ that coincide with ML estimates of the same quantities in the constant intensity model. Other corresponding variables differ from each other. For instance, for the constant intensity model the occurrence/exposure rate of mortality is equal to the corresponding intensity, i.e. $m = \mu$. For the linear integration model definition (1) leads to

$$m = \quad - b/(1 + \tfrac{1}{2}bh) ,$$

since $l(t) = 1 + bt$ and $\mu(t) = - b/(1+bt)$ for $0 \leqslant t \leqslant h$. This implies that for the constant intensity model the estimated occurrence/exposure rate of mortality equals

$$\hat{m} \quad = -\{ \ln(1 - D/Y)\} / h$$

whereas for the linear integration model it is

$$\hat{m} \quad = D / \{ h(Y-\tfrac{1}{2}D)\} \quad .$$

Denoting the former estimator by $\hat{m}_1$ and the latter by $\hat{m}_2$ we find

$$\tfrac{1}{2}h\hat{m}_2 \quad = \tanh(\tfrac{1}{2}h\hat{m}_1)$$

when $\hat{m}_1$ and $\hat{m}_2$ are estimated from the same data set (D,Y). In this expression, the hyperbolic tangent of u is defined by

$$\tanh(u) \quad = \{ \exp(u) - \exp(-u) \} / \{ \exp(u) + \exp(-u)\} \ .$$

Since for small u, tanh(u) may be approximated by $u - u^3/3$, we have

$$\hat{m}_2 \approx \hat{m}_1 - h^2(\hat{m}_1)^3/12 \ .$$

Hence there are only minor numerical differences between results produced when the two methods are applied to the same data in the case of short duration intervals and/or small mortality risks.

3. MULTIDIMENSIONAL MODELS

After the simple death model with two states and only one decrement we now turn to the more general case of several intercommunicating states. A number of names are used for these types of models: multidimensional models, increment-decrement models and multistate models. We shall treat them as being equivalent.

Much of what follows will be described in terms of $P_{ij}(x,y)$ and $l_i(x)$, as in section 2. Note that these two quantities are defined for general stochastic processes, not necessarily Markovian. In this section, i and j may take on values greater than two, provided they are integers and bounded.

One often assumes that the process under consideration is a continuous-time, finite state-space Markov process with absolutely continuous transition probabilities. Then the transition probabilities satisfy the Kolmogorov forward differential equations

$$\frac{\partial}{\partial y} P_{ij}(x,y) = - P_{ij}(x,y)\mu_j(y) + \sum_{k \neq j} P_{ik}(x,y)\mu_{kj}(y) , \tag{2}$$

with $P_{ij}(x,x) = 1$ for $i = j$ and $P_{ij}(x,x) = 0$ otherwise, and with $\mu_j = \sum_k \mu_{jk}(x)$, $k \neq j$. A complete knowledge (for all $x \leq y$) of the matrix $\mathbf{P}(x,y)$ with elements $P_{ij}(x,y)$ and the vector $\mathbf{l}(0)$ with elements $l_i(0)$ enables one to compute such model variables as the model occurrence/exposure rates, the expected duration in state j over the duration interval $[x,x+h]$ for those who were in state i at duration x, the mean duration elapsed since x for those who make a jump from i to j during $[x,x+h]$, the expected number of (i,j)-jumps during $(x,x+h)]$, and so on. Formulas for such indicators are given by Rogers (1975) for multiregional models, Krishnamoorthy (1979), Schoen and Land (1979), Willekens et al. (1982) for marital status models, by Hoem (1977), Willekens (1980), Schoen and Woodrow (1980), Brouard (1981) for working life tables, by Oechsli (1975), Chiang and Van den Berg (1982), Feichtinger and Lutz (1983) for fertility tables including parity, and surely by many others. Hoem and Funck Jensen (1982) give a comprehensive review of the increment-decrement model and its Markov process formulation.

The $\mathbf{P}(x,y)$ may be solved from the Kolmogorov differential equations when the matrix $\boldsymbol{\mu}(x)$ with elements $\mu_{ij}(x)$ $(i \neq j)$ is specified (and $\mu_{ii}(x) = - \mu_i(x)$). In some special situations a compact analytic solution for $\mathbf{P}(x,y)$ in terms of $\boldsymbol{\mu}(x)$ may be found. This is the case for any hierarchical model, in which all $\mu_{ij}(x)$'s are zero whenever $i > j$. The simple mortality model discussed in section 2 is an example of such a hierarchical model. Other special models for which an analytic solution exists are given by Hoem and Funck Jensen (1982, p. 178).

In general, an analytic solution for equation (1) is unknown (although we always have $\mathbf{P}(x,y) = \prod_{x \leqslant t \leqslant y} (\mathbf{I} - \boldsymbol{\mu}(t)dt)$ which may be computed as accurately as desired by numerical methods). Hence additional assumptions are required in order to find $\mathbf{P}(x,y)$. As we did with the simple mortality model, we shall here treat two cases: the assumption of constant intensities and the linear integration hypothesis.

The assumption of constant intensities figures prominently in statistical literature. Models of this type are known as "time-homogeneous Markov processes". They are much more simple than the general Markov process described by (1). Assume that $\mu_{ij}(x+t) = \mu_{ij}$ for $i \neq j$ and $0 < t \leqslant h$. Let $\mu_{ii} = -\sum_{i} \mu_{ij}$, $i \neq j$, and let $\boldsymbol{\mu}$ be the matrix with elements μ_{ij}. Then

$$\mathbf{P}(t) \qquad = \exp(\boldsymbol{\mu} t)$$

where $\mathbf{P}(t)$ denotes the matrix of transition probabilities with elements $P_{ij}(x,x+t)$, $0 \leqslant t \leqslant h$, see for instance Çinlar (1975, p. 255). (The exponential of a matrix is defined by the usual power series). Note that we adopt the notation generally used in the statistical literature in the definition of $\boldsymbol{\mu}$. This leads to row vectors with elements $l_i(x)$. It is much more convenient than the notation in demographic literature, which employs column vectors (see, for instance, Rogers and Ledent, 1976). In the latter notation one would use a matrix $\boldsymbol{\mu}^*$which is equal to $(-\boldsymbol{\mu})^T$ and arrive at a transposed transition matrix. With the results for the transition probabilities further calculations of expected sojourn times, survival probabilities and so on are straightforward.

The estimation method of the constant intensities model depends on the observational plan at hand. When complete life histories of all individuals are observed, any intensity μ_{ij} may be estimated by the corresponding observed occurrence/exposure rate $\hat{m}_{ij}$. The numerator of $\hat{m}_{ij}$ consists of all (i,j)-jumps recorded during the relevant interval; its denominator is the total amount of time spent in state i. Under quite weak conditions, the observed o/e rate $\hat{m}_{ij}$ is the Maximum Likelihood estimator of the unknown intensity μ_{ij}. It is asymptotically, normally distributed with known asymptotic variance (Aalen and Hoem, 1978, p. 97).

