Bridging the micro-macro gap in population forecasting

Contract no. SP23-CT-2005-006637

Revised version of Deliverable D1

Description of the multistate projection model (Multistate model for biographic analysis and projection)

Work Package 1
Multistate Methods

Author:

Frans Willekens

Netherlands Interdisciplinary Demographic Institute
P.O. Box 11650
2502 AR the Hague
The Netherlands

Multistate model for biographic analysis and projection

Frans Willekens*
Netherlands Interdisciplinary Demographic Institute (NIDI)
The Hague

December 2006

* I like to thank dr. Sergej Prokopenko, Max Planck Institute for Demographic Research (MPIDR), Rostock, for his extensive and very helpful comments on an earlier draft.

Content

1. Introduction

> "Our reading of current policy literature suggests there is emerging agreement on the need to focus on individuals as the unit of analysis in social policy. There is much recent emphasis on longitudinal data that track people over time and on analysis at the micro level, that is, at the level of individuals as opposed to predetermined groups of people."
> In: "A life course approach to social policy analysis. A proposed framework." Policy Research Initiative, Government of Canada, 2004, p. 6.

A population is composed of individuals and demographic changes are consequences of changing lives of men and women. Individuals differ in a multitude of attributes that vary with age and in time. Attributes relate to living status (alive or dead), health status, educational status, marital status, employment status, place of residence, living arrangement, identity, religious denomination, political views and party membership, personality, etc. Several of the attributes and traits can be measured directly or indirectly, but individuals have also latent attributes that are difficult to uncover. Attributes may be grouped into domains of life such as family, work, health and living space. Some attributes are time-invariant but most change with age and over time and as a result the composition of the population changes. When an attribute changes an event occurs. The change is often associated with a transition to a new stage of life. Events and the associated transitions underlie the lifepath, which is often used to characterize lives.

Attributes and events are the two basic building blocks of the human life course. An individual's biography is described in terms of the objective sequence of attributes and events and the subjective interpretation of attributes and events. If the biography is described in terms of attributes, the biographic perspective is said to be *attribute-based*. If, on the on the hand, the biography is described in terms of attributes at birth and subsequent events, the biographic perspective is *event-based*. The distinction is particularly relevant in biographic modeling. In biographic studies, both perspectives are often used simultaneously. The timing of events and the reasons for change are important concerns in biographic analysis. In this paper, population dynamics is described in terms of biographies of individuals and groups of individuals who share characteristics. A particularly useful group consists of individuals born during a same period.

In biographic modeling, an attribute or a combination of attributes is referred to as a (functional) state. The set of possible attributes is the *state space*. At any age, an individual occupies one of multiple states. The state variable indicates which state is occupied at a given age or time. If the attribute is hidden, the state variable is a latent variable. In most biographic studies, the state variable is a discrete variable; the number of values is equal to the size of the state space. A change in value of the state variable implies a transition from one state to another. The biographic model connects individuals and populations. Two types of state variables are distinguished: individual state variables (i-state variables) and population state variables (p-state variables)[1]. Population state variables describe population characteristics, such as the number of individuals in a given state at a given age or time. In this paper, an individual's biography is defined as the sequence of state occupancies (attributes) and transitions (events). The sequences are presented in two time scale: age and calendar time. Individual biographies are aggregated into cohort biographies and different cohorts constitute a population. The biography of a cohort or another group of individuals is fully documented if the number of individuals is known for each state of the state space and for any age and point in time. The state occupancies and state transitions represent two fundamental measures in quantitative biographic analysis. They are the basis for other biographic measures, such as the expected sojourn time or duration of stay in a state and the expected number of transitions.

Observations on individual biographies are generally incomplete since it is not feasible to follow individuals from birth to death. Information is generally limited to segments of life. These *observation windows* may extend over a narrow period of less than one year or over several years. Information on several individuals may be combined to produce a *synthetic biography* that is characteristic for a group of people observed during a same period. The empirical information on biographies may come from different sources and integrated using statistical techniques for combining data from different sources. The empirical information may be augmented by other information, such as expert opinions and/or behavioural assumptions to produce synthetic biographies that extend beyond the observation window. Examples include biographic projections that extend biographies into the future. Synthetic biographies constitute the main subject of this paper. The aim of the paper is a mathematical model of human biographies. The model should serve as an instrument for biographic analysis and forecasting. The main features of the model are:

i. It is a microdemographic model: it describes biographies of individuals and includes rules for aggregating individual biographies into cohort biographies and cohort biographies into population trajectories. A microdemographic model connects individuals and populations by focusing on individuals. It starts with the state of the individual, represented by i-state variables and derives characteristics of the population, represented by p-state variables. The microdemographic or individual-based approach to population studies is not entirely new. It was already considered by Lotka (1925) more than 80 years ago. Lotka outlined two basic approaches to population studies. One was the study of population dynamics by modelling processes at the population level through differential equations involving average individuals. The second was to derive population dynamics from individual behaviour. The first approach has dominated the literature (see Wilson, 1996). The two approaches to population studies outlined by Lotka, were also distinguished by Coleman (1990) in sociology[2]. The microdemographic approach is fully consistent with

[1] The distinction between i-state variables and p-state variables is due to Metz and Diekmann (1986) (see Keyfitz and Caswell, 2005, p. 64).

[2] In his book *Foundations of social theory*, Coleman (1990, p. 2) advances two modes of explanation of behaviour of a system. One depends on observation of the behaviour of the system as a whole over a period of time. The other examines processes internal to the system involving components of the system and

methodological individualism developed by Coleman. It is also consistent with the multistate/multistage modelling approach proposed by Metz and Diekmann (1986) in population ecology. Metz and Diekmann present conditions that guarantee straightforward aggregation of i-state variables in p-state variables (see also Keyfitz and Caswell, 2005, pp. 64ff). The rules of aggregation are dependent on real or assumed patterns of interaction between individuals. Interactions may involve both competition and cooperation. It may result in connectedness of individuals and the emergence of social institutions that facilitate and constrain the behaviour of individuals. In this paper individuals are assumed to be independent and interactions to be absent.

ii. It is a multistate model: it describes biographies in terms of state occupancies and state transitions. In mathematical system theory, the multistate model is known as the state-space model (see e.g. Director and Rohrer, 1972). The state of a system is represented by a state vector and the dynamics by a system of differential equations (continuous time) or difference equations (discrete time).

iii. It is a continuous-time model: the model describes changes in individuals and populations in continuous time. Differences in characteristics at the start and the end of an interval of age or time are related to continuous change during the interval. Modelling in continuous time allows a more precise analysis of changes than modelling in discrete time. For a detailed treatment of continuous-time models in conventional demography, see Pressat (1995).

iv. It is a model in two time scales, age (time of life) and calendar time (historical time). The model is consistent with the generalized population dynamics proposed by Arthur and Vaupel (1984).

v. It is an integrative model that encompasses the life table and the projection model. Traditionally the life table and the projection model are viewed as two separate models of demography. Keyfitz describes the life table as 'a scheme for expressing the facts of mortality in terms of probabilities' (Keyfitz, 1968, p. 3). These facts are expressed by following a (synthetic) cohort or group of people through successive ages until they die. A projection may be approached as a scheme for expressing the facts of mortality and fertility in terms of expected number of persons at points in time subsequent to a census or other starting point. This formulation incorporates Keyfitz' description of a population projection (Keyfitz, 1968, p. 27). What both the life table and the projection have in common is that they represent ways to think effectively about observed birth and death rates by asking what they would lead to if continued. The life table follows a cohort while the projection model follows a population that consists of several cohorts. In this paper, the distinction between life table and projection model is no longer maintained. In essence, the life table incorporates a projection model and that model is focused on in this paper. The life table and the projection model have a different perspective and a different focus, however. The life table adopts a cohort perspective and

individual units. The rationale for studying the behaviour of a system by examining individual behaviour is that individual-level actions and system-level behaviour are linked. A similar view is held by Metz and Diekmann (1986) who recognized the need to study populations by starting with individuals.

focuses on the experience of a cohort by reporting characteristics at consecutive ages.. The predominant time scale is individual time (age). The projection model adopts a period perspective and focuses on the experience of different cohorts evolve in time. It reports population characteristics at consecutive points in time. The focus is on calendar time. Cohort and period perspectives are basic perspectives adopted in traditional demography and also in this paper.

vi. It is a matrix model: the system of differential equations is written in matrix terms and matrix methods are used to solve systems of equations. Matrix population models were introduced in the 1940s by Bernardelli (1941) and Leslie (1945). They were largely neglected until the mid-1960s when demographers and ecologists rediscovered them (see Keyfitz and Caswell, 2005, p. 49). By relying on the theory of matrices, Rogers (1975) could pioneer the new subfield of demography that became known as multistate demography. Multistate demographic models produce several interesting demographic indicators that cannot be calculated without matrix theory. Matrix models continue to occupy a central role in demography and biology. For an illustration, see Ellner and Rees (2006).

vii. It is a probability model: the state vector consists of probabilities of state occupancy (state probabilities). The multistate model is a continuous-time Markov chain (CTMC) that is represented by a system of differential equations. The changes of the state probabilities with age and over time are governed by a set of instantaneous rates of transition (transition intensities) that depend on current characteristics and may depend on the entire life history. In physics, the equations that describe the time evolution of state probabilities are known as the *master equations*. The differential equations are deterministic, not stochastic. The *expected* state occupancies change deterministically with age and time. Idiosyncrasies of the individual journey through the state space are introduced using microsimulation techniques. Akushevich et al. (2005) used stochastic differential equations to model individual lifepaths and the resulting population changes.

viii. The model belongs to the class of exponential models. Applications of continuous-time models and the estimation of the parameters (intensities) from empirical observations require that assumptions are made about the age and time dependence of the transition intensities. Two general approaches exist. One is to parameterise the age- and time-dependence of the transition intensities. The other is a non-parametric approach. The non-parametric approach involves assumptions about the time dependence of the transition rates (hazard function) or state probabilities (survival function) during intervals of age and/or time. In demography and actuarial sciences, the most common assumption is that the survival function is piecewise linear, implying that the events are uniformly distributed during age intervals. In demography, the model that results from this assumption is known as the *linear model*. The approach has been criticized by Hoem and Funck Jensen (1982) and others, who advocated the alternative assumption of piecewise constant intensities implying a piecewise exponential survival function. The alternative assumption is common in the statistical analysis of duration data including lifetime data (survival analysis). Some demographic models such as LIPRO (van Imhoff and Keilman, 1991) implemented the assumptions in what became known as the *exponential model*, whereas most demographic models adopt piecewise linear survival functions (see e.g. Preston et al., 2001; Willekens and Drewe, 1984). The predominance of the linear model is also illustrated in Wunsch et al. (2002) on the life table. This paper represents a break with the past. It follows the exponential model and treats the linear model as an approximation to the exponential model. The exponential model is covered in the main text and the linear model is described in an appendix.

The report is organised as follows. Section 2 describes the approach that is adopted in biographic forecasting. The basic idea is to view life as a journey through a state space, manifested in sequences of states and transitions (events). The states and events may pertain to one particular domain of life or to a combination of domains such as work and family. A sequence in one domain is referred to as a career.

In demographic and biographic analyses, time is a key concept. Time scales position individuals and events in time and different time scales may be used simultaneously. Examples are calendar time (historical time), age (individual time) and duration of stay (e.g. marriage, duration of residence, duration of unemployment). Time scales measure the time elapsed since a reference event. The unit of time is generally fixed (day, month, year) but may vary. Time may be measured continuously or at discrete intervals. Although most events occur in continuous time, they are studied in continuous time or discrete time. The distinction between continuous age and time and discrete age and time is critical. If time is a discrete variable, time intervals are considered. The intervals can be of any length but in social and health sciences, they are usually one month, one year, or five years. If time is a continuous variable, the intervals are infinitesimally small. Section 3 presents a generic framework that locates events and transitions in two time scales, age and calendar time. The Lexis diagram and the Lexis surface show to be very useful devices to visualize different windows of observation that are delineated by age and time. An individual is positioned in two time scales: individual time (age), which is denoted by x, and historical time or calendar time, denoted by t. The position may be indicated by the exact or instantaneous age and time (continuous time) or by the age and/or time interval (discrete time). In continuous time, an individual of exact age x at exact time t is born at t-x. An unambiguous relation exists between age, current date and date of birth. In discrete time, an individual aged x to x+h at exact time t is born during the period from t-x-h to t-x. On the other hand, individuals born between t-x-h and t-x reach their x-th birthday, i.e. exact age x, somewhere during the period from t-h to t (see later).

The proposed age-time framework has a number of advantages. First, it clarifies the distinction between cohort analysis and period analysis. Cohort analysis focuses on cohort membership and studies how cohorts differ, i.e. to what extend members of different cohort behave differently. The subject of study is the life course of cohort members as it unfolds with age. The unfolding is generally documented by reporting the characteristics of cohort members at consecutive birthdays. Life-table analysis is an illustration of cohort analysis (real or synthetic cohort). In cohort analysis, age is the dominant time scale. Period analysis focuses on how population structures and demographic processes and behaviour evolve in time. The change is documented by reporting the characteristics of the population at consecutive points in time along the calendar time axis. Age is one of the characteristics, which implies that birth cohorts are distinguished. Calendar time is the dominant time scale and the characteristics of cohort members are described at consecutive points in time. Population projection may be viewed as an illustration of period analysis. Second, it clarifies and resolves a number of issues at the interface between continuous time and discrete time.

The multistate probability model that describes individual and cohort biographies is the continuous-time, finite-state Markov chain. The basic parameters of the model are the instantaneous rates of transition or probabilities of transition during an infinitesimally small interval (transition intensities). The initial condition, i.e. the state probabilities at birth, and the transition intensities for every age between birth and death determine the entire biography. In practical work, the transition intensities are assumed to be piecewise constant. The transition rates that result and the transition probabilities that are obtained from the rates are unambiguously related to the transition intensities. The relation is documented in Section 4.

The multistate model is presented in Sections 5 and 6. Section 5 addresses a closed population, i.e. a population in the absence of entries from and exits to positions outside the boundaries of the population system. In Section 6, an open population is considered. Section 7 introduces covariates in multistate population analysis and projection. The approach relies on transition rate models and transition probability models and follows the perspective on causal analysis adopted by Blossfeld and Rohwer (2002).

