Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

R.D. Gill, N. Keilman

On the estimation of multidimensional demographic models
with population registration data

On the Estimation of Multidimensional Demographic Models

with Population Registration Data

R.D. Gill

Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

N. Keilman

Netherlands Interuniversity Demographic Institute (NIDI)
P.O. Box 11650, 2502 AR The Hague, The Netherlands

In this paper the estimation of multidimensional demographic models 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 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.

TABLE OF CONTENTS   page

LIST OF NIDI WORKING PAPERS

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), namely the assumption of a uniform distribution of events over age or time intervals. This linear integration hypothesis implies that the total amount of time a large 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 these assumptions) under quite weak conditions. But not all types of data can be handled by this approach.

In this paper we investigate the estimation of multidimensional demographic models when population registration data are available. Although the registration takes place on the level of the individual, most statistical bureaus make them available in the form of tables (aggregated over individuals and applying to a certain period) only. It is argued that such aggregated data do not permit maximum likelihood estimation under the assumption of constant intensities. As an alternative, the linear integration approach is discussed in detail and reasons for implausible results are given. In particular we argue that a rather stronger form of this hypothesis than that given above is needed to justify the approach.

Since neither the assumption of constant intensities nor the linear integration hypothesis facilitates the estimation by traditional methods when population registration data are available, a new estimation algorithm for the constant intensities model is proposed. This method has recently appeared in a paper for mathematical statisticians (Gill, 1986). One of the main purposes of the present paper is to make the new estimator better known among demographers, and to add to the earlier results. 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 multidimensional demographic models, considered from a stochastic process point of view. We introduce the general random jump process (section 2.1), the more specific Markov jump process and the even more specific constant intensities Markov process (section 2.2). Then we discuss the general jump process in which the linear integration hypothesis is assumed in both a weak and a strong form (section 2.3). Section 3 contains a description of the type of data we focus on in this paper: aggregate data from population registration systems. Estimation algorithms for the linear integration hypothesis model and the constant intensities model are given in sections 4 and 5 respectively. A numerical example with Dutch nuptiality data is given in section 6. 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 7, where they are placed in a proper perspective.

The framework chosen for our analysis is that of a continuous-time random jump process, that is a stochastic process describing instantaneous jumps made by one individual between discrete states. The reason for this choice was that many issues which have been discussed in the literature on deterministic multidimensional demographic models may be studied with greater precision by considering an underlying individual-level stochastic process. In particular, this separates the issues of modelling a conceptually unlimited population from estimation with a finite sample. However, we present our findings mainly in terms familiar to demographers; for technical details we refer the reader to the references.

2. MULTIDIMENSIONAL DEMOGRAPHIC MODELS

2.1. Random jump processes

Let $x \geq 0$ denote a continuous duration variable describing age or time and let $E = \{1,...., p\}$ be a collection of discrete states. Consider a random individual jumping from one state to another as x varies; we suppose the individual stays in each state a positive length of time and then jumps instantaneously to a new one, making a finite (possibly zero) number of jumps in each bounded time interval. The system is supposed to be closed (one state is always being occupied; immigration and emigration should be modelled by adding extra states to the system). We can describe the process by the collection of random variables $I_i(x)$ (i = 1,...,p; $x \geq 0$) such that $I_i(x) = 1$ if the individual is in state i at time x+ and zero otherwise.

From this process various probabilities and expectations can be defined. For instance the <u>transition probability</u> of being in state j at duration y, given the individual was in state i at duration x <y is

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

If this conditional probability is a sufficiently smooth function of y there is a corresponding <u>mean intensity</u>.

$$\mu_{ij}(x) = \lim_{h \downarrow 0} P_{ij}(x,x+h) / h \quad , \quad i \neq j. \tag{2}$$

This intensity equals the mean of the <u>stochastic intensity</u> of making a jump from i to j in (x,x + h) given the complete past history of the process (and that it is in i at duration x).

Note that these transition probabilities and intensities are well defined whether or not the jump process is Markovian (i.e. its behaviour at times $y \geq x$ given its history up to duration x only depends on the state occupied at duration x, and not on any information prior to this). However it is only in the Markovian case that they satisfy the well known Chapman-Kolmogorov equations and the Kolmogorov differential equations (discussed in the next subsection).

Some further key quantities are the initial distribution

$$l_i(0) = \Pr \{ I_i(0) = 1 \} \tag{3}$$

and the distribution at duration x

$$l_i(x) = \Pr \{ I_i(x) = 1 \} . \tag{4}$$

We gather the elements $l_i(x)$ together into a row vector $l(x)$, and the transition probabilities $P_{ij}(x,y)$ into a matrix $P(x,y)$. Defining $\mu_{ii}(x) = -\sum_{j \neq i} \mu_{ij}(x)$ we also collect the intensities $\mu_{ij}(x)$ into a matrix $\mu(x)$. We can also call $l(x)$ the (expected) stock vector at duration x, per individual.

We have the following result, showing how the probability distribution over the states is transformed by the transition probabilities:

$$l(y) = l(x) P(x,y). \tag{5}$$

This is proved by simply writing out the definition of the conditional probabilities $P_{ij}(x,y)$ as a ratio of two unconditional probabilities.

We also consider the expected total time spent in each state (exposure time) over various periods, and the expected number of (direct) jumps between each two states in a given period.

Because the individual spends an amount of time $\int_x^y I_i(s)ds$ in state i between durations x and y, we have for its expected value

$$L_i(x,y) = \int_x^y l_i(s)ds. \tag{6}$$

We let $d_{ij}(x,y)$ denote the expected number of jumps between i and j in the period (x,y), for $i \neq j$. We define $d_{ii}(x,y) = - \sum_{j \neq i} d_{ij}(x,y)$, minus the expected number of jumps **out** of state i. From these we form the row vector $L(x,y)$ and the matrix $D(x,y)$.

Since the difference between $I_i(y)$ and $I_i(x)$ equals the number of jumps **into** i minus the number of jumps **out** of i during $(x,y]$, on taking expected values we obtain

$$l_i(y) = l_i(x) + \sum_{j \neq i} d_{ji}(x,y) - \sum_{j \neq i} d_{ij}(x,y). \tag{7}$$

In matrix notation this can be written as

$$l(y) = l(x) + l\ D(x,y) \qquad (8)$$

where l is a row vector of one's. This is called the <u>accounting equation</u> (see Hoem and Funck Jensen, 1982, p. 174). Under mild regularity conditions[1] there is an interesting relation between the derivative of $P_{ij}(x,x+h)$ with respect to h (at h = 0) and the derivative of $d_{ij}(x,x+h)$ with respect to h at h = 0, namely

$$\lim_{h \downarrow 0} d_{ij}(x,x+h)\,/\,h = l_i(x)\,\mu_{ij}(x),\ i \neq j. \qquad (9)$$

Intuitively, when h is small, there can only be one jump (at most) from i to j in the interval $(x,x+h]$, and this corresponds to the individual being in state i at duration x and in state j at duration x + h. The quantity $l_i(x)\,P_{ij}(x,x+h) \approx l_i(x)\mu_{ij}(x)\,h$ is the probability of this occurrence, which must be approximately equal to the expected value of the number of jumps from i to j in $(x,x+h]$.