In many cases individual life histories are not available. For instance the data may instead consist of counts of individuals $N_{ij}(x,x+h)$ who were in state i at some exact duration x and in state j at duration x+h (transition data). Since the set $N_{ij}(x,x+h)$ is multinomially distributed, given the number $Y_i(x) = \sum_j N_{ij}(x,x+h)>0$ of individuals in state i at duration x, an ML-estimator of the transition probability $p_{ij}(x,x+h)$ is

$$\hat{p}_{ij}(x,x+h) \quad = N_{ij}(x,x+h) \; / \; Y_i(x)$$

(provided that the corresponding matrix **P** is of the form exp(Q) for some "intensity matrix" Q). Its variance is $p_{ij}(1-p_{ij}) \; / \; Y_i(x)$, which can be estimated when p_{ij} is replaced by $\hat{p}_{ij}(x,x+h)$. The assumption of constant intensities comes in when **P**(t) must be determined for durations t different from integer multiples of h. One may try to find a matrix **C** which satisfies the exponential relation

$$\mathbf{P}(h) \qquad = \exp(\mathbf{C}\ h)$$

for a given **P**(h) and next calculate functions derived from the transition intensities. Neither existence nor uniqueness of such a matrix **C** are guaranteed. Problems arising in this context were discussed by Singer (1981).

Still other observational plans are possible. For instance, Lee, Judge and Zellner (1970) give a full account of the situation in which a sequence of stock vectors with elements $Y_i(x)$ is available. They estimate the matrix **P**(h) by choosing the value which would have produced the best approximation of the observed stocks. Similar types of aggregate data are treated by Kalbfleisch, Lawless and Vollmer (1983), by Van der Plas (1983) and by Kalbfleisch and Lawless (1984). In this paper we discuss the estimation of the constant intensities model when we have stock data and flow data (direct transitions) from a population registration system. No estimator is available in the literature for this type of data, as will be argued in section 4. Therefore, a new estimation algorithm was developed. However, we shall first discuss the linear integration hypothesis.

Recall the forward Kolmogorov differential equations (2). Integrating these between y = x and y = x+h, noting that $P_{ij}(x,x) = 1$ for i = j and $P_{ij}(x,x) = 0$ for

$i \neq j$ and multiplying the result by $l_i(x)$ leads to a set of equations, which, added over i yields

$$l_j(x+h) \quad = l_j(x) - \sum_{k \neq j} d_{jk}(x,x+h) + \sum_{k \neq j} d_{kj}(x,x+h) \, ,$$

see Hoem and Funck Jensen (1982, p. 174). This so-called accounting equation for state j may be written simultaneously for all p states using matrix representation

$$\mathbf{l}(x+h) \quad = \mathbf{l}(x) + \mathbf{1}\,\mathbf{D}(x,x+h) \tag{3}$$

In this expression, $\mathbf{l}(x)$ and $\mathbf{l}(x+h)$ denote row vectors for the initial and the final distribution of the process, respectively. Further, $\mathbf{1}$ is a row vector of 1's and $\mathbf{D}(x,x+h)$ is a matrix with elements $d_{ij}(x,x+h)$, being the expected number of direct transitions from state i to state j during $(x,x+h]$, $i \neq j$ and $d_{ii}(x,x+h) = - \sum_{j \neq i} d_{ij}(x,x+h)$ (observe that property (3) holds in general for stochastic processes which exclude the possibility of leaving the system).

The model occurrence/exposure rate can be written as

$$m_{ij}(x,x+h) = d_{ij}(x,x+h) \Big/ \int_0^h l_i(x+t)dt = d_{ij}(x,x+h) \,/\, L_i(x,x+h) \, .$$

The variable $L_i(x,x+h)$ represents the total exposure time in state i during $(x,x+h]$. Define the row vector $\mathbf{L}(x,x+h)$ with elements $L_i(x,x+h)$ $(i = 1, 2, \ldots, p)$. Then the linear integration hypothesis involves the assumption

$$\mathbf{L}(x,x+h) \quad = \tfrac{1}{2}h \, \{ \mathbf{l}(x) + \mathbf{l}(x+h) \} \, . \tag{4}$$

In view of expression (3), a uniform distribution of the events $d_{ij}(x,x+t)$ for $x < t \leqslant x+h$ is a sufficient condition for (4) for a certain h. Note that an estimate of $\mathbf{L}(x,x+h)$ according to (4) may be obtained with knowledge of the stock vectors at durations x and x+h only.

The occurrence/exposure rates $m_{ij}(x,x+h)$ may be collected in a matrix $\mathbf{M}(x,x+h)$ with the same configuration as the matrix of transition intensities $\mathbf{C}$. Then (3) can be written as

$$\mathbf{l}(x+h) \quad = \mathbf{l}(x) \{ \mathbf{I} + \tfrac{1}{2}h \, \mathbf{M}(x,x+h) \} \; \{ \mathbf{I} - \tfrac{1}{2}h \, \mathbf{M}(x,x+h) \}^{-1} \; , \tag{5}$$

see Rogers and Ledent (1976). Note that in this derivation we did not use the Markov assumption. Because of (5), authors using the linear integration hypothesis estimate the matrix of transition probabilities by

$$\hat{\mathbf{P}}(h) = \{\mathbf{I} + \tfrac{1}{2}h\,\hat{\mathbf{M}}(x,x+h)\}\,\{\mathbf{I} - \tfrac{1}{2}h\,\hat{\mathbf{M}}(x,x+h)\}^{-1}, \qquad (6)$$

where $\hat{\mathbf{M}}(x,x+h)$ is the matrix of observed occurrence/exposure rates under the linear integration hypothesis and $\mathbf{I}$ is the identity matrix.

It has been pointed out that this approach contains a number of drawbacks (Hoem and Funck Jensen, 1982, p. 157, p. 160, pp. 194-201; Land and Schoen, 1982, pp. 316-320; Nour and Suchindran, 1984, p. 325; Waters, 1984, pp. 364-366). One is that $\hat{\mathbf{P}}(h)$, as given by expression (6), may contain negative elements. This can be the case, when $\sum_{j\neq i} m_{ij}(x,x+h) > 2/h$, or equivalently under the linear integration hypothesis, when $\sum_{j\neq i} d_{ij}(x,x+h) > l_i(x) + \frac{1}{2}\sum_{j\neq i} d_{ji}(x,x+h)$ for some i, see Appendix 1 for a proof and an illustration. This result generalises the findings of Ledent (1980, p. 555) for the two-state model. Hence negative estimated transition probabilities can occur in the linear integration model when the total number of direct transitions out of state i during $(x,x+h]$ exceeds the number initially present at duration x in state i by an amount of more than half the number of direct transitions into state i. This may be the case, for instance, when: (i) all individuals $l_i(x)$ leave state i before duration x+h and (ii) more than 50% of the individuals that enter (possibly re-enter) state i during $(x,x+h]$ leave state i once again before duration x+h. It will be clear that the chance of encountering the situation $\sum d_{ij}(x,x+h) > l_i(x) + \frac{1}{2}\sum d_{ji}(x,x+h)$ is greater, the longer the interval $[x,x+h]$ is, provided that μ_i (the "combined" intensity of leaving state i) is sufficiently large. Hence the linear integration hypothesis may lead to unrealistic model values, even when immigration is absent, and independently of the estimation algorithm. (We emphasise that immigration and emigration (and death) are to be taken into account by adding more states to the model). This constrasts with the findings for the simple mortality model described in section 2.3.