The empirical base for demographic forecasting consists of data of various types. Two broad data types are distinguished: transitions measured in continuous time and transitions measured in discrete-time. When transitions are measured in discrete time, the positions of individuals are recorded at two consecutive points in time. In that case, transition rates cannot be estimated directly from the data and the number of events that occur during an interval cannot be determined. Section 8 describes a method for estimating transition rates from discrete-time transition probabilities. Section 9 considers another missing data problem. In this case, data on exposures are missing. A method is described to estimate transition rates from data on events and initial population only.

The cohort biography may be obtained by two methods. One method relies on expected values of model parameters. That method is used in most of this paper. The other method constructs individual biographies for all cohort members or for a sample of the cohort, and combines the individual biographies to generate the cohort biography, which coincides with the biography of an average cohort member. If the sample size is sufficiently large, then the biography of an average cohort member is the expected value of the biography of any individual cohort member. Transition rates provide a bridge between the population level and the individual level. At the individual level of analysis, e.g. to predict individual biographies, use is made of transition rates that depend on individual attributes. At the population level, e.g. to forecast cohort biographies, expected values of transition rates across individuals are used. The micro-macro link is further described in Section 10. Section 11 concludes the report.

2. An introduction to multistate models for biographic analysis and projection

Biographic models describe how state occupancies and events evolve with age for individuals, groups of individuals and entire populations. Biographic models are transition models that consider multiple states and describe trajectories through the state space. They are known as multistate models. This section situates the biographic model within the context of demographic projection and highlights the main features of the biographic model.

2.1. From demographic to biographic projection

Demographic projections are usually confined to populations disaggregated by age, sex and sometimes race/ethnicity, household status or region of residence. The general methodology, the cohort-component method, is well-established. The basic approach is to distinguish birth cohorts, to determine the number of survivors in the base year and to determine *for each cohort* and for each future year the number of persons by age and sex that (1) enter a population because of birth or immigration and (2) leave a population because of death or emigration. The number of entries and exits are based on *rates* of birth, death and emigration by age and sex, and number of immigrants by age and sex. The estimation of empirical rates from data (often incomplete or defective data) and the prediction of future rates involve important methodological issues.

The projected population by age and sex serves as an input in functional population projections, which are related to particular functions or activities in society. They include projections of the population by functional status such as labour force status, educational status, health status and status in the household. These projections are made to determine the projected need for some ‘function’ – a product, a service, an allowance, an activity or a facility (Kono, 1993). Examples include:

- The future size of the labour force to determine the supply of labour and the demand for jobs.
- The future size of the population retired from the labour force to determine the demand for pensions.
- The future size of school enrolments to determine the demand for teachers and buildings and to determine the population composition by level of education and hence the human capital.
- The future size of the population by health status and/or disability status to determine the demand for health care including the number of physicians and hospital beds.

- The future number of households by size and type to determine the demand for housing and durable consumer goods.
- The future number of people eligible for assistance of different type. Eligibility criteria frequently include age, sex and functional status (e.g. level of income, health status, labour force status).
- The future size of vulnerable groups in society.

Despite the wide variety of functions, from a methodological viewpoint traditional functional projections differ from each other only in minor detail. Traditionally, people of a given age and sex are allocated to functional states using a set of prevalence rates, a distribution function or another allocation mechanism. The method is referred to as distribution method. In the forecasting literature the method is also known as *static* as opposed to methods that are based on transition rates (or incidence rates) and that are *dynamic* (such as the multistate model) (see e.g. Zeng Yi et al., 1997). Examples of the static method based on distributions are the headship rate method for household projections (see e.g. Kono, 1987; Linke, 1988), labour force projections based on labour force participation rates (see e.g. Carone, 2005), educational projections based on enrolment rates, the ratio method of subnational population projection, and health status projections based on prevalence rates. The latter method is commonly referred to as the 'Sullivan method' (see e.g. Goldman et al. 2004). The distribution function may change over time to capture real or assumed shifts in behaviour or conditions. For illustrations of the static method in functional population projections, see Bogue et al. (1993, Chapter 18).

In the dynamic method, the distribution of people among functional states is not imposed by a distribution function but is the outcome of transitions people make in life. Individuals move between functional states and as a consequence, the structure of the population changes. Individuals also enter the population through birth or immigration and leave the population through death and emigration. Transitions between functional states are sometimes referred to as *internal events* (Van Imhoff and Keilman, 1991, p. 20). Entries and exits, i.e. transitions that involve a crossing of the system boundaries, are referred to as *external events*. The rates of transition determine the population dynamics and the rates may change over time. In demographic models they always vary by age and sex, and they may vary between subpopulations. In the dynamic method, several states of existence are distinguished and the transitions between the states are considered. The method is known as the multistate method. Because of the pivotal role of transitions, multistate models picture more closely the *mechanism of demographic change* taking place in the real world. As a result, they are better suited for integrated population projections in which functional states and interactions between functional states play a crucial role. In addition, the transitions provide a way to assess the impact on population dynamics of behavioural changes brought about by technological, economic or cultural change, or policies. The transitions are age-specific. As a result the multistate model gives at each age the distribution of cohort members among functional states. Formulated differently, it indicates how the characteristics of the cohort vary with age and in time.

The multistate method has become the standard methodology among demographers (Rogers, 1975, 1995; Hoem and Funck-Jensen, 1982; Willekens and Drewe, 1984; Keyfitz, 1985; Schoen, 1988; Ahlburg et al, 1999; Van Imhoff, 1990; Van Imhoff and Keilman, 1991; Zeng Yi, 1991; Wilson and Rees, 2005). It has been applied for projecting regional populations, and populations by educational status, household status, labour force participation and health/disability status. The multistate model is currently receiving much interest in epidemiology and public health (for a review, see Commenges, 1999; Hougaard, 1999, 2000). A major difference between multistate models in demography and epidemiology is that in demography age is a key variable, whereas in epidemiology it is not (yet). Influenced by medical statistics, the situation is changing in epidemiology. For a discussion of systematic inclusion of time (and age) in epidemiological studies, see Keiding (1999) and Pencina et al. (2006)

The choice of static versus dynamic method has been the subject of long debates in demographic analysis and forecasting. In labour force projections, the debate was most intensive in the early 1980s after the publication in 1982 by the Bureau of Labor Statistics in Washington D.C. of multistate tables of working life (Smith, 1982). In health status projections, the debate is of a more recent date (see e.g. Crimmins et al., 1994; Mathers and Robine, 1997). Some authors attempted to reconcile the two approaches (see e.g. Newman, 1988). In his review of functional population projections, Kono asserts that 'Because of complex and precise data demands, however, almost no multistate models which could be used reliably in official national projections, beyond regional projections, have been developed.' (Kono, 1993, p. 18.2).

Multistate models extend the cohort-component model that project a population by age and sex to a biographic projection model that projects the population by age, sex and other characteristics. The transitions that cohort members experience in life and the ages at which they experience the transitions determine how the distribution of members among functional states changes as the cohort ages[3]. The result is a *cohort biography*[4] that gives for each age the distribution of cohort members among functional states. The cohort biography is a collective biography, i.e. the combined biography of a group of individuals. The same multistate model can be applied to a single individual and generate an *individual biography*. It gives, for each age, the probability that an individual cohort member occupies a state. The probability is the expected value of an individual's state occupancy. For a given individual, the state occupied at a given age is determined by chance, more specifically by drawing a random number from a cumulative probability distribution. The chance mechanism is implemented using Monte Carlo simulation.

The research that leads to this paper is part of the MicMac project. The aim of the project is to bridge the gap that exists today between population-based models of population dynamics and individual-based models. The strategy adopted in the project is to view population-level measures as *expected values* of individual-level measures. In other words, the values of individual measures are distributed around the values at the population level. Probability theory is used to characterise the distributions. At the outset,

[3] Multistate models of the life course are also known as multistate life tables.
[4] The concept of cohort biography was introduced by Ryder (1965).

two models are distinguished. Mac concentrates on the cohort level, whereas Mic concentrates on individuals. Together Mac and Mic are able to generate the very detailed demographic forecasts that are required for the development of sustainable (elderly) health care systems and pension systems. For an introduction to the MicMac project, see Van der Gaag et al. (2005). Willekens et al. (2005) present an application of a prototype of the MicMac projection model to fertility using survey data. Van der Gaag et al. (2006) present an application to living arrangements using population register data.

The biographic model presented in this paper and developed as part of the MicMac project has several innovative features that distinguish it from other multistate projection models in use today. Several features are listed in the introduction. Here a few features are considered in more detail.

a. The model is individual-based and projects individual lifepaths. Individuals are members of birth cohorts and a combination of birth cohorts constitutes a population. As in traditional demographic projection models, such as the cohort-component model, population figures are obtained by combining data from different cohorts. Cohort figures are obtained from individual data (micro-data) in two different ways. The first is by using *expected values* of the demographic parameters (transition rates) that govern the dynamics of the population. The expected values are estimated from individual observations using statistical techniques, in particular the maximum likelihood method. The second is by aggregating individual biographies. The first method is entirely consistent with the multistate cohort-component model generating cohort biographies. The transition rates used in multistate models are expected values (by birth cohort and/or age and sex). The second method is used in microsimulation. To reliably project cohort biographies from individual lifepaths, it is not necessary to project the lifepaths of all cohort members. It suffices to project the lifepaths of a sample of cohort members. The size of the sample that is required to be representative for the birth cohort depends on the cohort heterogeneity. If cohort members are identical, it suffices to project the lifepath of a single member of the birth cohort. In the case of heterogeneous cohorts, the sample means will be closer to the expected values the larger the sample sizes. In the MicMac project we follow the view of Wolf (2000) that microsimulation is essentially an exercise in sampling (see below).
b. The model is founded on probability theory. In 1982, Hoem and Funck-Jensen asserted that multistate demographic models should be based less on accounting principles and more on the theory of probability. Accounting principles dominated demographic modeling for decades. Traditionally demographic models, in particular the life table and the projection model, are derived from accounting equations as demonstrated by any standard text (see e.g. Preston et al., 2001, p. 42). Until recently, relatively few texts adopt a probabilistic perspective (for an exception, see Namboodiri, 1991, Namboodiri and Suchindran, 1987, Chiang, 1984 and Biswas, 1988, 1995). In the probabilistic perspective, the parameters and variables of the model and the demographic indicators that result may take on multiple values as a simple consequence of chance. The variables of the projection model are random variables characterized by probability

distributions. The expected value of the variable is only one of many plausible values. It is interesting to note that the probabilistic population projections, which are receiving considerable attention in demography, are generally based on projection models that are derived from accounting equations.

c. The third innovative feature of the model is related to the first. It is its focus on biographies. Since the 1970s, researchers in multistate demography emphasized the potential multistate models have for the analysis and reconstruction of life histories and the multistate life table has extensively been used to generate *synthetic* cohort biographies (see e.g. Willekens and Rogers, 1978; Schoen, 1988). A biographic projection extends the notion of synthetic biography to the domain of projection and forecasting. Biographic projections differ from demographic projections in one important way. Demographic projections focus on population counts (population size, number of people) and multistate demographic projections focus on counts of people by functional state. Biographic projections focus on population counts by functional state, sequences of states **and** durations of stay in each functional state. Hence, whereas the focus of demographic projections is counts ('how many?'), the focus of biographic projections is counts of people ('how many?'), counts of transitions ('how often?') and durations ('how long?').

d. A fourth feature of the model that is not innovative in itself but that paves the way to significant innovations in demographic projections is the pivotal role of (instantaneous) transition rates. Transition rates are the basic parameters of the demographic and biographic models, both the macrosimulation model Mac and the microsimulation model Mic. These rates are the dependent variables of regression models that link the rates to causal factors, covariates and other explanatory factors. The regression models are known as survival models, duration models and event history models. Blossfeld (1998, p. 245) offers the following link between event history models that predict transition rates and theories of behaviour, illustrating the causal reasoning underlying transition rate models. A sequence of state occupancies, i.e. the lifepath or life trajectory, represents a course of action of the individual under study. The transition rates describe the propensity of the individuals under study to change their course of action. The effect of causal and other factors is to change the transition rates. That causal reasoning underlying transition rate models is adopted in the model presented in this paper. It makes the model ideally suited to assess whether better behavioural theories lead to an increased forecasting performance. The question whether better knowledge improves forecasts and what the most effective use is of substantive knowledge on demographic and behavioural processes have preoccupied demographers for decades (see e.g. Keyfitz, 1982; Sanderson, 1998). In the context of the forecasting of the health status in an ageing population, Manton et al. formulated the issue as follows: "Current forecasting procedures are often based on empirical extrapolations and do not directly reflect physiological processes at the individual level or the mixture of individuals in a cohort. The failure to deal with individual aging trajectories, and their cohort mix, makes it difficult to use epidemiological and biomedical evidence on the impact of health changes on the organism in forecasts." (Manton et al., 1993, p. 25). The major strength of the model is the pivotal role of the transition rates. The fundamental nature of the transition rates facilitates the accommodation in biographic and demographic forecasting of substantive knowledge on both causal behavioural mechanisms and random mechanisms.

2.2. Main features of the biographic model

In its simple form, the multistate model describes the collective life history or biography of a cohort and disregards differences between cohort members. The biography of a single cohort is usually obtained using the multistate life table method. The life table is among the oldest models applied in demography. The period life table has two interpretations (see e.g. Preston et al., 2001, pp. 51ff). The first is that of a description of a hypothetical cohort. The traditional life table describes the mortality experience and the multistate life table describes the mortality and mobility experience of a cohort. The second perspective on the life table is that of a stationary population. The life table presents several features of a population in static equilibrium. The second perspective is extended in projection models. Examples of multistate projection models include LIPRO (Van Imhoff and Keilman, 1991), MUDEA (Willekens and Drewe, 1984), and PROFAMY (Zeng Yi et al., 1997). The models are designed to describe and project changes in the population composition at the macro level. They concentrate on the position individuals occupy in the collective biography at consecutive points in time and the population structure that results. They do not address the prognosis of durations of the stages of life or episodes in the lives of people, which is the subject of biographic projections.

The link between the traditional macro-level models and the new micro-level models is the multistate life table, and more particularly the transition rates by age, state of origin and state of destination. Transition rates measure the propensity to change attributes[5]. Differences in transition rates are often expressed as ratios of transition rates, which are known as relative risks. The biographic consequences of unobserved differences between individuals may also be assessed. They are described by mixture models and random effect models. Mixture models classify people in a finite number of categories. Random effect models assume a continuous distribution of unobserved individual characteristics.