Further interesting quantities are the <u>occurrence/exposure rates</u> $m_{ij}(x,x+h)$, defined as the expected number of jumps from i to j in $(x,x+h]$ divided by the expected length of time (exposure to the risk of an $i \rightarrow j$ jump) spent in i during this duration interval. Thus

$$m_{ij}(x,x+h) = d_{ij}(x,x+h)\,/\,L_i(x,x+h). \qquad (10)$$

By (9), $d_{ij}(x,x+h) = \int_x^{x+h} l_i(s)\,\mu_{ij}(s)ds$, so together with (6) we have

$$m_{ij}(x,x+h) = \frac{\int_x^{x+h} l_i(s)\ \mu_{ij}(s)ds}{\int_x^{x+h} l_i(s)ds}. \qquad (11)$$

Thus $m_{ij}(x,x+h)$ is a weighted average of the mean intensity $\mu_{ij}(s)$ over the duration interval; weighted by the probability of being in i at each duration. Moreover

$$\lim_{h \downarrow 0} m_{ij}(x,x+h) = \mu_{ij}(x). \qquad (12)$$

Later we will need further quantities such as <u>origin dependent</u> expected numbers of jumps, expected exposure, and occurrence/exposure rates. Here we condition on a certain state being occupied at the start of the duration interval; thus $d_{(i)jk}(x, x + h)$ is the expected number of jumps from j to k in $(x, x + h]$ given the individual was in i at duration x, $L_{(i)j}(x, x + h)$ is the expected exposure in j given the same event, and

$$m_{(i)jk}(x, x + h) = d_{(i)jk}(x, x + h) / L_{(i)j}(x, x + h).$$

None of these collections of quantities fully determine the probabilistic behaviour of the individual. For this we need to know all the probabilities of being in any sequence of states at any finite sequence of durations x_i or all the conditional distributions of times and types of the k'th jump (k = 1,2,...) given the times and types of the previous ones (together with the initial distribution l(0)).

2.2. Markov jump processes

In a Markov jump process we assume that the probability distribution of the development of the process from some duration x given its complete history up to x only depends on the state occupied at x. This is a very severe restriction and puts a lot of interesting structure into the quantities we have been describing. The transition probability matrices satisfy the Chapman-Kolomogorov equations

$$P(x,y)\, P(y,z) = P(x,z) \tag{13}$$

from which we get the Kolomogorov forward differential equations

$$\frac{\partial}{\partial y}\, P(x,y) = P(x,y)\mu(y). \tag{14}$$

Since the probability the process occupies states i_0, i_1,....., i_r at durations 0, x_1,..., x_r is, under the Markov assumption, $l_{i_0}(0)\, P_{i_0 i_1}(0,x_1)....\, P_{i_{r-1} i_r}(x_{r-1}, x_r)$, we see that l(0) and the P(x,y) completely describe the process. Moreover by (14) the P(x,y) are determined by the intensities $\mu(x)$ (here I is the identity matrix):

$$P(x,y) = \prod_{s=x}^{y} (I + \mu(s)ds).$$

(15)

(Here we use product-integral notation; see e.g. Johansen, 1986; Gill and Johansen, 1987). Thus the initial distribution $l(0)$ and the intensities $\mu(x)$ determine the whole process and hence all the quantities we described in Subsection 2.1. and others: the 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 in the demographic literature 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.

In some special situations a compact analytic solution for $P(x,y)$ in terms of $\mu(x)$ may be found. This is the case for any hierarchical model, in which all $\mu_{ij}(x)$'s are zero whenever $i > j$. Other special models for which an analytic solution exists are given by Hoem and Funck Jensen (1982, p. 178).

However for many purposes the general Markov model is still too general, and a further restriction is made: the assumption of (piece-wise) constant intensities. This gives us a so-called time-homogeneous Markov process; it figures prominently in the statistical literature.

Assume that $\mu_{ij}(x+t) = \mu_{ij}$ for $i \neq j$ and $0 < t \leqslant h$. Let again $\mu_{ii} = -\sum_{i} \mu_{ij}$, $i \neq j$, and let μ be the matrix with elements μ_{ij}. Then

$$P(t) = P(x,x + h) = \exp(\mu t) = \prod_{0}^{t} (I + \mu \, ds)$$

(16)