A more important point applies to the derivation of (6) from (5). Since $\mathbf{l}(x+h)$ and $\mathbf{l}(x)$ are vectors, the matrix $\mathbf{P}(h)$ cannot be solved from $\mathbf{l}(x+h) = \mathbf{l}(x)\mathbf{P}(h)$. Therefore, application of (6) implicitly assumes that the occurrence/exposure rates contained in $\mathbf{M}(x,x+h)$ are independent of the distribution of individuals over the states at duration x. In other words, we assume that the observed origin-independent rates $\hat{m}_{jk}(x,x+h)$ are not only equal to their model counterparts, but also to the origin-dependent model rates. These rates, denoted by $m_{ijk}(x,x+h)$, are defined as

> the ratio of the expected number of direct transitions from j to k, conditional upon being in state i at duration x to the expected total time spent in state j during $(x,x+h]$, conditional upon being in state i at duration x.

In the special case of a Markov process the $m_{ijk}(x,x+h)$ can be given as

$$m_{ijk}(x,x+h) = \int_0^h P_{ij}(x,x+s)\mu_{jk}(x+s)ds \Big/ \int_0^h P_{ij}(x,x+s)ds, \; \forall i, \; j \neq k \, . \qquad (7)$$

The assumption that origin-dependent rates and origin-independent rates are equal, which was discussed by Ledent (1980, p. 550-554) and Land and Schoen (1982, p. 292), leads to the estimator given in (6) when at the same time the *strong linear integration hypothesis* is fulfilled: for some suitable number b_{ijt} we assume that (7) holds *and* that

$$d_{ijk}(x,x+t) = b_{ijk} \cdot t \; , \quad 0<t\leq h \; ,$$

where $d_{ijk}(x,x+t)$ is the origin-dependent number of direct transitions from j to k. In the case of a Markov process this variable is equal to

$$d_{ijk}(x,x+t) = l_i(x)\int_0^t P_{ij}(x,x+s)\,\mu_{jk}(x+s\,;\,i,x)ds, \quad j \neq k \, .$$

However, as will be argued below, these conditions together (strong linear integration hypothesis and origin-dependent rates equal to origin-independent rates) rule out that the process is Markov!

On the other hand, the *weak linear integration hypothesis*, which "only" assumes that

$$d_{ij}(x,x+t) = b_{ij} \cdot t \, , \; 0<t\leq h$$

leads directly to condition (4).

In Appendix 2 we prove that a stochastic process can be constructed with a given initial distribution and a given matrix of direct transitions, for which the two conditions discussed above hold. Hence they are mutually consistent, but the two conditions together are not consistent with the Markov assumption. For the constant intensities model, model origin-dependent o/e-rates are automatically equal to their origin-independent counterparts (and to the corresponding intensities). See Hoem and Funck Jensen (1982, p. 184).

Statistical properties of $\mathbf{P}(h)$ given in (6) in the context of population registration data are not known but could in principle be derived. However, non-parametric inference may be drawn using the combinatorial matrix procedures suggested by Dow (1985). The key idea of this approach is as follows. Suppose one estimates a matrix $\hat{\mathbf{B}}$ with elements $\hat{b}_{ij}$ for the linear integration model. If a second matrix $\overline{\mathbf{B}}$ can be constructed, the empirically obtained matrix $\hat{\mathbf{B}}$ can be tested against this "target" matrix using combinatorial matrix procedures. Such "target" matrices may be derived from different data sets, different time periods, or else they may be obtained according to a priori theoretical conjectures.

4. A NEW ESTIMATION ALGORITHM FOR POPULATION REGISTRATION DATA

In the situation of population registration data, consider the duration interval $[0,h]$. Then our data consist of the vectors $\mathbf{Y}(0)$ and $\mathbf{Y}(h)$ in which the population is distributed over the p states of the model. We are also given the matrix $\mathbf{N}$ with observations on direct transitions from i to state j during $(0,h]$, with $N_{ii} = -\sum_{j\neq i} N_{ij}$. Similarly to (3), we have

$$\mathbf{Y}(h) = \mathbf{Y}(0) + \mathbf{1}\,\mathbf{N}\,. \tag{8}$$

No entirely satisfactory estimator for the multidimensional model for data of this kind is known, neither when constant intensities are assumed, nor with the linear integration hypothesis. For instance, with constant intensities, a ML-estimator of the matrix of transition intensities would be based on the probability distribution of the data $(\mathbf{Y}(0)\,,\mathbf{N})$. But the joint distribution of $\mathbf{Y}(0)$ and $\mathbf{N}$ is so intractable that a maximum likelihood estimator of $\boldsymbol{\mu}$ cannot be computed. Therefore, one usually takes recourse to the "working approximation" $\mathbf{L} \approx \hat{\mathbf{L}} = \frac{1}{2}h(\mathbf{Y}(0) + \mathbf{Y}(h))$. Hence one doesn't consider the linear integration hypothesis as part of the model, as in (4). Rather, it is used to obtain a numerical approximation for **L**. Next, one estimates $\boldsymbol{\mu}$ by $\hat{\boldsymbol{\mu}} = (\text{diag}\,\hat{\mathbf{L}})^{-1}\,\mathbf{N}$, where "diag" of a vector denotes the diagonal matrix with the corresponding elements of the vector on its diagonal. But this estimator is generally inconsistent, since $\mathbf{L} \approx \hat{\mathbf{L}}$ is only a first-order approximation of the exposure time in the constant intensities model. Though in most situations its bias will be small compared with its standard deviation, and in any case the whole Markov process setup is itself only a "working approximation" to reality, it is felt that it is a failure that mathematical statistics did not yet provide a good statistical solution for this very common situation. Therefore, one of us developed a new algorithm, which will described only briefly here. Details can be found in Gill (1986).