[5] Transition rates are obtained by dividing the number of occurrences (events) during an age or time interval by the total duration of exposure during that interval. Transition rates are therefore also referred to as occurrence-exposure rates. Many models in demography and epidemiology rely on probabilities or on types of rates that differ from occurrence-exposure rates. Such rates are obtained by dividing the number of events during an interval not by the population exposed to the risk of the event but by a different population, usually a population that includes people that are not at risk of the event. For instance, the first birth rate is quite often obtained by dividing the number of women (of a given age) that give birth to a first child by the total number of women (of that age) rather then by the number of childless women. It should be stressed that transition rates are occurrence-exposure rates. In epidemiology, the method that allocates a key position to occurrence-exposure rates is referred to as the person-years approach because exposure is generally measured in person-years.

The multistate life table and the multistate projection model are adequately documented in the literature (see e.g. Rogers, 1975; Schoen, 1988; Manton and Stallard, 1988). In this report, no fundamental distinction is made between the life table and the projection model. In essence, the life table is a projection model. The multistate life table relates the characteristics of a cohort of people at a given birthday to the characteristics at the previous birthday and the events in the past year.

The combination of a transition rate model (hazard model) and a multistate life table constitutes the main ingredient of the proposed biographic model for functional population projections and the prediction of the life course. The combination of regression models and life tables was introduced more than thirty years ago by Cox (1972) and, for multistate life tables, more than 10 years ago by Gill (1992). Both authors give a central position to transition rates. Transition rates depend on risk factors and other determinants. A risk factor is defined as a factor that is causally related to an outcome. The concept originated in epidemiology, where the identification of the causal link is an important element of the etiology of a disease. In many cases, a causal link cannot be determined and the association between predictor and outcome is a statistical one. In the prediction of the life course, risk factors and other factors are evaluated in terms of their predictive performance and not their explanatory power. Two comments are warranted here. First, the link between a risk factor and the outcome is probabilistic. It means that the presence of a risk factor changes the *probability* of an event or the *expected* duration of an episode. The significance of an event lies in the consequences to the life history of an individual (Peeters et al., 2002). Second, several risk factors may change during the course of life. Modifiable risk factors are particularly relevant in the design of health policies and public health programmes. They should also be considered in forecasting since the health outcomes (and mortality) depend on modifiable risk factors. For instance, when more people stop smoking, start exercising and start eating healthy, the long-term consequences will be increased survival, possibly associated with longer periods of chronic disease. For a recent illustration of risk factor dynamics in survival analysis using multistate models, see Akushevich et al. (2005).

3. The age-time framework: continuous versus discrete

In demographic and biographic analysis, individuals and events are positioned in different time scales. Each time scale is related to a reference event, sometimes referred to as *event-origin*, and measures the time elapsed since the reference event in units of minutes, hours, days, months or years. Age (individual time) and calendar time (historical time) are common time scales. If time is a continuous variable, the position of individuals and events is given in exact time. Exact time is also referred to as instantaneous time (see e.g. Arthur and Vaupel, 1984). If time is a discrete variable, the position is given in completed time units (e.g. years). The two time scales may be visualised by the Lexis diagram and its extension, the Lexis surface. For an individual, the Lexis diagram shows the age (x) at a given calendar time (t). In this paper, the Lexis diagram is used to clarify the transition from continuous time to discrete time and the different types of discrete time intervals used in demographic analysis. Figure 1 displays a Lexis diagram. The unit of time is a h years. The diagram shows lifelines for three individuals (a, b and c). Individual a is exact age x at time t and is therefore born at instantaneous time t-x. At time t, individual c celebrates his (x+h)st birthday. He is born at t-x-h, exactly one year earlier than a. Individual b is born somewhere in the year that starts at t-x-h and ends at t-x. All the lifelines between lifeline a and c represent individuals born between t-x-h and t-x. Those individuals reach their x-th birthday between t-h and t, and at exact time t they are aged x in completed years. The lifelines that cross segment PV represent members of the birth cohort who reach exact age x and the lifelines that cross segment VQ represent members of the same birth cohort aged x at time t. The characteristics of the birth cohort may be updated at exact age x or at exact time t. In *cohort analysis*, it is customary to update the experience of members of a birth cohort at consecutive exact ages. In that case, age is a continuous variable and calendar time is a discrete variable. In *period analysis*, and more particularly in demographic projection, it is customary to update the cohort characteristics at consecutive calendar times. Age is a discrete variable (age interval) and historical time is a continuous variable.

Members of the birth cohort experience events at various points in time. The timing of events is generally not measured in continuous time (date of event, say) but in discrete time (month and/or year of occurrence). In case of discrete time, or a combination of continuous time and discrete time, different observation windows emerge. They are of fundamental importance in biographic analysis. Consider individual b. The individual experiences four events denoted by A, B, C and D. Event A occurs in year t-h while the individual is aged x in completed years. Event B occurs at the same age but one year later, i.e. in year t. Event C occurs in year t but at age x+h in completed years. The events during year t (i.e. between t and t+h) to members of the birth cohort t-x-h (i.e. born between t-x-h and t-x) cluster in the segment VQRS. At the time of the events some cohort members are aged x in completed years (events situated in triangle VQS) while others are aged x+h (events situated in triangle QSR). Similarly, the events experienced while the cohort members are aged x in completed years cluster in the segment PVSQ. Some experience the event in year t-h (events in triangle PVQ) while others experience the event in year t (events in triangle VQS). The combination of continuous and discrete time in cohort analysis and period analysis results in grouping of events in intervals that are delineated by age, cohort and time (period) segments. The grouping is particularly relevant in observational studies since the exact timing of events is often not available.

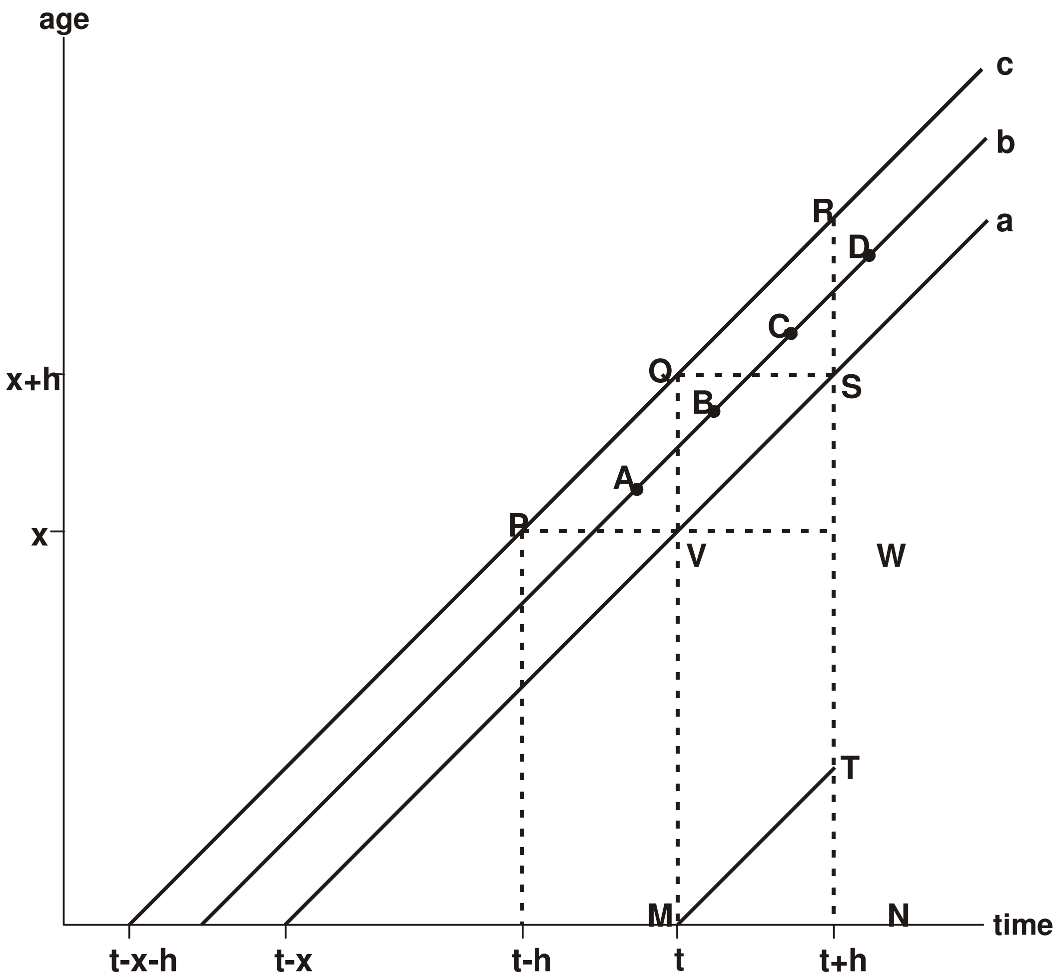

Figure 1 Lexis diagram

Four observation intervals or observation windows may be distinguished:

i. *Period observation (or period-age observation).* A period observational plan records the calendar year in which an event occurs as well as the age of the person at the time of the event (age at last birthday). The age is recorded in completed years. In Figure 1, the period observation interval is represented by QSWV. The interval covers two cohorts. The occurrence-exposure rates based on the period observational plan are referred to as period rates.

ii. *Cohort observation (or cohort-age observation).* A cohort observational plan records for a person experiencing an event the cohort to which the person belongs as well as the age in completed years at the time of the event. In Figure 1, it is parallelogram QSVP. The observation period extends over two calendar years. The rates based on the cohort observational plan are cohort rates.

iii. *Period-cohort observation.* A period-cohort observational plan records the calendar year in which an event occurs as well as the cohort to which the person belongs. In Figure 1, it is parallelogram QRSV. The observation interval covers two age groups. Recording the cohort in the period-cohort observational plan is equivalent to recording the age at the beginning of the interval (i.e. at time t) or at the end of the interval (i.e. at time t+h). It is also equivalent to recording age in period difference. The latter is obtained by subtracting the year of birth from the year of occurrence of the event under study (Wunsch and Termote, 1978, p. 10). Rates based on period-cohort observational intervals are period-cohort rates.

iv. *Age-period-cohort observation.* An age-period-cohort observational plan records for a person experiencing an event the age in completed years at time of the event, the calendar year in which an event occurs as well as the cohort to which the person belongs. In Figure 1, triangle VQS contains the events that occur in year t to individuals aged x, born in year t-x-h. The observation interval covers one age groups, one year and one birth cohort. Rates based on the age-period-cohort observational plan are age-period-cohort rates.

For cohort analysis, including life table analysis, the age-cohort observational plan is generally preferred because of the interest in the position of cohort members at exact ages (at their birthdays). For period analysis and more particularly for projection, a period-cohort observational plan is the ideal one. The period observational plan often serves as an approximation to one of the other plans. Many texts in demography implicitly assume that the data are period data (type i). See e.g. Keyfitz (1968, p. 9), Rogers (1975, 1995, p. 41) and Schoen (1988, p. 11).

The models presented in this paper are formulated in continuous time with intervals that are infinitesimally small. If time is denoted by t, an infinitesimally small interval is denoted by dt. Analysis in continuous time involves calculus[6]. The parameters of continuous-time transition models are instantaneous rates of transition or transition intensities. Many applications call for discrete age (age intervals) and/or discrete time. For instance, the age composition of people at a given point in time (e.g. 1st January 2006) is generally expressed in intervals of one or five years[7]. The very concept of a cohort implies discrete time. Although transition models may be formulated in continuous age and time, discrete age and time intervals are introduced when the models are applied or when their parameters are estimated from data. The above mentioned observation intervals represent different types of data and transition rates estimated from

[6] Readers not familiar with calculus may visualize an infinitesimally small interval dt as a pixel on a time axis. An interval from t to t+h consists of all pixels dt between t and t+h. The width of dt may be associated with the resolution. A low resolution is related to a relatively large dt and a high resolution required that dt is small. An observation window is defined by two time scales, i.e. calendar time and age. It may be thought off as a screen of finite size. An area of the observation window is defined the segment (t,t+h) along the time axis and the segment (x,x+h) along the age axis. Calculus is used to count the pixels in the area and to determine their characteristics. For instance, the interval VQS consists of pixels in the area delineated by the time segment (t,t+h), the age segment (x,x+h) and for which at time $t+\tau$, the age is between $x+\tau$ and x+h. Hence τ varies between 0 and h. A pixel may be represent a day in the life of an invidivual or a group of individuals of a given exact age at a given exact time. The screen is basically a contour map and a pixel the grid of the contour map. A contour map is basically a projection of the Lexis surface onto the age-time plane.

[7] In cohort studies and survival analysis, a group of cohort members in a particular state at a particular point in time is referred to as a *prevalent cohort*. A prevalent cohort is a group of people who satisfy a certain criterion when they enter the study. They generally experienced an initiating event.

the data. *With each observational plan is associated a type of discrete-time transition rate*. The empirical transition rate is an occurrence-exposure rate obtained by dividing the number of transitions (events) during a given observation window by the total duration of exposure during the same observation window. As a result, the rates are distinguished on the basis of the observation window. The distinction is between period rates, cohort rates (age-cohort rates), period-cohort rates and age-period-cohort rates. The estimation of the transition rates requires the measurement or estimation of exposure time. For an individual, exposure time is the time the individual is at risk of an event. The period of exposure being measured may extend over the entire lifetime or may be a segment fo the life course or an age interval. Consider the period-cohort observation window VQRS. The period-cohort rate is the ratio of the number of times a given event is recorded during the interval VQRS and the total time spent during the interval by individuals who are at risk of experiencing the event. Consider the individual with lifeline b. He experiences two events during year t, event B and event C. Event B occurs at age x and event C at age x+h (both ages in completed years). Individual with lifeline b becomes at risk of event B in year t when he crosses segment VQ. The duration of exposure to the risk of experiencing event B in year t is the time between entry into the risk period and event B. The duration at risk of experiencing event C during year t is the duration between event B and event C. Note that the exposure to C extends over two ages, which is typical for period-cohort observational plans. In determining an individual's duration of exposure to the risk of experiencing an event, several situations may be distinguished (Willekens, 2005, p. 111):

a. The individual is present at t and at t+h and does not experience the event during the period-cohort observation window from t to t+h. The duration of exposure is equal to 12 months.
b. The individual is present at t and experiences the event during the interval, at exact time (date) τ. The duration of exposure is τ - t.
c. The individual enters the population at risk at exact time ξ during year (t,t+h), does not experience the event, and is present at the end of the interval. The duration at risk is t+h-ξ.
d. The individual enters the population at risk at exact time ξ during the interval (t, t+h) and experiences the event at exact time τ. The duration at risk is ξ-τ.
e. The individual is present at t and leaves the population at risk at exact time ψ before reaching time t+h without experiencing the event. The duration at risk is ψ-t.
f. The individual enters the population in ξ and leaves the population at ψ. The duration at risk is ψ-ξ.