where $P(t)$ 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 μ. This leads to row vectors
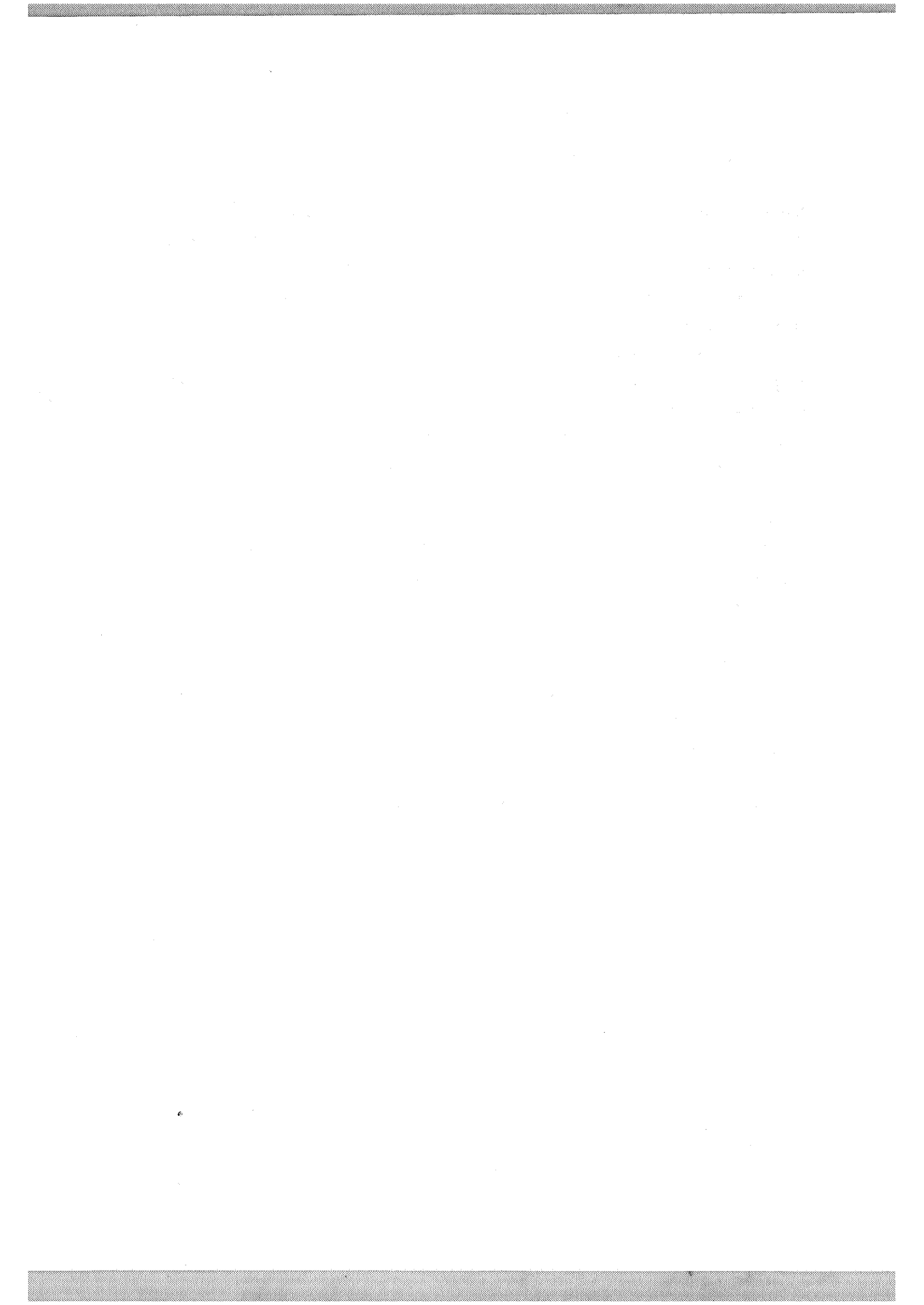
Also we have

$$L(x,y) \, 1^T = y-x. \tag{19b}$$

For given μ , (19a) and (19b) give a collection of linear equations from which L can be easily calculated.

Note that in a constant intensities Markov process the origin dependent occurrence/exposure rate $m_{(i)jk}(x,x + h)$, the origin-independent rate $m_{jk}(x,x + h)$, and the intensity μ_{jk} are all equal. This gives the result

$$P(x,x + h) = \exp(h \, M). \tag{20}$$

2.3. The linear integration hypothesis

We now describe a different way to specialize from the general random jump process described in Subsection 2.1, the so-called linear integration model.

A linear integration hypothesis usually involves the assumption

$$L(x,x+h) = \tfrac{1}{2}h \, \{l(x) + l(x+h)\} \,. \tag{21}$$

In view of expression (8), a uniform distribution of the events $d_{ij}(x,x+t)$ for $x < t \leqslant x+h$ is a sufficient condition for (21) for a certain h. Note that with this, $L(x,x+h)$ may be obtained with knowledge of the stock vectors at durations x and x+h only. We call the assumption "$d_{ij}(x,x + t)$ is linear in t for all i,j" the (weak) linear integration hypothesis.

The occurrence/exposure rates $m_{ij}(x,x+h)$ may be collected in a matrix $M(x,x+h)$ with the same configuration as the matrix of transition intensities $\mu(x)$. Then (8) can be rewritten after some easy algebra as

$$l(x+h) = l(x) \, \{ I + \tfrac{1}{2}h \, M(x,x+h) \} \, \{I - \tfrac{1}{2}h \, M(x,x+h)\}^{-1} \,, \tag{22}$$

see Rogers and Ledent (1976).

In view of (5) this suggests the relation
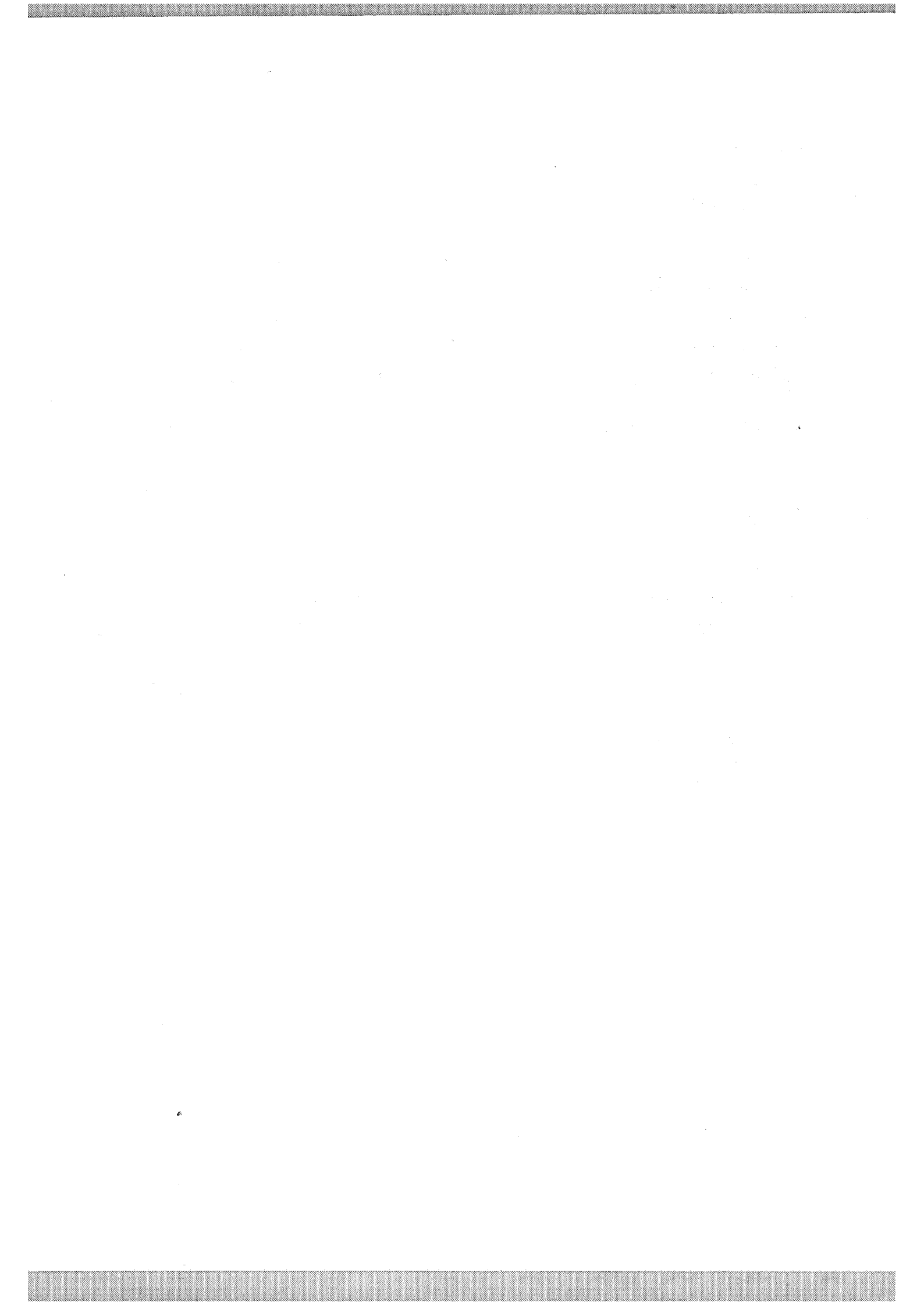
probability distribution over the states at duration x). The empirical counterparts of these two quantities are precisely the kind of data that are produced by population registration systems. By the accounting equation $l(x + h) = l(x) + 1 \, D(x,x + h)$ and because of the weak linear integration hypothesis we have $L(x,x + h) = \frac{1}{2}(l(x) + l(x + h))$; thus we can derive $M(x,x + h) = (\text{diag}(L(x,x + h)))^{-1} D(x,x + h)$ from $l(x)$ and $D(x,x + h)$ only. Our results are that, given any $l(x)$ and $D(x,x + h)$, one can find a random jump process with origin independent occurrence/exposure rates (over the duration interval $(x,x + h\,]$) and satisfying the strong linear integration hypothesis <u>if and only if</u> the matrix defined on the right hand side of (23) <u>is</u> a transition matrix (i.e. has non-negative elements, summing rowwise to one). Of course this is then <u>the</u> transition matrix for the process we have found. However, any such process is <u>not</u> Markov. So the two assumptions above are mutually consistent provided (23) is sensible; but they are not consistent with a Markov assumption. A sufficient condition for (23) to define a legitimate transition matrix by the way is