The statistical problem may be formulated as follows. We have stock data $\mathbf{Y}(0)$, $\mathbf{Y}(h)$ and flow data **N** for n individuals. We assume that each individual follows a Markov process during the interval $(0,h]$. These Markov processes are assumed to be independent of one another and time-homogeneous. We now have to estimate the matrix of transition intensities $\boldsymbol{\mu}$ on the basis of the observation of $\mathbf{Y}(0)$ and **N**, i.e. given the initial configuration and the total number of direct transitions during $(0,h]$. We assume that all other quantities, in particular the total exposure time $\int_0^h \mathbf{Y}(t)dt$, are not observed. We seek estimators which have good properties when n goes to infinity.

With the assumptions stated earlier we have

$$\mathbf{L}\boldsymbol{\mu} = \int_0^h \mathbf{l}(s)\,\boldsymbol{\mu}\,ds = \int_0^h \mathbf{l}(0)\exp(\boldsymbol{\mu}s)\boldsymbol{\mu}\,ds = [\,\mathbf{l}(0)\exp(\boldsymbol{\mu}s)\,]\Big|_0^h =$$

$$= \mathbf{l}(0)[\exp(\boldsymbol{\mu}h) - \mathbf{I}] = \mathbf{l}(h) - \mathbf{l}(0)\,.$$

Since the model is closed, the elements in each row of $\boldsymbol{\mu}$ add up to zero and the inverse of $\boldsymbol{\mu}$ does not exist. However, when there is at least one state to which all p states have access, then for given $\mathbf{l}(0)$ and $\boldsymbol{\mu}$ the equations in **L**,

$$\mathbf{L} = \int_0^h \mathbf{l}(0)\exp(\boldsymbol{\mu}s)ds \tag{9}$$

and

$$\mathbf{L}\boldsymbol{\mu} = \mathbf{l}(0)[\exp(\boldsymbol{\mu}h) - \mathbf{I}], \quad \mathbf{L}\,\mathbf{1}^T = h\,, \tag{10}$$

are equivalent. The condition for the equivalence of (9) and (10) is both necessary and sufficient (Gill, 1986). From now on we shall assume that this condition has been fullfilled. An example is when "dead" is one of the states of the Markov process. More complex cases can be handled by appropriate decompositions of the state space, cf. Funck Jensen (1982) and Gill (1986, Appendix III).

Let $\mathbf{D}$ denote the matrix with elements d_{ij} being the expected number of direct transitions from state i to state j during $(0,h]$, $i \neq j$ and $d_{ii} = - \sum_{j \neq i} d_{ij}$. Then $d_{ij} = \mu_{ij} L_i$, which we can rewrite (taking into account the diagonal elements of $\mathbf{D}$ and the definition of $\boldsymbol{\mu}$ given in section 3) as

$$\mathbf{D} = (\operatorname{diag} \mathbf{L})\ \boldsymbol{\mu} . \tag{11}$$

The approach proposed here is simply to equate the observed variables $\mathbf{Y}(0)$ and $\mathbf{N}$ to their expected values $n\,\mathbf{l}(0)$ and $n\,\mathbf{D}$ and to solve the resulting equations in $\mathbf{l}(0)$ and $\boldsymbol{\mu}$. This is equivalent to solving equations (10) and (11) considered for given $\mathbf{l}(0)$ and $\mathbf{D}$ (equal to $\mathbf{Y}(0)/n$ and $\mathbf{N}/n$) as equations in unknowns $\mathbf{L}$ and $\boldsymbol{\mu}$.

Various questions then arise.

1. When, for given $\mathbf{l}(0)$ and $\mathbf{D}$, do equations (9), (10) and (11) have a solution in $\mathbf{L}$ and $\boldsymbol{\mu}$, and when is the solution unique?
2. What is a good algorithm for finding a (the) solution?
3. What are the statistical properties of the resulting estimators?

Investigations of these ideas appear in Gill (1986). In general, a solution exists (question 1). Under a further quite simple condition the solution is unique; however, we can only verify this condition when p=2. When the process is hierarchical ($\mu_{ij} = 0$ for $i>j$) it can also be shown that there is exactly one solution. <u>We conjecture that there always exists exactly one solution</u>. This means that the first question has largely been answered, though not completely.

Regarding the second question, an obvious iteration method is based on cycling repeatedly through equations (10) and (11): first computing $\mathbf{L}$ for given $\mathbf{l}(0)$ and $\boldsymbol{\mu}$, then $\boldsymbol{\mu}$ for given $\mathbf{L}$ and $\mathbf{D}$. This resembles the EM-algorithm (cf. Dempster, Laird and Rubin, 1977), in that we compute in each cycle the expected total exposure time $n\,\mathbf{L}$ as a function of $\boldsymbol{\mu}$, given $\mathbf{Y}(0) = n\,\mathbf{l}(0)$; the EM-algorithm requires one to compute $n\,\mathbf{L}$ as a function of $\boldsymbol{\mu}$, given $\mathbf{Y}(0) = n\,\mathbf{l}(0)$ and $\mathbf{N} = n\,\mathbf{D}$. However, this superficial resemblance does not guarantee any convergence properties of the iterations. On the other hand, in every example yet considered these iterations converge quickly and independently of the starting value, to one limiting value. These findings lend support to both the usefulness of the iteration method and the conjecture about the uniqueness of the solution.

An alternative approach is to attempt to find a numerical solution, in $\mathbf{L}$, for given $\mathbf{l}(0)$, $\mathbf{l}(h)$ and $\mathbf{D}$, $\mathbf{l}(h)$ being defined as in (3), of the equations

$$\mathbf{l}(h) \quad = \mathbf{l}(0) \exp [(\operatorname{diag} \mathbf{L})^{-1} \mathbf{D} h] \ , \ \mathbf{L} \mathbf{1}^T = h$$

which is equivalent to solving $\mathbf{L}$ from

$$\mathbf{L} \quad = \int_0^h \mathbf{l}(0) \exp \ (\operatorname{diag} \mathbf{L})^{-1} \ \mathbf{D} s \ ds \ . \tag{12}$$

In all the examples we tried a standard quasi-Newton method worked excellently. Observe that the maximum likelihood estimator $\hat{\boldsymbol{\mu}}$ in the simple mortality model with constant intensity (section 2.3) is a special case of the solution of (9), (11) and (12).