The different situations represent different schemes of censoring and truncation (Klein and Moeschberger, 1997; Willekens 2005, pp. 105ff). In statistics, the method of estimating transition rates by careful measurement of exposure time is referred to as the Nelson-Aalen estimation method, which is part of the theory of counting processes (Andersen et al., 1993).

Whereas the Lexis diagram situates an individual in the age-time framework, the Lexis surface focuses on counts of people. The Lexis surface visualizes the number of individuals of age x at time t. The concept of Lexis surface was developed by Arthur and Vaupel (1984) to illustrate the change of population density defined over age and time.

The surface is defined by the population densities at different ages and different times. The "navigational" calculus, as they call it, encompasses fundamental relationships of the dynamics of a population and builds on the work by McKendrick (1926) and von Förster (1959), and the theory of partial differential equations (see Keyfitz and Keyfitz, 1997). The concept of population density used in this paper differs from that by Arthur and Vaupel in three respects. First, the population density is derived from the statistical theory for the analysis of survival and transition data. The population density at an instantaneous age and time is the outcome of an initial condition and the instantaneous rates of transition over a given period of time. Second, the Lexis surface is extended into the future. Third, not only the number of individuals at a given age and time is considered but also the characteristics of the individuals. As an outcome, the correspondence between individual life cycle behaviour and population characteristics, that is demonstrated by the generalized relationships developed by Arthur and Vaupel, is extended beyond standard fertility and mortality and includes other major transitions that people experience in the course of life (for a discussion on the correspondence, see Arthur and Vaupel, 1984, p. 223).

4. The individual biography

In some fields of study, such as medicine, insurance and criminology, the accurate prediction of individual life paths receives considerable attention. Person-centered or patient-centered computational methods constitute a rapidly developing field of study. In the field of cardiovascular disease epidemiology, the Framingham Heart Study (FHS) has been instrumental for the development of statistical methods. Hense (2004) reviews and discusses methods aimed at an accurate prediction of the probability of a cardiovascular event in individuals free of cardiovascular disease (CVD). The first attempt to predict an individual's risk of CVD on the basis of risk factors (age, sex, systolic blood pressure, cholesterol level, smoking and body mass) dates back to the 1970s. Current guidelines for e.g. cholesterol testing and management uses projections of 10-year coronary heart disease (CHD) absolute risk (probabilities of events). Threshold values of absolute risks are postulated to identify 'high risk' individuals who are eligible for intensified management and drug treatment. If the probabilities are not determined correctly, then part of the population will be either inappropriately elected or withheld from medical treatment (Hense, 2004, p. 236). The best strategy for the prevention of CVD is to take into consideration various risk factors simultaneously. Anderson et al. (1991a, 1991b; see also Mamun, 2003) developed a flexible model to predict a patient's 5- and 10-year probability of CHD or CVD. He took into account several risk factors and applied the method to FHS data. It is a multivariate regression model (accelerated failure time model). Later, new data were used and new methods were developed. In Europe, the SCORE project needs mentioning (Conroy et al., 2003). The technique is used extensively in medical sciences to determine risk profiles and the need for medication or other interventions that reduce the risk of an adverse outcome such as a disease. Although valid estimates of disease occurrence at the individual level are important to determine treatment, the expected benefit of treatment is important too. The expected benefit may be in terms of number of years of life gained by the treatment and these years may be weighted by the quality of life. One such measure is the disability-adjusted years of life (DALYs), a popular indicator in intervention analysis. This brief account demonstrates the need for biographic projections at the individual level and the biographic indicators considered: probabilities, counts of events, sojourn times and demographic and health indicators based on these basic measures. The basic indicators relate to three aspects of the life course: state occupancies, state transitions and durations of stay. These biographic indicators are the subject of this section.

The biographic indicators are derived for individual k with date of birth t_b. As k grows and develops, he experiences events and changes functional states. In Section 4.1, three basic random variables are introduced to denote state occupancies and state transitions. When k reaches age x the life course up to x is known (in principle). What it means is that the random variables have taken on specific values or realizations. Whether these values have been recorded or not, is a measurement issue. The complete history of state occupancies and times of change up to age x is often referred to as a *sample path* and denoted by ${}_k\mathbf{\Phi}(x,t)$ (Tuma and Hannan, 1984, p. 48). The realizations of the random variables at future ages remain unknown but the distribution of the random variables is assumed to be known. In other words, the value a random value will take on at some future date is unknown but the probability distribution of possible values is known. A prediction of future values uses the information on the distribution to determine plausible values and most likely values of the random variable given the available information such as age, time, current characteristics and past events and experience. The prediction of a future life path involves the distributions of the random variables mentioned and other random variables and the appraisal of the parameters of the distributions. The expected value of a random variable represents the single most important parameter of the distribution of the possible values of the random variable.

The life course of k is a stochastic process, i.e. a sequence of random variables. The stochastic process is governed by transition intensities that depend on age, time and the life history. Section 4.2 discusses the link between state probabilities, transition probabilities and transition intensities. The transition intensities, which are instantaneous rates of transition, describe k's behaviour in continuous time, i.e. the behaviour during infinitesimally small intervals following exact ages. All biographic indicators can be derived from the initial condition, i.e. the state occupied at birth, and the transition intensities. The relation between the biographic indicators and the instantaneous transition are presented in Sections 4.3 and 4.4. Section 4.3 discusses the exposure function and Section 4.4 the number of transitions. Section 4.5 covers the estimation of transition rates from data.

4.1. Random variables and their expected values

Individual k, who is born at instantaneous time t_b, reaches exact age x at $t = t_b+x$. At that age, individual k can be characterized by a set of attributes and state occupancies. The state space is S ={1, 2, 3, …, I}, with I the size of the state space. The state space includes all possible states. Two types of states are often distinguished: transient states that can be entered more than once and left more than once, and absorbing states that can be entered only. Death is an absorbing state. A system is said to be closed if absorbing states (e.g. death) are treated as part of the state space. In an open system, the state space is restricted to transient states or, if death is the only absorbing state, living states. If the state of dead is included in the state space, the state space is ${}^{D}S = \{1, 2, 3, \ldots, I, D\}$.

Let ${}_kY(x,t)$ be a random variable indicating the state occupied by k at exact age x and exact time t. The variable ${}_kY(x,t)$ is a discrete variable that can take on as many non-zero values as there are states in the state space. In an open system, the value of ${}_kY(x,t)$ at age x is zero if k dies before age x. The state variable ${}_kY(x,t)$ is a random variable. If the value of ${}_kY(x,t)$ is observed, ${}_kY(x,t) = {}_ky(x,t)$, the empirical value. The distribution of the possible values of the random variable ${}_kY(x,t)$ is assumed to be known. The sequence $\{{}_kY(x,t), x \geq 0; t \geq t_b; t = t_b+x\}$ is a stochastic process identifying the states occupied from birth to death. That sequence describes the life course of k. In the case of multiple states, a second approach is more convenient to denote the state occupied. It is the indicator variable ${}_kY_i(x,t)$ that is equal to 1 if k is in state i at age x at t and 0 otherwise. The indicator variable is a binary variable. A product of indicator variables is a new indicator variable that indicates whether or not individual k occupies particular states at consecutive ages or points in time. The product ${}_kY_i(x,t)* {}_kY_j(x+h,t+h)$ is one if individual k occupies state i at age x and time t, and j at age x+h and time t+h. It is zero otherwise[8]. The product is denoted by ${}_k^hY_{ij}(x,t)$. An alternative way of writing the variable is ${}_kY(x_1, x_2, t)$ where $x_1 = x$, $x_2 = x+h$ and t is the calendar time at x_1. The individual is born at $t-x_1$. The first notation is convenient when not more than two ages are considered and the length of the age interval is fixed. The latter notation is convenient when more than two ages are considered or the age intervals are of different length. For instance, ${}_kY_{ijr}(x_1,x_2,x_3,t)$ is one when individual k, born at $t-x_1$, is in state i at instantaneous age x_1, in state j at age x_2 and in state r at age x_3. Note that ${}_kY_{ijr}(x_1,x_2,x_3,t) = {}_kY_i(x_1,t)\ {}_kY_j(x_2,t_2)\ {}_kY_r(x_3,t_3)$. The transition indicator variable ${}_k^hY_{ij}(x,t)$ should be distinguished from ${}_k^hY_{j|i}(x,t)$ which indicates the state occupied at x+h *conditional* on the state individual k occupies at x. It is one if individual k who is born at t-x and occupies state i at exact age x, occupies state j at x+h.

A third indicator variable is introduced. It is also a binary (0,1)-variable. It indicates whether or not individual k makes a transition from state i to state j at instantaneous age x and time t. It is denoted by ${}_kY_{ij}(x,t)$. Note that the superscript h is missing. The indicator variable is one if at instantaneous age x and time t, k transfers (jumps) from i to j. Note the difference between this direct transition and the transition measured by comparing the states occupied at two points in time. The transition measured by comparing the states occupied at two consecutive ages (or points in time) is a *discrete-time transition*. The points in time define an interval (age or time interval). The probability of such a transition is a discrete-time transition probability. As the interval becomes smaller and tends to zero, the transitions are measured in continuous time. A transition measured in continuous time is referred to as a *direct transition* between states or immediate jump

[8] The product is zero if at least one of the state variables is zero.

between positions (van Imhoff and Keilman, 1991, p. 28; Rajulton, 1999). Discrete-time transitions are identified by the indicator variable ${}_{k}^{h}Y_{ij}(x,t)$ that is one if individual k occupies state i at instantaneous age x and state j at age x+h, and zero otherwise. Direct transitions are identified by the indictor variable ${}_{k}Y_{ij}(x,t)$ that takes on the value 1 if individual k whose date of birth is t-x, makes a move from state i to state j at instantaneous (exact) age x, i.e. in the infinitesimally small interval following x. It is zero otherwise. The interval is sufficiently small to exclude multiple transitions, i.e. at most one transition may occur during the interval. The indicator variable ${}_{k}Y_{j|i}(x,t)$ is one of individual k who is born at t-x and in state i at exact age x jumps to j during the infinitesimally small interval following x. The distinction between discrete-time transitions and transitions in continuous time is consistent with the traditional distinction between two approaches to microsimulation modeling: continuous-time modeling and discrete-time modeling (see e.g. Galler, 1997; O'Donoghue, n.d., p. 13).

In a number of applications, x_1 is a *reference age* and the life course beyond x_1 is conditioned on the state occupied at the reference age. For instance, one might be interested in the fertility career of women who are childless at age 30. Another illustration of the use of a reference age is the employment career of individuals who are unemployed or out of the labour force at age 30.

The indicator variables are binary variables that can take on two values only, zero and one, say. The expected value of the indicator variable is a probability. The probability that individual k occupies state i at exact (instantaneous) age x and exact time t is the *expected value* of the state variable ${}_{k}Y_i(x,t)$ and is denoted by ${}_{k}\ell_i(x,t)$:
${}_{k}\ell_i(x,t) = E\left[{}_{k}Y_i(x,t)\right] = \Pr\left\{{}_{k}Y_i(x,t)=1\right\} = \Pr\left\{{}_{k}Y(x,t)=i\right\}$.
It is the *state probability*.

The probability that k occupies state i at age x and time t depends on two processes: survival and mobility. The probability of surviving to age x is ${}_{k}\ell(x,t)$. The measure is irrespective of the state occupied at x and t. Note that ${}_{k}\ell(x,t) = \sum_{i=1}^{I} {}_{k}\ell_i(x,t) = {}_{k}\ell_{+}(x,t)$. The conditional probability that k occupies state i at age x, provided k is alive at x, is the conditional probability ${}_{k}\pi_i(x,t) = \Pr\left\{{}_{k}Y(x,t)=i \,\middle|\, {}_{k}Y(x,t)>0\right\}$ with $\sum_{i=1}^{I} {}_{k}\pi_i(x,t)=1$. The relation between the unconditional state probability ${}_{k}\ell_i(x,t)$ and the conditional state probability ${}_{k}\pi_i(x,t)$ is: ${}_{k}\ell_i(x,t) = {}_{k}\ell(x,t) * {}_{k}\pi_i(x,t)$. Note that the distinction between conditional and unconditional state probabilities is related to the treatment of absorbing states such as dead. If dead is one of the states, then $\sum_{i=1}^{I} {}_{k}\ell_i(x,t)=1$. If dead is taken into account but a separate state of dead is not considered, then $\sum_{i=1}^{I} {}_{k}\ell_i(x,t)\le 1$.