$$\sum_{j \neq i} m_{ij}(x,x + h) \leqslant 2/h \tag{24}$$

or equivalently (by construction of $m_{ij}(x,x + h)$)

$$\sum_{j \neq i} d_{ij}(x,x + h) \leqslant l_i(x) + \frac{1}{2} \sum_{j \neq i} d_{ji}(x,x + h): \tag{25}$$

the expected number of jumps out of i must be less than the probability of starting in i plus half the expected number of jumps into i; see Appendix 1.

Finally note that if the assumptions here are compatible with a given $l(x)$ and $D(x,x + h)$, there are actually very many <u>different</u> jump processes having these properties (see Appendix 2). This is unlike the Markov constant intensities model when $l(x)$ and μ completely determine the process.

3. AGGREGATE DATA FROM POPULATION REGISTRATION SYSTEMS

The data needed to estimate the parameters of stochastic process models 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, 1982, p. 168).

The observational plan influences the proper 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 the previously described models 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:

- although the data are obtained for individuals, most statistical bureaus make them available in the form of tables (aggregated over individuals) only;
- 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 or 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. One may have random left truncation (delayed entry) as well as right censoring without having to fundamentally revise the statistical procedure; see Andersen, Borgan, Gill and Keiding (1988).

In the situation of population registration data, consider the duration interval $(x, x + h)$. Then our data consist of the vectors $Y(x)$ and $Y(x + h)$ giving the numbers in which the population is distributed over the p states of the model at durations x and $x + h$. We are also given the matrix $N = N(x, x + h)$ with total numbers of direct moves from state i to state j during $(x, x + h]$, with $N_{ii} = -\sum_{j \neq i} N_{ij}$. Similarly to (8), we have

$$Y(x + h) = Y(x) + 1 \, N(x, x + h). \tag{26}$$

Let $n = \sum_i Y_i(x)$ be the total size of our population or sample.

We suppose that each of these n individuals have, independently of one another, moved from state to state according to the (probabilistically) same random jump process. We assume that other quantities, in particular the total exposure times $\int_x^{x+h} Y(s) ds$, are not available. Also the numbers of individuals starting in i and finishing in j, for each pair of states i and j, are not available either. This means

that direct estimation of the transition matrix $P(x,x + h)$ on the basis of observed 'net transitions' is not possible. Also the actual occurrence/exposure rates for the period $(x,x + h)$ are not available (these would be $(\text{diag}(\int_x^{x+h} Y(s)ds))^{-1} N(x,x + h))$.

4. ESTIMATION UNDER THE LINEAR INTEGRATION HYPOTHESIS

Because of (23), authors using the linear integration hypothesis estimate the matrix of transition probabilities by

$$\hat{P}(h) \quad = \{ I + \tfrac{1}{2}h \, \hat{M}(x,x+h)\} \, | \, \{I - \tfrac{1}{2}h \, \hat{M}(x,x+h)\}^{-1}, \tag{27}$$

where $\hat{M}(x,x+h)$ is the matrix of estimated occurrence/exposure rates using the linear integration hypothesis and I is the identity matrix:

$$\hat{M}(x,x+h) = (\text{diag}(\tfrac{1}{2}h(Y(x) + Y(x+h))))^{-1} \, N(x,x+h). \tag{28}$$

It has been pointed out that this approach contains a number of drawbacks (Ledent, 1980, p. 554; Hoem and Funck Jensen, 1982, p. 157, p. 160, pp. 194-201; Land and Schoen, 1982, pp. 316-320; Nour and Suchindran, 1984, pp. 325). One is that $P(h)$, as given by expression (27), may contain negative elements. This can only be the case, when $\sum_{j \neq i} \hat{m}_{ij}(x,x+h) > 2/h$, or equivalently when $\sum_{j \neq i} N_{ij}(x,x+h) > Y_i(x) + \tfrac{1}{2} \sum_{j \neq i} N_{ji}(x,x+h)$ for some i, see Subsection 2.3 and Appendix 1. 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 jumps 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 in $Y_i(x)$ 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 $\sum N_{ij}(x,x+h) > Y_i(x) + \tfrac{1}{2}\sum N_{ji}(x,x+h)$ is greater, the longer the interval $(x,x+h)$ is, provided that $\mu_i(y)$, $x < y < x + h$ is sufficiently large. Hence the (strong) linear integration hypothesis may be ruled out completely by the data, independently of the estimation method to be used, and even when immigration has been taken account of; for an example see Appendix 1.