For practical purposes, the answers to the first and second questions may be considered as satisfactory. As regards the third one, a satisfactory mathematical-statistical theory of the proposed estimators can be given, in which their asymptotic properties can be derived and in particular their asymptotic optimality (among estimators which use only the same aggregate data) can be proved. If $\hat{\mathbf{L}}$ is one solution for (12) for given $\mathbf{l}(0) = \mathbf{Y}(0)/n$ and $\mathbf{D} = \mathbf{N}/n$, then $(\hat{\mathbf{L}} - \mathbf{L})\, n^{\frac{1}{2}}$ is asymptotically multivariate, normally distributed with mean zero and with a covariance matrix which, in principle, can be computed. The same holds for $(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})n^{\frac{1}{2}}$. However, an estimate of the covariance structure of the statistical estimators $\mathbf{L}$ and $\hat{\boldsymbol{\mu}}$ would require rather complicated calculations. A useful practical solution is to use for $\hat{\boldsymbol{\mu}}$ the estimated covariance structure $\hat{\boldsymbol{\sigma}}^2$ for the occurrence/exposure rates applicable when the exposures $\mathbf{L}$ are observed too. This gives a lower bound to the asymptotic covariance matrix of the estimator actually used, since the matrix $\hat{\boldsymbol{\sigma}}^2$ gives the minimum variances when occurrences and exposures are observed. Therefore, we recommend using the off-diagonal elements of $\hat{\boldsymbol{\sigma}}^2 = (\operatorname{diag} n\, \hat{\mathbf{L}})^{-2}\, \mathbf{N}$ as a lower bound to, and a rough estimate of, the variances of the corresponding elements of $\hat{\boldsymbol{\mu}}$.

5. NUMERICAL APPLICATIONS

We consider here the marital status model with five states which was studied earlier by Schoen and Nelson (1974), Willekens et al. (1982) and by many others (see figure 1).

Our example is based on nuptiality patterns of Dutch females born in 1945, observed during the period 1965-1969. The data used are of the population-registration type, generously made available to us by the Netherlands Central Bureau of Statistics. The duration interval $[0,h]$ now represents the period between 1 January 1965 and 31 December 1969. The input data are given in table 1. Net immigration was disregarded and therefore the final population distribution **Y**(h) is not observed, but merely calculated. The initial number of persons in the state "dead" was taken as zero for the sake of convenience.

Figure 1. Marital status model

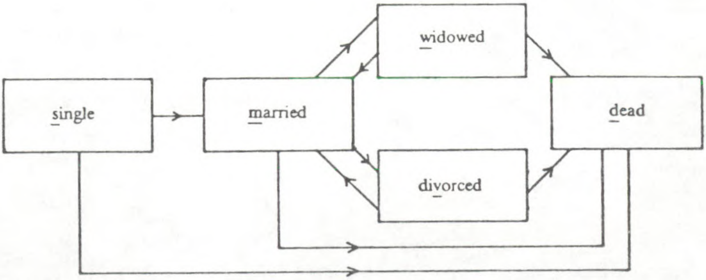

Table 1. Input data, nuptiality patterns of female birth cohort 1945, the Netherlands, 1965-1969

			s	m	w	v	d
Initial distribution	**Y**(0)		81773	10296	5	57	0
Direct transitions	**N**	single	58673	58540	0	0	133
		married	0	1116	136	921	59
		widowed	0	30	31	0	1
		divorced	0	342	0	343	1
		dead	0	0	0	0	0
Final distribution	**Y**(h)		23100	68092	110	635	194

Source: The Netherlands Central Bureau of Statistics.

Three different estimates of the matrix **P** of transition probabilities were computed, namely the "linear solution" $\hat{\mathbf{P}}_1 = (\mathbf{I} + \frac{1}{2}h\,\hat{\mathbf{M}}_1)(\mathbf{I} - \frac{1}{2}h\,\hat{\mathbf{M}}_1)^{-1}$, where $\hat{\mathbf{M}}_1 = (\text{diag } \hat{\mathbf{L}}_1)^{-1}\mathbf{N}$ and $\hat{\mathbf{L}}_1 = \frac{1}{2}h(\mathbf{Y}(0) + \mathbf{Y}(h))$; the "exact exponential solution" $\hat{\mathbf{P}}_2 = \exp(\hat{\mathbf{M}}_2\, h)$ where $\hat{\mathbf{M}}_2 = (\text{diag } \hat{\mathbf{L}}_2)^{-1}\mathbf{N}$ and $\hat{\mathbf{L}}_2$ is the (a?) solution to the equations $\mathbf{l}(h) = \mathbf{l}(0)\exp(\mathbf{M}_2\, h)$, $\mathbf{L}_2 . \mathbf{1}^T = h$; and the "approximate exponential solution" $\hat{\mathbf{P}}_3 = \exp(\hat{\mathbf{M}}_1\, h)$. Besides these three transition matrices we also present the corresponding exposure times per individual. From these results $\hat{\mathbf{M}}_1$ and $\hat{\mathbf{M}}_2$ can be easily computed for all 92131 women.

Table 2. Solutions: exposures and transition probability matrices

Marital status			s	m	w	v	d
Approximate exposures	$\hat{\mathbf{L}}_1$	(yrs)	2.8458	2.1271	.0031	.0188	.0053
Exact exposures	$\hat{\mathbf{L}}_2$	(yrs)	2.5189	2.4561	.0027	.0167	.0057
		s	28.25	70.88	.10	.56	.22
		m	0	98.02	.27	1.56	.15
Linear solution	$\hat{\mathbf{P}}_1$	w	0	40.67	57.60	.32	1.41
(per cent)		v	0	65.43	.09	34.23	.25
		d	0	0	0	0	1
		s	28.25	70.83	.10	.60	.22
Exact exponential		m	0	98.43	.22	1.21	.13
solution	$\hat{\mathbf{P}}_2$	w	0	44.81	53.30	.35	1.54
(per cent)		v	0	66.34	.10	33.32	.25
		d	0	0	0	0	1
		s	32.66	66.35	.11	.67	.21
Approximate exponential		m	0	98.11	.27	1.47	.15
solution	$\hat{\mathbf{P}}_3$	w	0	39.86	58.39	.37	1.38
(per cent)		v	0	61.94	.10	37.72	.24
		d	0	0	0	0	1

Since both $\hat{\mathbf{P}}_1$ and $\hat{\mathbf{P}}_2$ are exact solutions, we see that they fit the data exactly, i.e. $\mathbf{l}(0)\hat{\mathbf{P}}_1 = \mathbf{l}(0)\hat{\mathbf{P}}_2 = \mathbf{l}(h)$.

Table 2 shows differences in exposure times of 12-14 per cent except for the state "dead", where the exact exposure is 7 per cent higher than the approximate exposure. Transition probabilities obtained by the "linear solution" are very close to those obtained by the "exact solution", the largest difference being 4 percentage points (cf. the remarriage probability and the retention probability of widows). The latter is caused by the relatively high remarriage intensity of widows, for which a value of 12 per cent was found. Differences between "exact exponential" transition probabilities and "approximate exponential" transition probabilities are of the same magnitude, although they appear somewhat more frequently.

When considering the differences between the three methods one should, however, bear in mind that variances of estimated intensities for the states widowed and divorced are rather high. When using the off-diagonal elements of $\hat{\boldsymbol{\sigma}}^2 = (\text{diag } n\,\hat{\mathbf{L}}_2)^{-2}\mathbf{N}$, one finds, for example, for the remarriage intensity of widows (having a value of .12) a lower bound of the corresponding standard deviation of 0.02. The extremely short exposure times spent in the widowed state are responsible for these and other large standard deviations.