If k's life history up to age x is known, the future life path may be conditioned on it. Let ${}_{k}\boldsymbol{\Phi}(x,t)$ denote the life course of k up to x and t. The information includes current characteristics, characteristics at a much earlier age (e.g. childhood), current and previous living conditions or living environment such as housing and economic conditions, and features of the entire life history. If the life path beyond age x and time t depends on the life history ${}_k\mathbf{\Phi}(x,t)$, the life path beyond x is given by the stochastic process $\{{}_kY(x,t, {}_k\mathbf{\Phi}(x,t)), x \geq 0; t=t_b+x\}$. The expected value of the random variable ${}_kY(x,t)$ at x and t depends on the characteristics of individual k at x and t, and the entire life history up to age x or part of that life history. Assume the state occupied by k at three instantaneous ages and associated points in time are known, at (x_1,t_1), (x_2,t_2) and (x_3,t_3) . The individual is born at t_3-x_3 (which is equal to t_1-x_1, etc). The probability that individual k who was in state ${}_ky(x_1,t_1)$, at age x_1 at time t_1, in state ${}_ky(x_2,t_2)$ at age x_2 at time t_2 and in state ${}_ky(x_3,t_3)$ at age x_3 at time t_3, and with a set of characteristics denoted by ${}_k\mathbf{\Phi}^*(x,t)$, is in state j at age x_4 at time t_4, is

$$Pr\{\, {}_kY(x_4,t_4) = j \mid {}_ky(x_3,t_3),\ {}_ky(x_2,t_2),\ {}_ky(x_1,t_1);{}_k\mathbf{\Phi}^*(x,t)\} \qquad x_4 > x_i \quad i = 1, 2, 3$$

where ${}_k\mathbf{\Phi}^*(x,t)$ denotes a selection of contemporary and prior characteristics and experiences. It is often assumed that only the most recent state occupancy is relevant:

$$Pr\{\, {}_kY(x_4,t_4) = j \mid {}_ky(x_3,t_3),\ {}_ky(x_2,t_2),\ {}_ky(x_1,t_1);{}_k\mathbf{\Phi}^*(x,t)\} = Pr\{\, {}_kY(x_4,t_4) = j \mid {}_ky(x_3,t_3);{}_k\mathbf{\Phi}^*(x,t)\}$$

If the state occupied at x_3 is i, then ${}^{x_4-x_3}_{\;k}p_{ij}[x_3,t_3,{}_k\mathbf{\Phi}^*(x,t)] = \Pr\{\,{}_kY(x_4,t_4) = j \mid {}_ky(x_3,t_3) = i;\ {}_k\mathbf{\Phi}^*(x,t)\}$. The probability of occupying state j at a given age depends on the state occupied at a previous age and characteristics and experiences. The probability that an individual who occupies state i at age x_3 and time t_3 occupies state j at age x_4 and time t_4 is the *transition probability*. In the remainder of this section, ${}_k\mathbf{\Phi}^*(x,t)$ is omitted for convenience. In addition it is assumed that the interval between x_3 and x_4 is of length h.

The transition probability is denoted by ${}^h_kp_{ij}(x,t)$:

$${}^h_kp_{ij}(x,t) = \Pr\{{}_kY(x+h,t+h) = j \,|\, {}_ky(x,t) = i\}$$

It is the probability that individual k, who is born at instantaneous time t-x **and** occupies state i at x, occupies state j at age x at time t. The transition probability is a discrete-time transition probability, because the transition is measured by comparing the state occupancies at two consecutive ages and points in time. The interval can be of any length. For instance, t could be the time at birth ($t=t_b$) and t+h 40 years later. In that case

$${}^{40}_{\;k}p_{ij}(0,t) = \Pr\{{}_kY(40,t_b+40) = j \,|\, {}_ky(0,t_b) = i\}$$

Transitions and transition probabilities may be situated within the framework of the Lexis diagram. At birth (at time t_b) individual k starts his journey in the Lexis diagram and, together with other members of his cohort, on the Lexis surface. The observed state occupancies ${}_ky(0,t_b)$, ${}_ky(x_1,t_1)$, ${}_ky(x_2,t_2)$ and ${}_ky(x_3,t_3)$ indicate the attributes of individual k at different points of the journey. The random variable ${}_kY(x,t)$ indicates the yet unknown attribute of k at position x and t. The probability of reaching age x is the survival probability. Individual k cannot move freely over the surface but must follow a predetermined direction, namely along the cohort direction (in the "age and time" diagram [see Vandeschrick, 2001, p. 105]). Individuals born during the same interval stay together on the same path over the Lexis surface[9].

The transition probability ${}^{h}_{k}p_{ij}(x,t)$ is a conditional probability. It is the probability that individual k occupies state j at age x+h and time t+h, provided k is in i at x and t. The unconditional transition probability may be expressed as the product of a state probability and a conditional transition probability:

$$\Pr\{\,{}_kY(x+h,t+h)=j\,,\ {}_kY(x,t)=i\}=\Pr\left\{{}^{h}_{k}Y_{ij}(x,t)=1\right\}={}^{h}_{k}p_{ij}(x,t)*{}_k\ell_i(x,t)={}^{h}_{k}\ell_{ij}(x,t)$$

where ${}_k\ell_i(x,t)$ is the state probability and ${}^{h}_{k}\ell_{ij}(x,t)$ is the unconditional transition probability, i.e. the probability that individual k, who is born at t-x, occupies state i at instantaneous age x (and time t) and j at instantaneous age x+h (and time t+h).

Now consider direct transitions. The expected value of ${}_kY_{ij}(x,t)$ is the probability that individual k born at t-x makes a transition from i to j at exact age x. It depends on being alive at x and being in i at that age. The density of a direct transition from i to j at age x and time t is ${}_k\ell_{ij}(x,t)$. It is

$${}_k\ell_{ij}(x,t)=\lim_{h\to 0}\frac{\Pr\left\{{}_kY(x,t)=i,\ {}_kY(x+h,t+h)=j\right\}}{h}=\lim_{h\to 0}\frac{\Pr\left\{{}^{h}_{k}Y_{ij}(x,t)=1\right\}}{h}=\lim_{h\to 0}\frac{{}^{h}_{k}\ell_{ij}(x,t)}{h}$$

for $j \neq i$

The product ${}_k\ell_{ij}(x,t)dx$ is the probability that individual k, who is born at t-x, makes a move from i to j during the infinitesimally small interval dx following age x and time t. It is $E[{}_kY_{ij}(x,t)]\,dx$. The conditional density of a move from i to j provided individual k is alive and in state i at age x is

$${}_k\mu_{ij}(x,t)=\lim_{h\to 0}\frac{\Pr\left\{{}_kY(x+h,t+h)=j \mid {}_kY(x,t)=i\right\}}{h}=\lim_{h\to 0}\frac{{}^{h}_{k}p_{ij}(x,t)}{h} \qquad \text{for } j \neq i$$

The measure ${}_k\mu_{ij}(x,t)$ is the rate at which individual k moves from i to j during the infinitesimally small interval following exact age x and time t. The measure is known as the instantaneous rate of transition or *transition intensity*[10] at age x and time t. The

---

[9] Graphical methods to visualize individual and cohort life histories have not received much attention. The software for the Lexis surface was developed by Andreev (1999, 2002). References to the visualization of the life course include Francis and Pritchard (2000) who developed three-dimension Lexis pencils. For an overview of attempts, see Maltz and Klosak-Mullany (2000).

[10] The intensity may also be written as follows:

$${}_k\mu_{ij}(x,t)=\lim_{h\to 0}\frac{P(x\le X > x+h, t\le T > t+h, J=j \mid X\ge x, I=i)}{h}$$

where X, T, I and J are random variables denoting age, time, state of origin and state of destination, respectively.

probability that individual k advances from state i to state j during the small interval (x, x+dx) is ${}_k\mu_{ij}(x,t)dx$. It is $E[{}_kY_{j|i}(x,t)]\,dx$. The probability density of a direct transition at x and t and the transition intensity are related:

${}_k\ell_{ij}(x,t)={}_k\mu_{ij}(x,t)\;{}_k\ell_i(x,t)$

where ${}_k\ell_i(x,t)$ is the state probability. The expression may also be written as ${}_k\ell_{ij}(x,t)={}_k\mu_{ij}(x,t)\;{}_k\ell_i(x,t)$. The quantity ${}_k\mu_{ij}(x,t_b+x)\,dx$ is the probability that individual k makes a direct transition to j during the infinitesimally small interval (x, x+dx), provided k is in state i at instantaneous age x. It is a conditional probability.

4.2. Transition intensities, state probabilities and transition probabilities

The life course of k is governed by transition intensities that depend on age x, time t, and k's life history up to x: ${}_k\mu_{ij}[x,t,{}_k\mathbf{\Phi}(x,t)]$. The transition intensity ${}_k\mu_{ij}[x,t,{}_k\mathbf{\Phi}(x,t)]$ describes the local behaviour of k at x and t. The argument ${}_k\mathbf{\Phi}(x,t)$ is omitted for convenience. The sum ${}_k\mu_{i+}(x,t)=\sum_{j\neq i}^{I}{}_k\mu_{ij}(x,t)$ is the rate of leaving state i at age x and time t. It is known as the escape rate from i.

The instantaneous rates of transition may be combined in a transition intensity matrix ${}_k\boldsymbol{\mu}(x,t)$:

$${}_k\boldsymbol{\mu}(x,t)=\begin{bmatrix} {}_k\mu_{11}(x,t) & -{}_k\mu_{21}(x,t) & . & . & -{}_k\mu_{I1}(x,t) \\ -{}_k\mu_{12}(x,t) & {}_k\mu_{22}(x,t) & . & . & -{}_k\mu_{I2}(x,t) \\ . & . & . & . & . \\ . & . & . & . & . \\ -{}_k\mu_{1I}(x,t) & -{}_k\mu_{2I}(x,t) & . & . & {}_k\mu_{II}(x,t) \end{bmatrix}$$

The off-diagonal elements are $-{}_k\mu_{ij}(x,t)$. In a closed system, e.g. a system in which the absorbing state of dead is one of the states of the state space, the diagonal elements ${}_k\mu_{ii}(x,t)$ are defined as minus the sum of the off-diagonal elements in the same column: ${}_k\mu_{ii}(x,t)=\sum_{j\neq i}^{I}{}_k\mu_{ij}(x,t)={}_k\mu_{i+}(x,t)$. ${}_k\mu_{ii}(,x,t_b+x)\,dx$ is the probability that k leaves state i during the interval (x,x+dx). The diagonal elements ${}_k\mu_{ii}(x,t)$ is defined such that ${}_k\mu_{ii}(x,t)-\sum_{j\neq i}^{I}{}_k\mu_{ij}(x,t)=0$. As a result of the definition of the diagonal elements, each column of ${}_k\boldsymbol{\mu}(x,t)$ sums to zero. The quantity ${}_k\mu_{ii}(x,t)$ is non-negative. It is sometimes referred to as the intensity of passage because it relates to the transition from i to any other state different from i (see e.g. Namboodiri and Suchindran, 1987, p. 138). Schoen (1988, p. 65) refers to $-{}_k\mu_{ii}(x,t)$ as the 'force of retention'. The quantity $1-{}_k\mu_{ii}(x,t_b+x)\,dx$ is the probability that k remains in state i during the interval (x,x+dx).

The transition intensity matrix ${}_k\boldsymbol{\mu}(x,t)$ is known as the generator of the stochastic process $\{{}_kY(x,t);\ x \geq 0;\ t = t_b + x\}$ that describes the position at every age of individual k born at time t_b (Çinlar, 1975, p. 256).

In demographic studies, the state of dead is often considered but excluded from the state space and deaths are exits from the system (open system). In that case, ${}_k\mu_{ii}(x,t)$ is defined such that ${}_k\mu_{ii}(x,t) = {}_k\mu_{id}(x,t) + \sum_{j\neq i}^{I} {}_k\mu_{ij}(x,t)$ where ${}_k\mu_{id}(x,t)$ is the instantaneous rate of death at age x and time t. In this paper, the state of dead is not considered unless stated otherwise.

The following relation exists between ${}_k\mu_{ii}(x,t)$ and the transition probabilities:

$$ {}_k\mu_{ii}(x,t) = \sum_{j\neq i}^{I} {}_k\mu_{ij}(x,t) = \sum_{j\neq i}^{I} \lim_{h\to 0} \frac{{}_k^h p_{ij}(x,t)}{h} = \lim_{h\to 0} \frac{1 - {}_k^h p_{ii}(x,t)}{h} $$

Because of the definition of ${}_k\mu_{ii}(x,t)$, we may write the matrix equation

$$ \lim_{h\to 0} \frac{{}_k^h\mathbf{P}(x,t) - \mathbf{I}}{h} = -{}_k\boldsymbol{\mu}(x,t) \qquad (4.1) $$

with

$$ {}_k^h\mathbf{P}(x,t) = \begin{bmatrix} {}_k^h p_{11}(x,t) & {}_k^h p_{21}(x,t) & . & . & {}_k^h p_{I1}(x,t) \\ {}_k^h p_{12}(x,t) & {}_k^h p_{22}(x,t) & . & . & {}_k^h p_{I2}(x,t) \\ . & . & . & . & . \\ . & . & . & . & . \\ {}_k^h p_{1I}(x,t) & {}_k^h p_{2I}(x,t) & . & . & {}_k^h p_{II}(x,t) \end{bmatrix} $$

the matrix of discrete-time transition probabilities. An element ${}_k^h p_{ij}(x,t)$ of ${}_k^h\mathbf{P}(x,t)$ denotes the (conditional) probability that an individual who is born at t-x and occupies state i at exact age x, is in state j exactly h years later. Each column of ${}_k^h\mathbf{P}(x,t)$ sums to unity. The system of differential equations that governs the change in state probabilities may be derived from (4.1) (see e.g. Taylor and Karlin, 1994, p. 363). We write ${}_k^h\mathbf{P}(x,t)$ as ${}_k\mathbf{P}(x, x+h, t)$. The Markov property leads to the Chapman-Kolmogorov equation: ${}_k\mathbf{P}(x, x+h+\Delta h, t) = {}_k\mathbf{P}(x+h, x+h+\Delta h, t+h)\ {}_k\mathbf{P}(x, x+h, t)$. Subtracting ${}_k\mathbf{P}(x, x+h, t)$ from both sides gives

$$ \frac{{}_k\mathbf{P}(x, x+h+\Delta h, t) - {}_k\mathbf{P}(x, x+h, t)}{\Delta h} = \frac{\left[{}_k\mathbf{P}(x+h, x+h+\Delta h, t+h) - \mathbf{I}\right] {}_k\mathbf{P}(x, x+h, t)}{\Delta h} $$

and

$$\lim_{\Delta h \to 0} \frac{{}_k\mathbf{P}(x, x+h+\Delta h, t) - {}_k\mathbf{P}(x, x+h, t)}{\Delta h} = \lim_{\Delta h \to 0} \frac{\left[{}_k\mathbf{P}(x+h, x+h+\Delta h, t+h) - \mathbf{I}\right]}{\Delta h} {}_k\mathbf{P}(x, x+h, t)$$

$$\frac{d\ {}_k\mathbf{P}(x, x+h, t)}{dh} = -{}_k\boldsymbol{\mu}(x+h, t+h)\ {}_k\mathbf{P}(x, x+h, t)$$

To show the equivalence between the matrix expression and the above equations, consider the equation for the diagonal elements

$$\lim_{h \to 0} \frac{{}_k^h p_{ii}(x,t) - 1}{h} = -\sum_{j \neq i}^{I} {}_k\mu_{ij}(x,t) = -{}_k\mu_{i+}(x,t)$$

and for the off-diagonal elements

$$\lim_{h \to 0} \frac{{}_k^h p_{ij}(x,t)}{h} = {}_k\mu_{ij}(x,t)$$

The intensities are the basic parameters of a continuous-time multistate process. Under the restrictive Markov assumption, the probability that an individual leaves a state depends only on the present age and state and is independent of other characteristics, in particular the life history up to but not including the present. In that case, the continuous-time multistate process reduces to a continuous-time Markov chain (CTMC). The CTMC is a stochastic process on a discrete state space in continuous time, $\{Y(t,t_b+x);\ x \geq 0\}$, for which the conditional distribution of a future state at age x + h, given the present state at age x and all past states, depends only on the present state and is independent of the past. The conventional multistate life table is a CTMC (for a review of the life table from the perspective of probability theory, see Hoem and Funck Jensen (1982) and Namboodiri and Suchindran (1987). The Markov chain is a general approach to describe the movement of individuals through the life course (see e.g. Keyfitz and Caswell, 2005, p. 246). Multistate models are extensions of the Markov chain model. Hougaard (2000, p. vi) viewed multistate models as 'the most classical way of analyzing life history data'.