Even if the strong linear integration hypothesis is assumed, it is not possible to derive useful statistical properties of the estimator (27), because for instance the large sample variances of the elements of $P(h)$ depend on more than on just the quantities $I(x)$ and $D(x,x+h)$ (which are basically all we can estimate). This is because many different random jumps processes exist (satisfying the hypothesis)

all with the same values of $l(x)$ and $D(x, x + h)$: the model is not identifiable with the given data. In principle one could calculate large sample variances for the special case of the model mentioned in Appendix 2: Markov with hyperbolic intensities <u>given</u> each initial state.

One may ask why (27) still gives reasonable results even if its derivation depended on strong assumptions which can hardly ever be true. In our opinion this is because for short duration intervals (which is precisely when the method is known to be reasonable) the strong linear integration hypothesis is true to a good approximation, as also is the assumption of origin independent occurrence/exposure rates. The latter is of course moving towards a Markov assumption. For the same reasons the 'Markov based estimate' $\exp(h \hat{M})$ (cf.(20)) gives similar, reasonable results when h is small.

5. ESTIMATION UNDER THE CONSTANT INTENSITIES ASSUMPTION

Had one observed the total exposures $\int_x^{x+h} Y(s)ds$ one would have calculated the actual occurrence/exposure rates

$$\widetilde{M}(x,x+h) = (\text{diag}(\int_x^{x+h} Y(s)ds))^{-1} \, N(x,x+h) \tag{29}$$

and hence

$$\widetilde{P}(x,x+h) = \exp(h\,\widetilde{M}(x,x+h)).$$

Large sample properties of these estimators are well known; see e.g. Aalen and Hoem (1978, p. 97). When only $Y(x)$ and $N(x,x+h)$ is available, an obvious compromise is to approximate $\widetilde{M}$ by using (28): thus a hybrid approach between those appropriate under the linear integration and the homogeneous Markov models. This is rather unsatisfactory and hence one of us derived a new estimator working consistently in the assumed model; see Gill (1986) and for further results Wieseman (1987).

The new estimator is actually based simply on the rather old method of moments. That is, for convenience replacing $(x,y]$ by the interval $(0,h]$, we write down the expected values of $Y(0)$ and $N = N(0,h)$; these are functions of the model parameters $l(0)$ and μ ; we then solve the estimating equations "observed = expected". This gives immediately $\hat{l}(0) = Y(0) \, / \, n$; $\hat{\mu}$ then has to be derived from the equation (cf.(18a,b))

$$N \, / \, n = \text{diag}(\int_0^h \hat{l}(0) \exp(\hat{\mu} s)ds) \quad . \tag{29}$$

This looks hard to solve but from (19a,b) we can discover that an equivalent set of equations is:

$$\hat{\mu} = (\text{diag } L)^{-1} \, N \, / \, n \tag{30}$$

$$\hat{L} = \hat{l}(0)(\exp(\hat{\mu} h) - I) \tag{31a}$$

$$\hat{L} \, 1^T = h. \tag{31b}$$

This suggests a very easy iterative solution method: given an initial guess of $\hat{L}$, compute $\hat{\mu}$ by (30). Now recompute $\hat{L}$ by substituting $\hat{\mu}$ in (31a,b) and solving

for $\hat{L}$. (One would at first think that (31a) is enough: $\hat{\mu}$ is a pxp matrix and $\hat{L}$ contains p unknowns. However $\hat{\mu}$ generally has rank p-1 or less, and then we need the extra equation (31b); see Gill (1986) for more details). This iterative method resembles the EM-algorithm (cf. Dempster, Laird and Rubin, 1977), in that we compute in each cycle the expected total exposure time n L as a function of μ, given Y(0) = n l(0); the EM-algorithm for computing the maximum likelihood estimator would require one to compute n L as a function of μ , given Y(0) = n l(0) and N = n **D**.

There are some questions which should be answered: do these equations always have a solution, is it unique, do the suggested iterations converge, and finally what are the statistical properties of the resulting estimators (e.g. as n → ∞)?

At least the first question has a positive answer. There always is a solution -that is, there always is a homogeneous Markov process which matches exactly a given l(0) and d(0,h). Compare this with the unsatisfactory result on the strong linear integration hypothesis.

On uniqueness and convergence there are very few theoretical results. Practical experience is however excellent. We conjecture uniqueness, and believe that the iteration technique given here is very useful.

Finally, because this is an identifiable model the distribution of the estimator around the parameter to be estimated can also be estimated: it only depends on n, l(0) and μ. In Gill (1986) it is shown that $\sqrt{n}(\hat{\mu} - \mu)$ is approximately normally distributed for large n with mean zero and with a variance which in principle can be computed. Wieseman (1987) has supplied the (horrific) formulas for this approximate variance which all the same are quite easy to calculate in practice. We give some practical results in Section 6 and the formulas themselves in Appendix 3.

Note that these results on μ carry over to similar results for $\hat{P} = \exp(\hat{\mu})$. One has the nice identity l(h) = l(0)P, just as in the linear integration hypothesis case.

An approximate variance of $\hat{\mu}$ and hence of $\hat{P}$ is also rather easy to find: since our estimator cannot be better than the maximum likelihood estimator of μ when more data is available, the actual occurrence/exposure rates, estimate the variance of the latter by $\hat{\sigma}^2 = (\text{diag } n\ \hat{L})^{-2}$ N (see Aalen and Hoem (1978)) and use it as a lower bound to the variance of our estimates: thus use the off-diagonal elements of $\hat{\sigma}^2$ as a lower bound to the variances of the corresponding elements of $\hat{\mu}$. In Appendix 3 we mention how to go from here to an approximate

variance of **P**. Our practical experience is that this works well for states with large exposures and low occurrence/exposure rates. In the opposite situation the lower bound grossly underestimates the actual variance. This is an important finding: it shows that population registration data can severely impair estimation of some important intensities.

We remark that it is computationally impossible to derive maximum likelihood estimators for $l(0)$, μ based on $Y(0)$, **N**. However in large samples the new method of moment estimators is just as good (the asymptotic variances are the same).