Besides the nuptiality example reported here, we also analysed a multiregional model. Observed population registration data on migration and mortality from the year 1980 for Dutch males born in 1956 were used for the 11 provinces of the Netherlands. Together with the state "dead" this resulted in a 12-state Markov model. Again, three solutions were computed. Bot they did not differ much, due to the low migration intensities of the males involved. For instance, the maximum value of the observed migration rates (approximated by the linear integration hypothesis) was 27 per thousand. However, it is worth mentioning that the convergence properties of the exact exponential algorithm were comparable to those of the 5-state marital status model.

6. CONCLUSIONS

In this paper we investigated the estimation of multidimensional demographic models when population registration data are available. Data of this kind consist of stocks and flows. The latter are aggregated over time and over individuals. In particular, this means that for each event information on the total number of occurrences (direct transitions) is known, but that we have no exposures. This means that the Markov model with constant intensities cannot be estimated by methods developed earlier. On the other hand, we argued that the linear integration model, based upon the assumption of a uniform distribution of events, may lead to transition probabilities outside the *[0,1]*-range. Conditions which lead to such unreasonable parameter values were discussed. Moreover, it was argued, that the derivation of an expression for the transition probabilities in the linear integration model requires the so-called strong linear hypothesis (for each type of direct transition, for each initial state subpopulation, direct transitions are uniformly distributed over time) and the assumption of origin-independent occurrence/exposure rates. But his rules out the Markov assumption. Since one makes projections by multiplying successive transition probability matrices, this makes the linear integration model aesthetically unsatisfactory.

Because of the deficiencies of the linear integration model and the fact that known estimation methods for the constant intensities model cannot be applied to population registration data, we presented a new estimation method. It is based on a Markov model with constant transition intensities. It uses an iterative algorithm for which a solution always exists; uniqueness of the solution could only be verified for some special cases (and convergence not at all). However, in all practical examples the iterations converged quickly to one limiting value, independently of the initial value. Therefore, we conjecture that the new estimation method always has exactly one solution. Asymptotic distributions of the proposed estimators were given.

Our findings indicate that in most practical cases there is little difference between the solution of the proposed method and that of the linear integration method. Moreover, the solution of a third approach, the approximate exponential model, is very similar to these two solutions. The approximate exponential model involves a Markov process with constant intensities. It is estimated by approximating the exposures using the linear integration hypothesis.

When a multidimensional demographic model and its underlying Markov process have to be estimated from population registration data we recommend the following approach. When transition intensities are moderate or small, it is a good strategy to approximate the exposures by the linear integration hypothesis and next to estimate the intensities of the Markov model (with constant intensities) using "observed-approximated" occurence/exposure rates. When particular emphasis is placed on an exact solution of the model, i.e. an accurate simulation by the model of the observed data, the linear integration hypothesis may be used. However, large transition intensities will produce an unrealisitc behaviour of this model, and its non-Markovian character makes a projection by multiplication of successive transition probability matrices aesthetically unsatisfactory. In such situations, application of the estimation algorithm proposed in this paper will be useful.

Finally, the main findings of this paper should be placed in proper prospective. First, we only considered two types of models: the constant intensities model and the linear integration model. Many more possibilities exist, for which the accompanying models exhibit different local behaviour of the intensity functions over the duration interval. For instance, Land and Schoen (1982) investigate Markov-generated multidimensional models with polynomial direct transitions, leading to rational polynomial intensity functions. Such an assumption may be useful for long duration intervals or strongly fluctuating intensities (e.g. infant mortality in the first few months after birth or seasonal patterns in marriage behaviour). But in most practical cases the solution of a more refined model of this type will be very close to that of the constant intensities model or the linear integration model.

Second, when small data sets are used, the randomness of estimates of model parameters may be important and a statistical approach should be used. However, for large data sets, estimates of standard deviations will generally be small compared to those of model parameters and functions thereof and a strictly deterministic analysis will often be sufficient.

Third, we want to stress the fact that a Markov model with constant intensities should often be seen as no more than a rough approximation of reality. Therefore, an "exact" statistical solution for this model is usually not very important. However, the contribution made in this paper has a methodological nature. We hope that it helps to clarify the controversy surrounding the linear integration hypothesis. Also, this paper illustrates the value of choosing a mathematical model with the aid of which elements of the probabilistic model, questions of numerical approximations and problems of data availability can be discussed systematically.

REFERENCES

Aalen, O.O. and J.M. Hoem (1978) Random time changes for multivariate counting processes, Scandinavian Actuarial Journal, 2, 81-101.

Berman, A. and R.J. Plemmons (1979) Nonnegative matrices in the mathematical sciences, Academic Press, New York etc.

Brouard, N. (1981) Les tables de vie active déduites de la mobilité par age, Population, 36(6), 1085-1103.

Chiang, C.L. and B.J. Van den Berg (1982) A fertility table for the analysis of human reproduction, Mathematical Biosciences, 62, 237-251.

Çinlar, E. (1975) Introduction to stochastic processes, Prentice-Hall Inc., Englewood Cliffs, New Jersey.

Dempster, A.P., N.M Laird and D.R. Rubin (1977) Maximum likelihood from incomplete data via the EM-algorithm (with discussion), Journal of the Royal Statistical Society B, 39, 1-38.

Dow, M.M. (1985) Nonparametric inference procedures for multistate life table analysis, Journal of Mathematical Sociology, 11, 245-263.

Feichtinger, G. and W. Lutz (1983) Eine Fruchtbarkeitstafel auf Paritätsbasis, Zeitschrift für Bevölkerungswissenschaft, 9(3), 363-376.

Funck Jensen, U. (1982) The Feller-Kolmogorov differential equation and the state hierarchy present in models in demography and related fields, Research Report no. 9, Section of Demography, University of Stockholm.

Gill, R. (1986) On estimating transition intensities of a Markov process with aggregate data of a certain type: occurrences but no exposures, forthcoming in Scandinavian Journal of Statistics.

Hoem, J.M (1977) A Markov chain model of working life tables, Scandinavian Actuarial Journal, 1-20.

Hoem, J.M. and U. Funck Jensen (1982) Multistate life table methodology: a probabilist critique, in: K.C. Land and A. Rogers (eds.), Multidimensional mathematical demography, Academic Press, New York etc., 155-264.

Kalbfleisch, J.D., J.F. Lawless and W.M. Vollmer (1983) Estimation in Markov models from aggregate data, Biometrics, 39, 907-919.

Kalbfleisch, J.D. and J.F. Lawless (1984) Least-squares estimation of transition probabilities from aggregate data, The Canadian Journal of Statistics-La Revue Canadienne de Statistique, 12(3), 169-182.