Note that the notation of ${}_k\boldsymbol{\mu}(x,t)$ does not follow the convention in matrix methods. In conventional notation, the first subscript denotes the row and the second the column. In the notation adopted in this paper, the first subscript denotes the column and the second the row. The notational convention adopted in this paper for the intensity matrix also differs from the convention in other fields of study. The configuration of the intensity matrix (and the discrete-time analogue, the matrix of transition rates) used in this paper follows that used in demography (see e.g. Keyfitz, 1968; Keyfitz and Caswell, 2005; Rogers, 1975; Willekens and Drewe, 1984). In several applications, slightly different configurations are used. In probability theory, including the theory of stochastic processes, the intensity matrix is minus the matrix shown above, i.e. $-{}_k\boldsymbol{\mu}(x,t)$. In that matrix, the off-diagonal elements are nonnegative (see e.g. Çinlar, 1975, p. 256; Lancaster, 1990, p. 110; Alho and Spencer, 2005, pp. 168ff). Other authors on multistate models use the transpose of the ${}_k\boldsymbol{\mu}(x,t)$ matrix shown above (with or without a minus sign in fort of it) (see e.g. Schoen, 1988, p. 88; Van Imhoff and Keilman, 1991, p. 26). We use the configuration shown above because (1) the multistate survival function resembles the conventional survival function and (2) it is the convention in multistate demography.

The transition probabilities of the CTMC are found by solving a system of linear differential equations. Let ${}_k^{\tau}p_{ij}(x,t)$ denote the probability that individual k who is born at t-x and occupies state i at exact age x, occupies state j at exact age x+τ. At that age, the time is t+τ. The probability is

$${}_k^{\tau}p_{ij}(x,t) = \Pr\{\, {}_kY(x+\tau, t+\tau) = j \mid {}_ky(x,t) = i\}$$

The transition from i to j may follow an indirect path through any other state r. The transition of the stochastic process from state i at x and t to state j at x+h and t+h allows for different intermediate states. The Chapman-Kolmogorov equation relates the discrete-time transition to the intermediate states. It is:

$${}_k^{\tau}p_{ij}(x,t) = \sum_{r=j}^{I} {}_k^{\tau-\Delta\tau}p_{ir}(x,t)\ {}_k^{\Delta\tau}p_{rj}(x+\tau-\Delta\tau, t+\tau-\Delta\tau)$$

The Chapman-Kolmogorov equation shows how the transition probability from x to x+τ can be decomposed in conditional transition probabilities for smaller intervals. The conditional transition probability is sometimes referred to as the propagator of the stochastic process (Weidlich and Haag, 1988, p. 320). The probability that a process that starts in i at x and t is in j at x+τ is propagated by conditional probabilities (propagators). The above equation may be written as:

$${}_k^{\tau}p_{ij}(x,t) - {}_k^{\tau-\Delta\tau}p_{ij}(x,t)\ {}_k^{\Delta\tau}p_{jj}(x+\tau-\Delta\tau, t+\tau-\Delta\tau) = \sum_{r\neq j}^{I} {}_k^{\tau-\Delta\tau}p_{ir}(x,t)\ {}_k^{\Delta\tau}p_{rj}(x+\tau-\Delta\tau, t+\tau-\Delta\tau)$$

Since

$$\lim_{\Delta\tau\to 0} \frac{{}_k^{\Delta\tau}p_{rj}(x+\tau-\Delta\tau, t+\tau-\Delta\tau)}{\Delta\tau} = {}_k\mu_{rj}(x+\tau, t+\tau)$$

and

$${}_k^{\Delta\tau}p_{jj}(x+\tau-\Delta\tau, t+\tau-\Delta\tau) = 1 - \sum_{r\neq j}^{I} {}_k^{\Delta\tau}p_{jr}(x+\tau-\Delta\tau, t+\tau-\Delta\tau)$$

When Δτ→0, the above equation becomes

$$\lim_{\Delta\tau\to 0}\left[{}_k^{\tau}p_{ij}(x,t) - {}_k^{\tau-\Delta\tau}p_{ij}(x,t)\right] + {}_k^{\tau}p_{ij}(x,t)\sum_{r\neq j}^{I} {}_k\mu_{jr}(x+\tau, t+\tau) = \sum_{r\neq j}^{I} {}_k^{\tau}p_{ir}(x,t)\ {}_k\mu_{rj}(x+\tau, t+\tau)$$

which results in the forward Kolmogorov equation

$$\frac{d\ {}_k^{\tau}p_{ij}(x,t)}{d\tau} = \sum_{r\neq j}^{I} {}_k^{\tau}p_{ir}(x,t)\ {}_k\mu_{rj}(x+\tau, t+\tau) - {}_k^{\tau}p_{ij}(x,t)\ {}_k\mu_{jj}(x+\tau, t+\tau) \qquad \tau \geq 0$$

The expression may be written differently:

$$\frac{d\ {}_k p_{ij}(x,x+\tau,t)}{d\tau} = \sum_{r\neq j}^{I} {}_k p_{ir}(x,x+\tau,t)\ {}_k\mu_{rj}(x+\tau,t+\tau) - {}_k p_{ij}(x,x+\tau,t)\ {}_k\mu_{jj}(x+\tau,t+\tau)$$

where ${}_k p_{ir}(x,x+\tau,t)$ is ${}^{\tau}_{k}p_{ir}(x,t)$ and denotes the probability that individual k, who is born at t-x and occupies state i at x, occupies state r at x+τ.

An intuitive explanation of this equation goes as follows. Suppose individual k, born at t-x and in state i at exact age x, is currently aged x+τ and in state r, $r \neq j$. The instantaneous rate of change into state *j* at x+τ is ${}_k\mu_{rj}(x+\tau,t+\tau)$. If the current state is state *j*, the escape rate from state j is $\sum_{r\neq j} {}_k\mu_{jr}(x+\tau,t+\tau)$. The probability that k, who occupies state i at x, is in state *r* at age $x+\tau$ is ${}^{\tau}_{k}p_{ir}(x,t)$ or ${}_k p_{ir}(x,x+\tau,t)$. The rate of change in the probability of being in state *j* at time $x+\tau$, $\left(d\,{}_k p_{ij}(x,x+\tau,t)/d\tau\right)$, is simply the sum over all r, $r \neq j$, of the probability of being in state *r* times the rate of change into state *j* from state *r*, minus the probability of being in state *j* times the rate out of state *j*. Remembering that we have defined ${}_k\mu_{ii}(x,t)$ as ${}_k\mu_{ii}(x,t) = \sum_{j\neq i} {}_k\mu_{ij}(x,t)$ we arrive at the forward differential equations as stated above.

The differential equations for all i and j may be written in matrix form

$$\frac{d\ {}^{\tau}_{k}\mathbf{P}(x,t)}{d\tau} = -{}_k\boldsymbol{\mu}(x+\tau,t+\tau)\ {}^{\tau}_{k}\mathbf{P}(x,t)$$

where ${}_k\boldsymbol{\mu}(x+\tau,t+\tau)$ is the transition intensity at age x+τ for k. An equivalent expression is

$$\frac{d\ {}_k\mathbf{P}(x,x+\tau,t)}{d\tau} = -{}_k\boldsymbol{\mu}(x+\tau,t+\tau)\ {}_k\mathbf{P}(x,x+\tau,t)$$

where an element $p_{ij}(x,x+\tau,t)$ of ${}_k\mathbf{P}(x,x+\tau,t)$ denotes the probability that individual k, who is born at instantaneous time t-x and occupies state i at exact age x, occupies state j at x+τ. The equation is the forward Kolmogorov differential equation (See also Taylor and Karlin, 1994, p. 363; Schoen, 1988, p. 66; Çinlar, 1975, p. 255).

The system of differential equations may be derived similarly but using matrices. Because of the Markov property, we may write

$${}^{\tau}_{k}\mathbf{P}(x,t) = {}^{\Delta\tau}_{k}\mathbf{P}(x+\tau-\Delta\tau,t+\tau-\Delta\tau)\ {}^{\tau-\Delta\tau}_{k}\mathbf{P}(x,t)$$

where the second term on the right-hand-side is the matrix of discrete-time transition probabilities during interval from x to x+τ-Δτ and the first term is the transition probability during the interval x+τ-Δτ to x+τ.

Subtracting ${}^{\tau-\Delta\tau}_{k}\mathbf{P}(x,t)$ from both sides gives

$${}^{\tau}_{k}\mathbf{P}(x,t) - {}^{\tau-\Delta\tau}_{k}\mathbf{P}(x,t) = \left[{}^{\Delta\tau}_{k}\mathbf{P}(x+\tau-\Delta\tau,t+\tau-\Delta\tau) - \mathbf{I}\right]\ {}^{\tau-\Delta\tau}_{k}\mathbf{P}(x,t)$$

Taking the limit

$$\lim_{\Delta\tau\to 0}\left[{}^{\tau}_{k}\mathbf{P}(x,t) - {}^{\tau-\Delta\tau}_{k}\mathbf{P}(x,t)\right] = \lim_{\Delta\tau\to 0}\left[{}^{\Delta\tau}_{k}\mathbf{P}(x+\tau-\Delta\tau, t+\tau-\Delta\tau) - \mathbf{I}\right] {}^{\tau-\Delta\tau}_{k}\mathbf{P}(x,t)$$

leads to the differential equation that describes the dynamics of the multistate system:

$$\frac{d\ {}^{\tau}_{k}\mathbf{P}(x,t)}{d\tau} = -{}_{k}\boldsymbol{\mu}(x+\tau, t+\tau)\ {}^{\tau}_{k}\mathbf{P}(x,t)$$

The solution to the system of differential equations is

$${}^{h}_{k}\mathbf{P}(x,t) = \exp\left[-\int_0^h {}_{k}\boldsymbol{\mu}(x+\tau, t+\tau)\, d\tau\right]$$

We now turn to state probabilities and return to the solution of the system of differential equations later. Recall the state probability ${}_{k}\ell_i(x,t)=E[{}_{k}Y_i(x,t)]$. It denotes the probability that individual k, who is born at t-x, is in state i at exact age x and exact time t. Let ${}_{k}\mathbf{k}(x,t)$ denote the vector of state probabilities ${}_{k}\ell_i(x,t)$:

$${}_{k}\mathbf{k}(x,t) = \begin{bmatrix} {}_{k}\ell_1(x,t) \\ {}_{k}\ell_2(x,t) \\ . \\ . \\ {}_{k}\ell_I(x,t) \end{bmatrix}$$

The state probabilities at x+h and t+h is

$${}_{k}\mathbf{k}(x+h, t+h) = {}^{h}_{k}\mathbf{P}(x,t)\ {}_{k}\mathbf{k}(x,t) = \exp\left[-\int_0^h {}_{k}\boldsymbol{\mu}(x+\tau, t+\tau)\, d\tau\right] {}_{k}\mathbf{k}(x,t)$$

The state probabilities at x+h and t+h depend on the state probabilities at x and t ${}_{k}\mathbf{k}(x,t)$ and the transition intensities during the interval (x,x+h) and (t,t+h). The state probabilities at x and t are related to the state occupancies at birth (initial condition) and the transition intensities from age 0 to age x:

$${}_{k}\mathbf{k}(x,t) = \exp\left[-\int_0^x {}_{k}\boldsymbol{\mu}(\tau, t_b+\tau)\, d\tau\right] {}_{k}\mathbf{k}(0, t_b)$$

where t_b is the date of birth of individual k. The expression is the solution of the differential equation

$$\frac{d\ {}_{k}\mathbf{k}(x, t_b+x)}{dx} = -{}_{k}\boldsymbol{\mu}(x, t_b+x)\ {}_{k}\mathbf{k}(x, t_b+x)$$

which describes the behaviour of k at instantaneous age x and time t in terms of the density of the different transitions. The system of differential equations where ${}_{k}\mathbf{k}(x,t)$ is a vector of state probabilities at time t and ${}_{k}\boldsymbol{\mu}(x,t)$ is a matrix of transition intensities between the states, is a fundamental equation of probability theory. In the physics literature that description of the CTMC is known as the *master equation*. The term has later been adopted by other disciplines, such as sociology (Weidlich and Haag, 1983), geography (Weidlich and Haag, 1988) and economics (Aoki, 1996, 2004). Aoki gives a description of the master equation that fits well in this paper: 'The master equations describe time evolution of probabilities of states of dynamic processes in terms of the probability transition rates and state occupancy probabilities.' (Aoki, 1996, p. 116). The equation is a flow equation that accounts for the flows into and out of states. For a very brief overview of applications in the social sciences, see Aoki (1996, p. 116). For an introduction to the master equation from a physics perspective, see Hizanidis (2002) and for an elaborate but technical treatment, see Gardiner (2004). Note that in this paper, the evolution of the state probabilities (or state occupancy probabilities) is studied in two time scales, calendar time and age. In the social sciences the CTMC was introduced by Coleman (1964), one of the founding fathers of mathematical sociology, and has been used extensively by Bartholomew (1982, pp. 85ff).