6. APPLICATION

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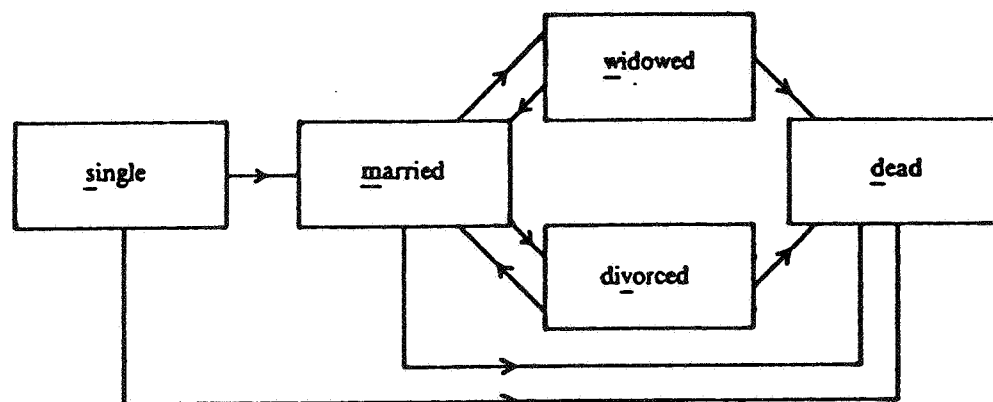
Table 1. Input data, nuptiality patterns of female birth cohort 1945, the Netherlands, 1965-1969

			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.

1. The "linear solution" $\hat{P}_1 = (I + \frac{1}{2}h\,\hat{M}_1)(I - \frac{1}{2}h\,\hat{M}_1)^{-1}$, where $\hat{M}_1 = (\text{diag }\hat{L}_1)^{-1}N$ and $\hat{L}_1 = \frac{1}{2}h(Y(0) + Y(h))$, see expressions (27) and (28). As was argued in section 4, it is not possible to give standard deviations of these estimates.

2. The "exact exponential solution" $\hat{P}_2 = \exp(\hat{M}_2\,h)$, where $\hat{M}_2 = (\text{diag }\hat{L}_2)^{-1}N$ and $\hat{L}_2$ is the (a?) solution to the equations $l(h) = l(0)\exp(M_2\,h)$, $L_2 \cdot 1^T = h$, compare expressions (20), (18a), (5) and (19b) respectively. Standard deviations of the estimators $\hat{P}_2$ and $\hat{M}_2$ were computed using the expressions in Appendix 3. The standard deviations of the estimated matrix of occurrence/exposure rates $\hat{M}_2$ (or, equivalently, of the estimated intensities $\hat{\mu}$ in this exponential model) are compared to those that would be obtained when occurrences as well as exposures would be observed. The latter standard deviations serve as a lower bound to the standard deviations of our estimates $\hat{M}_2$, as was argued in Section 5.

3. The "approximate exponential solution" $\hat{P}_3 = \exp(\hat{M}_1\,h)$.

Both $\hat{P}_1$ and $\hat{P}_2$ are exact solutions, and hence they fit the data exactly, i.e. $l(0)\hat{P}_1 = l(0)\hat{P}_2 = l(h)$. When checking the results it should be kept in mind that the length of the interval is $h = 5$ years, and that the total number of observed women is $n = 92131$.

Table 2 shows estimation results for the matrices of occurrence/exposure rates (expressed in occurrences per year) and for exposure times. To begin with the latter, differences between the linear estimator $\hat{L}_1$ and the exponential estimator $\hat{L}_2$ are 12-14 per cent, except for the state "dead" where the exact exposure is estimated to be 7 per cent higher than the approximate exposure. Percentage differences between exposure times are directly reflected in differences between estimations of occurrence/exposure rates contained in $\hat{M}_1$ and $\hat{M}_2$. For instance, the estimated approximate exposure time spent in the single state is 13 per cent higher than the corresponding estimated exact exposure time. Hence we see an estimate of the first marriage rate which is 13 per cent lower in the linear estimator $\hat{M}_1$ than in the exact exponential estimator $\hat{M}_2$.

When we consider exact estimations of standard deviations corresponding with elements of the exact exponential estimator $\hat{M}_2$ it becomes clear that even a relatively large cohort of more than 90,000 persons, observed over a 5-year period, results in reliable estimates of nuptiality intensities only for those women who were initially single. We can have some confidence in the divorce

Table 2. Solutions: occurrence/exposure rates and exposures[1]

			Marital status				
			s	m	w	v	d
Linear solution	$\hat{M}_1$	s	-.2238	.2233	0	0	.0005
		m	0	-.0057	.0007	.0047	.0003
		w	0	.1050	-.1085	0	.0035
		v	0	.1975	0	-.1980	.0006
		d	0	0	0	0	0
Exact exponential solution	$\hat{M}_2$	s	-.2529	.2523 (.0010) (.0011)	0	0	.0006 (.0001) (.0001)
		m	0	-.0050	.0006 (.0001) (.0003)	.0041 (.0001) (.0017)	.0003 (.0000) (.0001)
		w	0	.1206 (.0220) (.2341)	-.1246	0	.0040 (.0007) (.0085)
		v	0	.2223 (.0120) (.2421)	0	-.2230	.0007 (.0006) (.0009)
		d	0	0	0	0	0
Approximate exposures	$\hat{L}_1$	(yrs)	2.8458	2.1271	.0031	.0188	.0053
Exact exposures	$\hat{L}_2$	(yrs)	2.5189	2.4561	.0027	.0167	.0057

1) First figure in parentheses under estimated occurrence/exposure rates of exponential model is lower bound to corresponding standard deviation, second figure in parentheses is exact estimation of standard deviation.

rate (estimated to be .0041, with an estimated standard deviation of .0017), but no information can be obtained from other rates for married women, or from all rates for divorced women or widows. Observe that lower bounds of standard deviations give a misleading impression of the reliability of the estimated rates for all women, except for those who are single. Stated differently, population

registration data with information aggregated over individuals and over time would not permit us to estimate nuptiality rates of this five-state model. But when occurrences as well as exposures would be available, all rates but one (the death rate of divorcees) would have reliable estimates for this sample.

Table 3 contains estimated transition probabilities according to the "linear approach" $(\tilde{P}_1)$, the "exact exponential approach" $(\hat{P}_2)$ and the "approximate

Table 3. Solutions: transition probabilities[1]

			Marital status				
			s	m	w	v	d
Linear solution	$\tilde{P}_1$	s	.2825	.7088	.0010	.0056	.0022
		m	0	.9802	.0027	.0156	.0015
		w	0	.4067	.5760	.0032	.0141
		v	0	.6543	.0009	.3423	.0025
		d	0	0	0	0	1
Exact exponential solution	$\hat{P}_2$	s	.2825 (.0016)	.7083 (.0017)	.0010 (.0001)	.0060 (.0004)	.0022 (.0001)
		m	0	.9843 (.0008)	.0022 (.0003)	.0121 (.0010)	.0013 (.0005)
		w	0	.4481 (.6132)	.5330 (.6389)	.0035 (.0056)	.0154 (.0250)
		v	0	.6634 (.3876)	.0010 (.0008)	.3332 (.3900)	.0025 (.0025)
		d	0	0	0	0	1
Approximate exposures tial solution	$\hat{P}_1$	s	.3266	.6635	.0011	.0067	.0021
		m	0	.9811	.0027	.0147	.0015
		w	0	.3986	.5839	.0037	.0138
		v	0	.6194	.0010	.3772	.0024
		d	0	0	0	0	1

1) Figures in parentheses under elements of $\hat{P}_2$ are estimated standard deviations.

exponential approach" ($\hat{P}_3$). Estimated standard deviations for the elements of $\hat{P}_2$ are given as well. Transition probabilities in $\hat{P}_1$ are very close to those in $\hat{P}_2$, the largest differences being 4 percentage points (cf. the remarriage probability and the retention probability of widows). Differences between elements of $\hat{P}_2$ and $\hat{P}_3$ are of the same magnitude, although they appear somewhat more frequently. Observe that estimated standard deviations for transition probabilities of women who were initially widowed or divorced are so large, that no conclusions can be drawn about these probabilities.