Krishnamoorthy, S. (1979) Classical approach to increment-decrement life tables: an application to the study of the marital status of United States females, 1970, Mathematical Biosciences, 44, 139-154.

Land, K.C. and R. Schoen (1982) Statistical methods for Markov-generated increment-decrement life tables with polynomial gross flow functions, in: K.C. Land and A. Rogers (eds.), Multidimensional mathematical demography, Academic Press, New York etc., 265-346.

Ledent, J. (1980) Multistate life tables: movement versus transition perspectives, Environment and Planning A, 12(5), 533-562.

Ledent, J. and P. Rees (1986) Life tables, in: A. Rogers and F. Willekens (eds.), Migration and settlement - a multiregional comparative study, Reidel Press, Dordrecht, 385-418.

Lee, T.C., G.G. Judge and A. Zellner (1970) Estimating the parameters of the Markov probability model from aggregate time series data, North-Holland Publ. Co., Amsterdam-London.

Nour, E.-S. and C.M. Suchindran (1984) The construction of multistate life tables: comments on the article by Willekens et al., Population Studies, 38(2), 325-328.

Oechsli, F. (1979) A general method for constructing increment-decrement life tables that agree with the data, Theoretical Population Biology, 16(1), 13-24.

Rogers, A. (1975) Introduction to multiregional mathematical demography, Wiley, New York.

Rogers, A. and J. Ledent (1976) Increment-decrement life tables: a comment, Demography, 13(2), 287-290.

Schoen, R. and V.E. Nelson (1974) Marriage, divorce and mortality: a life table analysis, Demography, 11(2), 267-290.

Schoen, R. and K.C. Land (1979) A general algorithm for estimating a Markov-generated increment-decrement life table with applications to marital status patterns, Journal of the American Statistical Association, 74(368), 761-776.

Schoen, R. and K. Woodrow (1980) Labor force status life tables for the United States, 1972, Demography, 17(3), 297-322.

Singer, B. (1981) Estimation of non-stationary Markov chains from panel data, in: S. Leinhardt (ed.), Sociological Methodology 1981, Jossey Bass, San Francisco.

Van der Plas, A.P. (1983) On the estimation of the parameters of Markov probability models using macro data, The Annals of Statistics, 11(1), 78-85.

Waters, H.R. (1984) An approach to the study of multiple state models, Journal of the Institute of Actuaries, 111(II), 363-374.

Willekens, F.J. (1980) Multistate analysis: tables of working life, Environment and Planning A, 12(5), 563-588.

Willekens, F.J., I. Shah, J.M. Shah and P. Ramachandran (1982) Multi-state analysis of marital status life tables: theory and application, Population Studies, 36(1), 129-144.

APPENDIX 1

This appendix contains the proof of the following lemma given in section 3. For a matrix $\mathbf{M} = (m_{ij})$ of occurrence/exposure rates and a duration interval of length h, the condition

$$-m_{ii} = \sum_{j \neq i} m_{ij} \leqslant 2 / h$$

implies that

$$\mathbf{P} = (\mathbf{I} + \tfrac{1}{2}h\,\mathbf{M})(\mathbf{I} - \tfrac{1}{2}h\,\mathbf{M})^{-1} \tag{A1}$$

is a transition matrix i.e. a matrix with nonnegative elements that add up to one rowwise.

We use the following notation: vectors are column vectors, $\mathbf{1}$ and $\mathbf{0}$ are vectors with all elements equal to 1 and 0 respectively, and for a vector $\mathbf{x}$ or a matrix A we write

$\mathbf{x} \geqslant \mathbf{0}$ $(A \geqslant 0)$ if the inequality holds elementwise;
$\mathbf{x} > \mathbf{0}$ if $\mathbf{x} \geqslant \mathbf{0}$ and $\mathbf{x} \neq \mathbf{0}$;
$\mathbf{x} \gg \mathbf{0}$ if we have elementwise strict inequality.

Furthermore, $Z^{pxp} = \{ \mathbf{A} = (a_{ij}) \in R^{pxp} : a_{ij} \leqslant 0,\ i \neq j \}$, i.e. Z^{pxp} is the set of all real (pxp) - matrices with non-positive off-diagonal elements

Berman and Plemmons (1979), in chapter 6 on "M-matrices", give 50 equivalent necessary and sufficient conditions for $A \in Z^{pxp}$ to be a "non-singular M-matrix". Two are:

(I 27) $\exists\, \mathbf{x} \gg \mathbf{0}$ with $\mathbf{A}\,\mathbf{x} \gg \mathbf{0}$, and
(N 38) $\mathbf{A}^{-1}$ exists and $\mathbf{A}^{-1} \geqslant \mathbf{0}$.

Now the proof is as follows:

Let $\mathbf{M}$ be a pxp intensity matrix, that is $-\mathbf{M} \in Z^{pxp}$ and $\mathbf{M}\,\mathbf{1} = \mathbf{0}$. If $-m_{ii} \leqslant 2/h\ \forall i$, then $(\mathbf{I} + \frac{1}{2}h\,\mathbf{M})$ is a transition matrix. The product of two transition matrices is a transition matrix, so it suffices to show that $(\mathbf{I} - \frac{1}{2}h\,\mathbf{M})^{-1}$ exists and is a transition matrix. But we have $\mathbf{I} - \frac{1}{2}h\,\mathbf{M} \in Z^{pxp}$ and $\mathbf{1} \gg \mathbf{0}$. Therefore $(\mathbf{I} - \frac{1}{2}h\,\mathbf{M})\mathbf{1} = \mathbf{1} \gg \mathbf{0}$. So by the equivalence of conditions (I 27) and (N 38), we find that $(\mathbf{I} - \frac{1}{2}h\,\mathbf{M})^{-1}$ exists and that $(\mathbf{I} - \frac{1}{2}h\,\mathbf{M})^{-1} \geqslant \mathbf{0}$. Since $(\mathbf{I} - \frac{1}{2}h\,\mathbf{M})\mathbf{1} = \mathbf{1}$ we find by premultiplying with $(\mathbf{I} - \frac{1}{2}h\,\mathbf{M})^{-1}$ that $\mathbf{1} = (\mathbf{I} - \frac{1}{2}h\,\mathbf{M})^{-1}.\mathbf{1}$. Hence $(\mathbf{I} - \frac{1}{2}h\,\mathbf{M})^{-1}$ is a transition matrix and so is $(\mathbf{I} + \frac{1}{2}h\,\mathbf{M})(\mathbf{I} - \frac{1}{2}h\,\mathbf{M})^{-1}$. This completes the proof.