The transition intensities ${}_k\boldsymbol{\mu}(x,t)$ are the fundamental parameters of the multistate models for life history analysis and biographic projection. The transition intensity ${}_k\mu_{ij}(x,t)$ is decomposed into two components: a generation component and a distribution component. The generation component is the intensity of leaving the state of origin (exit rate, escape rate). The distribution component is the probability of a given destination, conditional on leaving the state of origin. The transition intensity may be written as ${}_k\mu_{ij}(x,t) = {}_k\mu_{i+}(x,t) * {}_k\xi_{ij}(x,t)$ with ${}_k\mu_{i+}(x,t)$ the instantaneous rate of leaving state i (escape rate, exit rate) and ${}_k\xi_{ij}(x,t)$ the probability that individual k who leaves state i selects j as the destination ($j \neq i$). It is the conditional probability of a *direct transition* from i to j (conditional on leaving i) (see e.g. Chiang, 1984, p. 250). In migration analysis and multiregional demography, direct transitions are generally referred to as *moves* (Rogers et al., 2002). Probabilities of a direct transition are estimated in e.g. LIFEHIST, a packaged developed by Rajulton (1999) at the University of Western Ontario. Note that ${}_k\xi_{ij}(x,t) = \dfrac{{}_k\mu_{ij}(x,t)}{{}_k\mu_{i+}(x,t)}$.

The off-diagonal elements of ${}_k\boldsymbol{\mu}(x,t)$ may be replaced by $-{}_k\mu_{i+}(x,t)\,{}_k\xi_{ij}(x,t)$. The diagonal elements are ${}_k\mu_{i+}(x,t)$. The $\boldsymbol{\mu}$-matrix may be written as

$$\begin{bmatrix} {}_k\mu_{11}(x,t) & -{}_k\mu_{21}(x,t) & . & . & -{}_k\mu_{I1}(x,t) \\ -{}_k\mu_{12}(x,t) & {}_k\mu_{22}(x,t) & . & . & -{}_k\mu_{I2}(x,t) \\ . & . & . & . & . \\ . & . & . & . & . \\ -{}_k\mu_{1I}(x,t) & -{}_k\mu_{2I}(x,t) & . & . & {}_k\mu_{II}(x,t) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -{}_k\xi_{21}(x,t) & . & . & -{}_k\xi_{I1}(x,t) \\ -{}_k\xi_{12}(x,t) & 1 & . & . & -{}_k\xi_{I2}(x,t) \\ . & . & . & . & . \\ . & . & . & . & . \\ -{}_k\xi_{1I}(x,t) & -{}_k\xi_{2I}(x,t) & . & . & 1 \end{bmatrix} \begin{bmatrix} {}_k\mu_{1+}(x,t) & 0 & . & . & 0 \\ 0 & {}_k\mu_{2+}(x,t) & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & {}_k\mu_{I+}(x,t) \end{bmatrix}$$

Note also that the expression ${}_k\mu_{ij}(x,t) = {}_k\mu_{i+}(x,t) * {}_k\xi_{ij}(x,t)$ is that of a competing risk model or a transition rate model with multiple destinations (Blossfeld and Rohwer, 2002).

In the terminology of competing risks, the first term is the escape rate from i and the second term (destination) indicates the destination, given departure from i. The escape rate determines the timing of the transition. If the waiting time to a transition and the direction of the transition are independent, then the timing of change and the direction of change can be studied separately (Çinlar, 1975, pp. 246ff). The timing of change is governed by the escape rates ${}_k\mu_{i+}(x,t)$. The probability of staying in i from x to x+h is the survival probability $\exp[-\int_0^h {}_k\mu_{i+}(x+\tau,t+\tau)\,d\tau]$. The probability of a direct transition (at least one) from state i to state j during the interval (x,x+h) is

$$\int_0^h {}_k\mu_{ij}(x+\tau,t+\tau)\; {}_k^\tau p_i^*(x,t)\,d\tau = \int_0^h {}_k\mu_{ij}(x+\tau,t+\tau)\exp[-\int_0^\tau {}_k\mu_{i+}(x+\xi,t+\xi)\,d\xi]\,d\tau$$

where ${}_k^\tau p_i^*(x,t)$ is the probability that individual k, who in i at exact age x, stays in i on a continuous basis at least till age x+τ. Schoen (1988, p. 81 and p. 103) refers to the probability of a direct transition as 'ξ-probabilities' or probabilities of first transfer. A direct transition from i to j is a transition in one step.

The successive states visited are determined by the transition matrix of the Markov chain ${}_k\boldsymbol{\xi}(x,t)$. If the escape rate is constant [${}_k\mu_{i+}(x+\tau,t+\tau)={}_k^h m_{i+}(x,t)$ for $0 \le \tau < h$] then the sojourn time in state i (time between direct transition or jumps) has an exponential distribution with parameter ${}_k m_{i+}$. If the escape rate is independent of the state of origin [${}_k^h\mathbf{m}_{i+}(x,t)={}_k^h\mathbf{m}(x,t)$ for all i] , then the probability of exactly n direct transitions (jumps, moves) between x and x+h is given by the Poisson distribution:

$$\Pr\left\{{}_k^h N(x,t)=n\right\}=\frac{\exp[-{}_k^h m(x,t)*h][{}_k^h m(x,t)*h]^n}{n!}$$

where ${}_k^h N(x,t)$ is the number of direct transitions experienced by individual k during the interval (x, x+h), provided k is alive at x.

The probability of a transition from state i to state j in n steps is $[{}_k\xi_{ij}(x,t)]^n$, which is the (i,j)-th element of the matrix of transition probabilities $[{}_k\boldsymbol{\xi}(x,t)]^n$ where n is the n-th power of ${}_k\boldsymbol{\xi}(x,t)$ which is the matrix of elements ${}_k\xi_{ij}(x,t)$ that denote the probability of a transition from i to j in a single step. These direct transition probabilities are constant in the interval (x,x+h). Hence (Çinlar, 1975, p. 236):

$$\Pr\left\{{}_kY(x+h,t+h)=j\,\middle|\,{}_kY(x,t)=i\right\}=\sum_{n=0}^{\infty}\frac{\exp[-{}_k^h m(x,t)*h][{}_k^h m(x,t)*h]^n}{n!}[{}_k\xi_{ij}(x,t)]^n$$

where the first term of the right-hand side describes a Poisson process and is the probability of n direct transitions during the interval of length h and the second term denotes a Markov chain and is the probability of a transition from i to j in n steps. In matrix terms:

$${}_k^h\mathbf{P}(x,t)=\sum_{n=0}^{\infty}\frac{\exp[-{}_k m(x,t)*h][{}_k m(x,t)*h]^n}{n!}\left[{}_k\boldsymbol{\xi}(x,t)\right]^n$$

For details on a Markov process as a Markov chain subordinated to a Poisson process, see Çinlar (1975). Note that the sequence of states visited by the process is independent of the sojourn times. During the interval (x,x+h), the sojourn times in the successive states are exponentially distributed with the same parameter ${}_km(x,t)$. The transition probabilities ${}_k^h\mathbf{P}(x,t)$ of the Markov process may also be expressed as a Markov chain subordinate to a Poisson process when the escape rates differ between states.

Çinlar (1975, p. 260) shows that the transition function **P**(t) of a Markov process can be expressed in terms of a Poisson process with rate c and a Markov chain with transition matrix **K** that is independent of t, provided the Poisson process and the Markov process are independent. The transition function is (for individual k, but k is omitted for convenience):

$$\mathbf{P}(t) = \sum_{n=0}^{\infty} \frac{\exp[-ct](ct)^n}{n!}\mathbf{K}^n$$

To characterise the Markov process, the generator $\boldsymbol{\mu}$ must be determined from c and **K**. Consider individual k in state i. The instantaneous rate of a transition from state i to state j is $\mu_{ij} = c\frac{K_{ij}}{1-K_{ii}}$, provided i is not an absorbing state. In the expression c is the escape rate from i, K_{ij} is the probability of moving to j in 1 step, and $1-K_{ii}$ is the probability of staying in state i for one step of the Markov chain. The sequence Y_0, Y_1, Y_2, … of successive states visited form a Markov chain with transition matrix **Q** with elements $Q_{ij} = \frac{K_{ij}}{1-K_{ii}}$ and $Q_{ii} = 0$. The probability that the Markov chain stays in state i for at least one step is K_{ii}. The probability that the chain stays in i for exactly n steps is $[K_{ii}]^{n-1}[1-K_{ii}]$. The sojourn time in i is the sum of n interarrival times with probability $[K_{ii}]^{n-1}[1-K_{ii}]$. That sojourn time has an exponential distribution with parameter μ_i. Hence μ_i is proportional to $1-K_{ii}$. The proportionality factor is c: $\mu_i= c\,[1-K_{ii}]$. The distribution of the sojourn times in state i is $\exp(-\mu_i t)$. The probability of moving from i to j at time t is $Q_{ij}\exp[-\mu_i t]$ where $\mu_i= c\,[1-K_{ii}]$. The relation between the transition intensity and the elements of the transition matrix **K** follows: $\mu_i = c[1-K_{ii}]$ and $\mu_{ij} = cK_{ij}$. In matrix terms:

$$\boldsymbol{\mu} = c\mathbf{I} - c\mathbf{K}$$

Hence $\mathbf{K} = \left[\mathbf{I} - \frac{1}{c}\boldsymbol{\mu}\right]$ where $\boldsymbol{\mu}$ is defined at the beginning of this section.

The separation of the transition intensity into two parts is particularly useful when the factors that affect the occurrence of an event differ from the factors that affect the type of event (direction of change or destination after the event). In that case the event occurrence and the direction of change are two distinct causal processes and ${}_k\mu_{i+}(x,t)$ and ${}_k\xi_{ij}(x,t)$ can be estimated independently (Hachen, 1988, p. 29; Sen and Smith, 1995, p. 372). The transition rate is studied using a transition rate model (or hazard model) whereas the destination probability is studied using a logit model, which is equivalent to a logistic regression model.

4.3. Estimation of transition rates from data

To estimation of ${}^{h}_{k}\mathbf{P}(x,t) = \exp\left[-\int_0^h {}_k\boldsymbol{\mu}(x+\tau, t+\tau)\,d\tau\right]$ from data requires assumptions. The common approach is to postulate a *piecewise constant intensity function* in the interval from x to x+h: ${}_k\boldsymbol{\mu}(x+\tau,t+\tau) = {}_k\boldsymbol{\mu}(x,t)$ for $(0 \le \tau < h)$. This implies an exponential distribution of time to event within each age interval of length h. The model that results is referred to as the exponential model. A second avenue is to postulates a *piecewise constant density function*, i.e. a piecewise linear survival function. A piecewise constant density function implies a piecewise linear survival function and is obtained when demographic events are uniformly distributed within the age intervals. The model that results is referred to as the linear model. In multistate modelling, the first avenue is followed by Hoem and Funck-Jensen (1982), Van Imhoff (1990) and Van Imhoff and Keilman (1991) among others; the second by Rogers (1975), Willekens and Drewe (1984) and Zeng Yi et al. (1997) among others. The exponential model is considered next. The linear model is discussed in Annex B.

In the exponential model, the transition intensities ${}_k\boldsymbol{\mu}(x,t)$ are assumed to remain constant during the age interval from x to x+h. The piecewise constant transition rates are denoted by ${}^{h}_{k}\mathbf{m}(x,t)$. Hence

$${}_k\boldsymbol{\mu}(x+\tau,t) = {}^{h}_{k}\mathbf{m}(x,t) \text{ for } 0 \le \tau < h.$$

The matrix of discrete-time transition probabilities during the interval from x and x+h, for individuals born at t-x, is

$${}^{h}_{k}\mathbf{P}(x,t) = \exp[-h\ {}^{h}_{k}\mathbf{m}(x,t)]$$

A number of methods exist to determine the value of $\exp\left[-{}^{h}_{k}\mathbf{m}(x,t)\right]$. Several methods are reviewed in Annex A. One method is the Taylor series expansion. Note that for matrix **A**, exp(**A**) may be written as a Taylor series expansion A

$$\exp(\mathbf{A}) = \mathbf{I} + \mathbf{A} + \frac{1}{2!}\mathbf{A}^2 + \frac{1}{3!}\mathbf{A}^3 + \ldots$$

Hence

$$\exp[-h\ {}^{h}_{k}\mathbf{m}(x,t)] = \mathbf{I} - \mathrm{h}\ {}^{h}_{k}\mathbf{m}(x,t) + \frac{\mathrm{h}^2}{2!}\left[{}^{h}_{k}\mathbf{m}(x,t)\right]^2 - \frac{\mathrm{h}^3}{3!}\left[{}^{h}_{k}\mathbf{m}(x,t)\right]^3 + \ldots = \sum_{\mathrm{p}=0}^{\infty} \frac{-[\mathrm{h}\ {}^{\mathrm{h}}_{\mathrm{k}}\mathbf{m}(x,t)]^p}{p!}$$

(see also Schoen, 1988, p. 72).

The transition rates are model rates, i.e. rates used in the demographic model (life table or projection model). The model rates are set equal to empirical occurrence-exposure rates for that age interval, and the empirical rates are estimated from available data. The distinction between model rates and empirical rates is common in demography and is made because empirical rates must often be adjusted (e.g. smoothed) before they can be used in cohort analysis (life tables) or period analysis. The assumption that model rates are equal to empirical rates is consistent with the general assumption in demography that life-table rates are equal to empirical rates (see e.g. Preston et al., 2001, p. 48).