7. CONCLUSIONS

In this paper we investigated the estimation of multidimensional demographic models when population registration data are available. Data of this kind are most often made available by statistical bureaus in the form 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 this 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 has 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 exact matching by the estimated model to the observed data, the linear integration hypothesis or our new method may be used. However, large transition intensities can prevent the former model from being applicable 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 very 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. However as we saw in our case a sample size of 100,000 is not 'large' at all.

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. We see that even with fairly large data sets, sampling variances can be very large and it is important to use models which enable them to be evaluated.

Notes

1) For instance if for each k the conditional distribution of the time of the k'th jump given the times and types (state i to state j) of the previous ones has a uniformly bounded hazard rate.

2) Available from Artemis Systems Inc., 125 Berry Corner Lane, Carlisle, Ma. 01741 USA or Datavision AG, P.O. Box 7250, Klosters, Switzerland.

REFERENCES

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

Andersen, P.K., Ø. Borgan, R.D. Gill and N. Keiding (1987) Censoring, truncation and filtering in statistical models based on counting processes, Report Centre for Mathematics and Computer Science, Amsterdam.

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.

Gill, R. (1986) On estimating transition intensities of a Markov process with aggregate data of a certain type: "occurrences but no exposures", Scandinavian Journal of Statistics, 13, 113-134.

Gill, R.D. and S. Johansen (1987) Product-integrals and counting processes, Report MS-R8707 Centre for Mathematics and Computer Science, Amsterdam.

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.

Johansen (1987), Product integrals and Markov processes, CWI Newsletter 12, 3-13; originally appeared (1977) as: Preprint no. 3, Institute of Mathematical Statistics, University of Copenhagen.

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. (1975) A population model based on a life table that includes marriage and parity, Theoretical Population Biology, 7 (2), 229-245.

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.

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.

Wieseman, M. (1987) Onderzoek naar de matrix van overgangsintensiteiten van een Markovproces (An investigation into the matrix of transition intensities of a Markov process) M.Sc Dissertation Mathematical Statistics, University of Amsterdam.

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. Properties of $P = (I + \frac{1}{2}hM)(I - \frac{1}{2}hM)^{-1}$

This appendix contains the proof of the following lemma given in section 3. For a matrix $M = (m_{ij})$ 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

$$P = (I + \frac{1}{2}h M)(I - \frac{1}{2}h M)^{-1} \qquad (A1)$$

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, 1 and 0 are vectors with all elements equal to 1 and 0 respectively, and for a vector x or a matrix A we write

$x \geqslant 0$ $(A \geqslant 0)$ if the inequality holds elementwise;

$x > 0$ if $x \geqslant 0$ and $x \neq 0$;

$x \gg 0$ if we have elementwise strict inequality.

Furthermore, $Z^{pxp} = \{ A = (a_{ij}) \epsilon R^{pxp} : a_{ij} \leqslant 0, i \neq j \}$, 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 $A \epsilon Z^{pxp}$ to be a "non-singular M-matrix". Two are:

(I 27) $\left.\right]$ x $\gg$ 0 with A x $\gg$ 0, and

(N 38) A^{-1} exists and $A^{-1} \geqslant 0$.

Now the proof is as follows:

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

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

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 $-m_{ii}(x,x+h) \leqslant 2/h$ is equivalent to

$$\sum_{j \neq i} d_{ij}(x,x+h) \leqslant l_i(x) + \tfrac{1}{2} \sum_{i \neq j} d_{ji}(x,x+h) . \tag{A2}$$

Hence, when in the linear integration model condition (A2) is fulfilled, then the matrix **P** according to (A1) is a transition matrix. Otherwise it may not be (but (A2) is not a necessary condition).

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:

$$M(20, 25) = \begin{bmatrix} -.2313 & .2291 & 0 & 0 & .0022 \\ 0 & -.0172 & .0007 & .0153 & .0012 \\ 0 & .1551 & -.1603 & 0 & .0052 \\ 0 & .4860 & 0 & -.4897 & .0037 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\quad\quad (1) \quad\quad (2) \quad\quad (3) \quad\quad (4) \quad\quad (5)$$

(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

$$P = \begin{bmatrix} .2672 & .7104 & .0009 & .0122 & .0093 \\ 0 & .9577 & .0025 & .0337 & .0061 \\ 0 & .5419 & .4285 & .0093 & .0203 \\ 0 & 1.0694 & .0013 & -.0824 & .0117 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

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

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

and that of the initial distribution

$$l(20) = (54177 \quad 41955 \quad 59 \quad 544 \quad 3265) .$$

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. On the compatability of the strong linear integration hypothesis, origin-independent occurrence/exposure rates and the Markov assumption

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 $l(0)$, D and $l(1) = l(0) + 1\,D$ are given, such that the transition matrix $P = (I + \frac{1}{2} M)(I - \frac{1}{2} M)^{-1}$ is a stochastic matrix, with M a matrix of occurrence/exposure rates for the interval $\begin{bmatrix} 0, 1 \end{bmatrix}$ defined as $M = (\mathrm{diag}\ L)^{-1} D$. Now a (stochastic) jump process can be constructed with initial distribution $l(0)$ and expected number of direct transitions 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 P, we also know $l^{(i)}(1)$, that is the final distribution conditional on state i at duration 0. For the rows of P are the conditional distributions of "state at duration 1" conditional on "state at duration 0". Now condition (i) implies that $L^{(i)} = \frac{1}{2}(l^{(i)}(0) + l^{(i)}(1))$ for each i. Here $l^{(i)}(0)$ is the vector with value $l_i(0)$ in the i-th element and zeros otherwise, whereas $L^{(i)}$ 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 $M^{(i)}$, is independent of i and therefore $M^{(i)} = M$ for each i. Then we compute a matrix with direct transitions conditional on state i at duration 0 as $D^{(i)} = (\mathrm{diag}\ L^{(i)})M$. For each i a jump process exists with this $l^{(i)}(0)$ and $D^{(i)}$ which satisfies the (weak) linear integration hypothesis, namely the Markov process with initial distribution $l^{(i)}(0)$ and time-dependent (hyperbolic) intensities $\mu^{(i)}(t) = \mathrm{diag}(l^{(i)}(0) + t\,1\,D^{(i)})^{-1} D^{(i)}$. (There also exist many non-Markov processes which could be used here: they just have to have the right values of $d_{(i)jk}(t)$, $0 \leqslant t \leqslant 1$.). 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.