It should be noted, that in the linear integration model

$$m_{ij}(x,x+h) \;=\; 2d_{ij}(x,x+h) \;/\; \{h(l_i(x) + l_i(x+h)\} \;.$$

Therefore, the condition $-m_{ij}(x,x+h) = \sum_{j\neq i} m_{ij}(x,x+h) \leq 2/h$ is in this model equivalent to

$$\sum_{j\neq i} d_{ij}(x,x+h) \leq l_i(x) + l_i(x+h) \;.$$

But $l_i(x+h)$ equals $l_i(x) - \sum_{j\neq i} d_{ij}(x,x+h) + \sum_{i\neq j} d_{ji}(x,x+h)$ and therefore $-m_{ii}(x,x+h) \leq 2/h$ is equivalent to

$$\sum_{j\neq i} d_{ij}(x,x+h) \leq l_i(x) + \tfrac{1}{2} \sum_{i\neq j} d_{ji}(x,x+h) \;. \qquad \text{(A2)}$$

Hence, when in the linear integration model condition (A2) is not fulfilled, then the matrix **P** according to (A1) is not necessarily a transition matrix.

As an illustration, consider a small part of the data set given by Schoen and Nelson (1974) which has recently been used by Nour and Suchindran (1984) to illustrate the occasional breakdown of the linear integration model (A1). The latter authors started from the following matrix **M** belonging to a five-state nuptiality model for the U.S. male population in the age interval (20, 25) in 1960:

$$\mathbf{M}(20,\,25) = \begin{array}{c} \left[\begin{array}{rrrrr} -.2313 & .2291 & 0 & 0 & .0022 \\ 0 & .0172 & .0007 & .0153 & .0012 \\ 0 & .1551 & .1603 & 0 & .0052 \\ 0 & .4860 & 0 & .4897 & .0037 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \\ \begin{array}{ccccc} (1) & (2) & (3) & (4) & (5) \end{array} \end{array}$$

(1) Never married
(2) Currently married
(3) Widowed
(4) Divorced
(5) Dead

(Note that we adapted their notation to ours). It may be observed that the diagonal element for divorced males $-m_{44}$ exceeds $2/h = .40$. Indeed, the corresponding transition matrix **P** computed according to (A1) is

$$\mathbf{P} = \begin{bmatrix} .2672 & .7104 & .0009 & .0122 & .0093 \\ 0 & .9577 & .0025 & .0337 & .0061 \\ 0 & .5419 & .4285 & .0093 & .0203 \\ 0 & 1.0694 & .0013 & -.0824 & .0117 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

which is not a transition matrix. This is caused by the value of the matrix of direct transitions

$$\mathbf{D}(20, 25) = \begin{bmatrix} -40176 & 40043 & 0 & 0 & 133 \\ 0 & -6537 & 373 & 5971 & 193 \\ 0 & 146 & -148 & 0 & 2 \\ 0 & 4009 & 0 & -4021 & 12 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and that of the initial distribution

$$\mathbf{l}(20) = (54177 \quad 41955 \quad 59 \quad 544 \quad 3265) \; .$$

We observe a number of remarrying or dying divorced males (4021) which exceeds the initial number of divorcees (544) by more than half the number of new divorcees (5971). This may be due to estimations carried out by Schoen and Nelson which were necessary because their initial data were partly defective.
They consider the accuracy and adequacy of their data and state that the number of divorced seems to be underreported and that their remarriage figures are probably high (Schoen and Nelson, 1974, p. 289). On the other hand, even when accurate figures could have been obtained, unrealistic remarriage patterns are likely to have been produced by the linear integration model, for remarriage rates may well have exceeded 40% for young divorced males in the 1960's. The constant intensities model will cure this anomaly; as an alternative, one may consider the linear integration model for single ages.

APPENDIX 2

In this appendix we prove that under a natural condition the two conditions (i) the strong linear integration hypothesis, and (ii) origin-independent occurrence/exposure rates, are consistent, but are not consistent together with the Markov assumption. Suppose $\mathbf{l}(0)$, $\mathbf{D}$ and $\mathbf{l}(1) = \mathbf{l}(0) + \mathbf{1}\,\mathbf{D}$ are given, such that the transition matrix $\mathbf{P} = (\mathbf{I} + \frac{1}{2}\mathbf{M})(\mathbf{I} - \frac{1}{2}\mathbf{M})^{-1}$ is a stochastic matrix, with $\mathbf{M}$ a matrix of occurrence/exposure rates for the interval $[0, 1]$ defined as $\mathbf{M} = (\text{diag } \mathbf{L})^{-1}\,\mathbf{D}$. Now a (stochastic) jump process can be constructed with initial distribution $\mathbf{l}(0)$ and expected number of direct transitions $\mathbf{D}$, such that for each initial state i the (weak) linear integration hypothesis holds and the occurrence/exposure rates are the same. However, this process is not Markovian. The argument is as follows.

Since we know $\mathbf{P}$, we also know $\mathbf{l}^{(i)}(1)$, that is the final distribution conditional on state i at duration 0. For the rows of $\mathbf{P}$ are the conditional distributions of "state at duration 1" conditional on "state at duration 0". Now condition (i) implies that $\mathbf{L}^{(i)} = \frac{1}{2}(\mathbf{l}^{(i)}(0) + \mathbf{l}^{(i)}(1))$ for each i. Here $\mathbf{l}^{(i)}(0)$ is the vector with value $l_i(0)$ in the i-th element and zeros otherwise, whereas $\mathbf{L}^{(i)}$ is the vector of exposures conditional on being in state i at duration 0. Furthermore, condition (ii) means that the matrix of occurrence/exposure rates conditional on state i at duration 0, denoted by $\mathbf{M}^{(i)}$, is independent of i and therefore $\mathbf{M}^{(i)} = \mathbf{M}$ for each i. Then we compute a matrix with direct transitions conditional on state i at duration 0 as $\mathbf{D}^{(i)} = (\text{diag } \mathbf{L}^{(i)})\mathbf{M}$. For each i a jump process exists with this $\mathbf{l}^{(i)}(0)$ and $\mathbf{D}^{(i)}$ which satisfies the (weak) linear integration hypothesis, namely the process with initial distribution $\mathbf{l}^{(i)}(0)$ and time-dependent (hyperbolic) intensities $\mu^{(i)}(t) = \text{diag } (\mathbf{l}^{(i)}(0) + t\,\mathbf{1}\,\mathbf{D}^{(i)})^{-1}\,\mathbf{D}^{(i)}$. Thus we can construct a stochastic process with properties described by conditions (i) and (ii), which means that these conditions are mutually consistent. But any process satisfying (i) and (ii) cannot be Markovian. For, conditions (i) and (ii) together with the Markov assumption lead to hyperbolic intensities, $\mu^{(i)}(t)$ given above, which are different for each initial state i, whereas for a Markov process intensities are independent of initial states.

Ontvangen: 15-08-86
Geaccepteerd: 28-10-86