APPENDIX 3. The variance of $n^{\frac{1}{2}}(\hat{\mu}-\mu)$ and of $n^{\frac{1}{2}}(\hat{P}-P)$

Consider a duration interval $\begin{bmatrix}0,1\end{bmatrix}$ and let $l(0)$, $l(1)$ and D be the expected initial distribution, the expected final distribution (at duration 1) and the matrix of expected numbers of moves between two states, all 'per individual'.

We estimate μ replacing $l(0)$, $l(1)$ and D by estimates $\hat{l}(0) = Y(0)/n$, $\hat{l}(1) = Y(1)/n$, $\hat{D} = N/n$ and then solving for $\hat{L}$ and $\hat{\mu}$ the equations:

$$\hat{\mu} = (\text{diag } \hat{L})^{-1} \hat{D}$$

$$\hat{l}(0)\exp \hat{\mu} = \hat{l}(1)$$

subject to $\hat{L}1^T = \hat{l}(0)1^T = \hat{l}(1)1^T = 1$.

Wieseman (1987), working out ideas of Gill (1986), show that, conditionally on $Y(0)/n(= \hat{l}(0))$, $n^{\frac{1}{2}}(\hat{\mu}-\mu)$ is approximately normally distributed with mean zero and variance-covariance matrix $\text{Asvar } \hat{\mu}$ which can be calculated (and hence estimated by substituting μ by $\hat{\mu}$ etc.) as follows.

Firstly we solve linear differential equations for three quantities $l(t)$, $K(t)$ and $\text{Var}_N(t)$ $(0 \leqslant t \leqslant 1)$; the first is a $1 \times p$ vector, the second a $p \times p \times p$ array, the third a $p(p-1) \times p(p-1)$ matrix induced by ordered pairs of different states i,j.

The initial conditions are

$$l(0) = \hat{l}(0) \text{ (given)}$$
$$K(0) = 0$$
$$\text{Asvar}_{\hat{D}}(0) = 0$$

Then the differential equations themselves are

$$l(t + dt) = l(t).(I + \mu dt)$$
$$K(t + dt) = K(t).(I + \mu \ dt) + l(t)^T \otimes I \ dt$$
$$\text{Asvar}_{\hat{D}}(t + dt) = (I + A \ dt)^T \text{Asvar}_{\hat{D}}(t).(I + A \ dt) + B(t)dt$$

where

$$(A)_{ij,kl} = (-\delta_{ik} + \delta_{jk})\mu_{kl} \left.\vphantom{\begin{matrix}a\\b\end{matrix}}\right\} \quad i{\neq}j,\ k{\neq}l$$
$$(B(t))_{ij,kl} = \delta_{ij,kl}\mu_{kl}l_k(t)$$

Also . is the ordinary matrix product and $\otimes$ is the Kronecker product (so in the expression $K(t).(I + \mu\,dt)$ we sum over the third index of K, which equals the first index of $(I + \mu\,dt)$). These equations can be solved by the difference scheme which they suggest. We then work further with $I(1)$, $K = K(1)$ and $AsVar_N = AsVar_N(1)$ respectively. As a check, one can verify that $I(1)$ computed this way also eqals $I(0) + ID$; that $\sum_k (K)_{iik} = L_i$; that $\sum_i (K)_{iik} = I(1)_k$; and that

$$\sum_j \mu_{ij}K_{ijk} = \sum_j K_{iij}\mu_{jk}.$$

Next we define matrices U,V,W and Z ($p{\times}p$; $p(p-1){\times}p$; $p{\times}p(p-1)$ and $p(p-1){\times}p(p-1)$ respectively) by

$$U_{ik} = -\frac{1}{L_i}(K.\mu)_{iik}$$

with the p'th colum replaced by 1's

$$V_{ij,k} = (\frac{1}{L_i}(K_{ijk} - K_{iik}) + \delta_{kj} - \delta_{ik}) \quad i{\neq}j$$

with the p'th column replaced by 0's;

$$W_{i,jk} = \frac{-\delta_{ij}\mu_{ik}}{L_i} \quad j{\neq}k$$

$$Z_{ij,kl} = \frac{\delta_{ij,kl}}{L_i} \quad i{\neq}j,\ k{\neq}l$$

Then we calculate

$$X = -V\,U^{-1}W + Z$$

and finally

$$\text{AsVar}_{\hat{\mu}} = X^T \text{AsVar}_{\hat{D}} X;$$

this $p(p-1) \times p(p-1)$ matrix contains the asymptotic variances and covariances of the offdiagonal elements of $n^{\frac{1}{2}}$ ($\hat{\mu} - \mu$).

As a final check, the matrix

$$\text{AsVar}_{\hat{\mu}} - \text{diag}(\text{vec}((\text{diag } L)^{-1} \mu))$$

should be positive definite (here vec means: form the $p(p-1) \times 1$ vector with the nondiagonal elements of the given matrix).

One can go on to compute the approximate value of $\text{AsVar}_{\hat{P}}$ where $\hat{P} = \exp(\hat{\mu})$.

Since $\dfrac{\partial P_{ij}}{\partial \mu_{kl}} = K_{iklj} - K_{ikkj}$

where K_{ijkl} is, for given i, the $p \times p \times p$ array $K(1)$ obtained ealier when we replace $I(0)$ by the vector having a 1 at the i'th place and zero's elsewhere, we can collect these partial derivatives into a $p^2 \times p(p-1)$ matrix S and apply the delta method giving

$$\text{AsVar}_{\hat{P}} = S \, \text{AsVar}_{\hat{\mu}} \, S^T.$$

These calculations appear complicated but can be radily carried out using a matrix oriented programming language such as APL (we actually recommend P.J. Huber's statistical programming environment ISP[2]) which combines APL-like features with a friendly, statistically orientated environment). The actual program is then not much longer than the number of formulas above.

A further check: compare var P with the multinomial covariances one would have had if transitions had been observed. We have less data but more structure (model assumptions), hence these two types of variances could be comparable.