The matrix of transition rates is ${}_k^h\mathbf{m}(x,t)$ and the matrix of empirical transition rates is denoted as ${}_k^h\mathbf{M}(x,t)$. An element ${}_k^hM_{ij}(x,t)$ of ${}_k^h\mathbf{M}(x,t)$ is equal to the ratio of the number of moves or direct transitions from i to j made during the interval from x to x+h by individual born at t-x, and the time spent in the state of origin i between x and x+h by individuals born at instantaneous time t-x:

$${}_k^hM_{ij}(x,t) = \frac{{}_k^hn_{ij}(x,t)}{{}_k^hPY_i\ (x,t)}$$

where ${}_k^hn_{ij}(x,t)$ is the observed number of moves from i to j during the interval by individual k and ${}_k^hPY_i\ (x,t)$ is the number of years (or other unit of time) individual k stays in state i between the ages x and x+h. It is the duration of exposure to the risk of leaving state i. Exposure is measured in person-months or person-years (PY). In case of two states, the rate equation may be written as follows:

$$\begin{bmatrix} {}_k^hM_{11}(x,\mathrm{t}) & -{}_k^hM_{21}(x,\mathrm{t}) \\ -{}_k^hM_{12}(x,\mathrm{t}) & {}_k^hM_{22}(x,\mathrm{t}) \end{bmatrix} = \begin{bmatrix} {}_k^hn_{11}(x,\mathrm{t}) & -{}_k^hn_{21}(x,\mathrm{t}) \\ -{}_k^hn_{12}(x,\mathrm{t}) & {}_k^hn_{22}(x,\mathrm{t}) \end{bmatrix} \begin{bmatrix} {}_k^hPY_1\ (x,t) & 0 \\ 0 & {}_k^hPY_2\ (x,t) \end{bmatrix}^{-1}$$

where ${}_k^hM_{11}(x,t) = {}_k^hM_{12}(x,t)$, ${}_k^hM_{22}(x,t) = {}_k^hM_{21}(x,t)$, ${}_k^hn_{11}(x,t) = {}_k^hn_{12}(x,t)$ and ${}_k^hn_{11}(x,t) = {}_k^hn_{12}(x,t)$. In matrix notation: ${}_k^h\mathbf{M}(x,\mathrm{t}) = {}_k^h\mathbf{n}(x,\mathrm{t})\left[diag\ {}_k^h\overline{\mathbf{PY}}(x,\mathrm{t})\right]^{-1}$

4.4. The exposure function: sojourn times

An important function of the transition probabilities is the exposure function that measures the expected sojourn time in the different states. Sojourn times may be determined for an entire lifespan or for the interval between two ages or the interval between two points in time. For instance, we may be interested in the number of years individual k may expect to be employed between ages 50 and 65. That figure depends on the age of k when the question is asked. The expected number of years in employment between 50 and 65 is lower at k's birth than at his 50th birthday because of the (small) probability that k does not reach age 50. The expected sojourn time in employment may also be conditioned on the state k occupies at birth, at age 50 or at any other age between 0 and 50. For instance, the sojourn time in employment between 50 and 65 is likely to be larger if k is employed at 30 than when he is not in the labour force at that age. Biographic indicators that condition on the state occupied are status-based measures. Indicators that do not condition on the state occupied are population-based measures

(Willekens, 1987). In this section, the age (or point of time) that serves as a reference is the reference age. The position of the observer may be at a different age. For instance, the sojourn time in employment between 50 and 65 by employment status at age 30 may be estimated at k's birth. In this case, 30 is the reference age and 0 is the age at estimation or the age at which the observer is positioned. The age at estimation may be 30 or at any age between 0 and 30. To determine the expected sojourn time, the state occupied at the reference age may be taken into account (status-based demographic indicator). If the reference age differs from the age at the start of the interval for which the sojourn time is measured, then the reference age and the state occupied at the reference age (if applicable) are shown as superscripts at the right. Age at estimation and, if applicable, the state occupied at that age, are shown as left subscripts. For example, ${}_{k\,b}^{\;h}L_i^{a,r}(x,t_b+x)$ denotes the number of years individual k, who is born at t_b, may expect to spend in state i between ages x and x+h, provided k is in state r at exact age a. The sojourn time is estimated at age b of k (for example at the age at entry into the labour market). If the age at estimation is omitted, it is the same as the reference age. If both the reference age and the age at estimation are omitted, they are 0 (age at birth).

Since the transition intensities fully determine the life path, they also determine the expected durations of stay in the different episodes of life. By way of introduction we first consider a single event (e.g. death) and two states (e.g. alive and dead). Multiple states are considered next. In case of two states the expected sojourn time is the expected waiting time to the event (e.g. death). We consider the case of two states first. In conventional survival analysis with a single event (death), the exposure function is obtained by integrating the survival function. If ${}_k\ell(x,t_b+x)$ is the probability that individual k, who is born at t_b, is alive at instantaneous age x and time t = t_b+x, then the expected lifetime is ${}_k^{\infty}L(0,t_b)=\int_0^{\infty}{}_k\ell(\tau,t_b+\tau)\,d\tau$ where ∞ is the highest age (and may be written as ω). The expected lifetime or life expectancy is the area under the survival function. The expected numbers of years spent alive during the interval (x,x+h) is the area under the survival function from x to x+h. It is ${}_k^hL(x,t)=\int_0^h{}_k\ell(x+\tau,t+\tau)\,d\tau$. It is the number of years k may expect to live between ages x and x+h. The number of years is measured at birth. It may also be written as ${}_{k\,0}^{\;h}L(x,t)$. If the individual is alive at x, the expected number of years is ${}_{k\,x}^{\;h}L(x,t)={}_k^hL(x,t)/{}_k\ell(x,t)$ where ${}_k\ell(x,t)$ is the probability that k survives to x. It is equal to ${}_k\ell(x,t)=\exp\left[-\int_0^x{}_k\mu(\tau,t_b+\tau)\,d\tau\right]$. The expected number of years a newborn child (k) may expect to live between ages x and x+h, provided the newborn reaches age x, is the product of a survival probability and a sojourn time conditional on surviving: ${}_{k\,0}^{\;h}L(x,t)={}_{k\,x}^{\;h}L(x,t)\;{}_k\ell(x,t)$. If the death rate is constant during the interval (x,x+h) then

$${}_{k\,x}^{\;h}L(x,t)=\frac{1}{{}_k\ell(x,t)}\int_0^h{}_k\ell(x+\tau,t+\tau)\,d\tau=\int_0^h{}_k^{\tau}p(x,t)\,d\tau=\int_0^h\exp\left[-\tau\;{}_k^hm(x,t)\right]d\tau$$

which is equal to $\left| -\frac{1}{{}_k^h m(x,t)} \exp\left[-\tau \, {}_k^h m(x,t)\right] \right|_0^h = \frac{1-\exp[-h \, {}_k^h m(x,t)]}{{}_k^h m(x,t)}$

Note that the numerator is the probability of an event (death) between x and x+h, provided the event has not yet occurred at age x. Hence the sojourn time is the ratio of the probability of the event over the event rate: ${}_{k\,x}^{\,h}L(x,t) = \frac{1-{}_k^h p(x,t)}{{}_k^h m(x,t)}$. The life expectancy at x is the expected number of years beyond x: ${}_{k\,x}^{\,\infty}L(x,t) = \frac{1}{{}_k\ell(x,t)} \int_0^\infty {}_k\ell(\tau, t_b + \tau)\, d\tau$. If the death rate is piecewise constant, then

$${}_{k\,x}^{\,\infty}L(x,t) = \frac{1}{{}_k\ell(x,t)} \sum_{y=x}^{\infty} \left[\frac{1-\exp[-h \, {}_k^h m(y, t_b + y)]}{{}_k^h m(y, t_b + y)} \, {}_k\ell(y, t_b + y) \right]$$

At the highest, open-ended age group, the probability of surviving is 0 and the number of years individual k who is currently x years, may expect to live beyond that age (z, say), is ${}_{k\,x}^{\,\infty}L(z, t_b + z) = \frac{{}_k\ell(z, t_b + z)}{{}_k\ell(x,t)} \frac{1}{{}_k^h m(z, t_b + z)}$ where ${}_k^h m(z, t_b + z)$ is the death rate of the highest age group.

The waiting time to the event (death) is not conditional on the event occurring during the interval. If the event does not occur individual k remains at risk for the entire interval, i.e. for h time units. The probability that k remains at risk for the entire interval is $1-{}_k^h p(x,t)$. If, during the interval, the event occurs, then the exposure time is shorter. Let ${}_k^h a(x,t)$ denote the ***fraction of the interval*** (x,x+h), individual k is exposed in case *the event occurs* during the interval. The conditional measure may be estimated from the unconditional waiting time to the event:

$${}_k^h L(x,t) = h \, {}_k^h p(x,t) + h \, {}_k^h a(x,t) \left[1 - {}_k^h p(x,t)\right] = \frac{1 - {}_k^h p(x,t)}{{}_k^h m(x,t)}$$

which gives

$${}_k^h a(x,t) = 1 - \frac{1}{1-{}_k^h p(x,t)} + \frac{1}{h \, {}_k^h m(x,t)} = 1 - \frac{1}{1-\exp[-h \, {}_k^h m(x,t)]} + \frac{1}{h \, {}_k^h m(x,t)}$$

Hence ${}_k^h a(x,t)$ is determined by both the conditional event probability $1-{}_k^h p(x,t)$ and the event rate ${}_k^h m(x,t)$. The time to event for a process that ends in an event during the interval (x,x+h) is h ${}_k^h a(x,t)$. The measure ${}_k^h a(x,t)$ is generally known as Chiang's "a". In empirical studies of mortality, Chiang (1960) found that ${}_k^h a(x,t)$ is more or less invariant with respect to sex, race, geographic location, other demographic variables, and cause of death. He suggested that, if ${}_k^h a(x,t)$ is determined for each age group in one population, it could be used for many populations. Chiang called ${}_k^h a(x,t)$ the *fraction of the last year of life* (h = 1) (Chiang, 1968, pp 190ff; 1984, pp. 142ff). Note that ${}_k^h a(x,t) = \frac{h}{2}$ if the events are uniformly distributed during the interval. If case of a constant event rate, the waiting time to the event is exponentially distributed. Consequently, the majority of events occur during the first half of the interval. The probability that the event occurs in the first half of the interval of length h, provided the event occurs during the interval, is $\dfrac{1-\exp[-\frac{h}{2}\ {}_k^h m(x,t)]}{1-\exp[-h\ {}_k^h m(x,t)]}$.

In multistate analysis, the exposure function is obtained by integrating the function of state probabilities, which is the multistate equivalent of the survival function. The time individual k may expect to spend in state i between ages x and x+h is ${}_k^h\overline{L}_i(x,t)=\int_0^h {}_k\ell_i(x+\tau,t+\tau)\,d\tau$ where ${}_k\ell_i(x+\tau,t+\tau)$ is the (unconditional) probability that k is in state i at age x+τ and time t+τ. The expected sojourn times in the different states during the interval (x,x+h) or (t,t+h) is ${}_k^h\overline{\mathbf{L}}(x,t)=\int_0^h {}_k\mathbf{k}(x+\tau,t+\tau)\,d\tau$, where ${}_k\mathbf{k}(x,t)$ has been defined above. The expected sojourn times in the different states between birth and death is ${}_k^\infty\overline{\mathbf{L}}(0,t_b)=\int_0^\infty {}_k\mathbf{k}(\tau,t_b+\tau)\,d\tau$ where t_b is the date of birth of individual k.

Exposure functions differ in the degree of conditionality imposed. The expected sojourn time in a given state during the interval (x, x+h) may be conditioned on surviving at x or at another age (reference age) and the state occupied at that age. The expected sojourn times in the different states beyond age x, provided k survives to age x is ${}_{k\,x}^{\infty}\overline{\mathbf{L}}(x,t)=\dfrac{1}{{}_k\ell_+(x,t)}\left[\int_x^\infty {}_k\mathbf{k}(\tau,t_b+\tau)\,d\tau\right]$. It is the number of years spent in the different states during the expected lifetime beyond age x, provided k reaches age x.

The expected sojourn time may also be conditioned on the state occupancy at current age, a previous age or at birth. Let ${}_{k\,ix}^{\;\;h}L_j(x,t)$ denote the number of years individual k, who is born at t_b = t-x and occupies state i at age x, may expect to spend in state j between ages x and x+h. The value does not depend on the state occupied at birth. The sojourn times in each state between x and x+h by state occupied at x are assembled in the matrix ${}_{k\,x}^{h}\mathbf{L}(x,t)$. Note that the matrix has the sojourn times by state occupied at exact age x. The sojourn times in each state by state occupied at birth is ${}_{k\,0}^{h}\mathbf{L}(x,t)$. An element ${}_{k\,i0}^{\;\;h}L_j(x,t)$ of ${}_{k\,x}^{h}\mathbf{L}(x,t)$ denotes the number of years k, who is born at t_b = t-x and occupies state i *at birth*, may expect to spend in state j between ages x and x+h. It is equal to ${}_{k\,x}^{h}\mathbf{L}(x,t)=\int_0^h {}_k^\tau\mathbf{P}(x,t)\ d\tau$ where ${}_k^\tau\mathbf{P}(x,t)$ is the matrix of probabilities of state occupancies at age x+τ by state occupied at x.

Since ${}_{k\,0}^{h}\mathbf{L}(x,t)={}_{k\,x}^{h}\mathbf{L}(x,t)\ {}_k^x\mathbf{P}(0,t_b)$, with ${}_k^x\mathbf{P}(0,t_b)$ the matrix indicating the probabilities of state occupancies at age x by state occupied at birth, the expected sojourn times conditional on state occupancy at age x may be obtained as ${}_{k\,x}^{h}\mathbf{L}(x,t)={}_{k\,0}^{h}\mathbf{L}(x,t)\left[{}_k^x\mathbf{P}(0,t_b)\right]^{-1}$ provided the inverse exists. This expression was first developed in multiregional demography to obtain the expected duration of residence in each region beyond age x by place of residence at age x (see Rogers, 1975; Willekens and Rogers, 1978, p. 36, Rogers and Willekens, 1986, p. 375). The matrix ${}^{h}_{k0}\mathbf{L}(x,t)$ contains measures of sojourn time that are conditional on the state occupied at birth whereas ${}^{h}_{kx}\mathbf{L}(x,t)$ contains measures that are conditional on the state occupied at age exact age x.

The sojourn time ${}^{h}_{k}\overline{\mathbf{L}}(x,t)$ is related to the transition intensities:

$$
\begin{aligned}
{}^{h}_{k}\overline{\mathbf{L}}(x,t) &= {}^{h}_{kx}\mathbf{L}(x,t)\ {}_{k}\mathbf{k}(x,t) = \left[\int_0^h {}^{\tau}_{k}\mathbf{P}(x,t)\ d\tau\right] {}_{k}\mathbf{k}(x,t) \\
&= \left[\int_0^h \exp\left[-\int_x^{x+\tau} {}_{k}\boldsymbol{\mu}(\xi, t_b+\xi)\, d\xi\right] d\tau\right] \exp\left[-\int_0^x {}_{k}\boldsymbol{\mu}(\tau, t_b+\tau)\, d\tau\right] {}_{k}\mathbf{k}(0,t_b)
\end{aligned}
$$

The quantity ${}^{h}_{k}\overline{\mathbf{L}}(x,t)$ is the expected sojourn time in each state during the (x,x+h)-interval irrespective of the survival status and state occupied at x and ${}^{h}_{kx}\mathbf{L}(x,t)$ is expected sojourn time in each state during the (x,x+h)-interval by state occupied at x. An element ${}^{h}_{kx}L_{ij}(x,t)$ of ${}^{h}_{kx}\mathbf{L}(x,t)$ is the number of years individual k who is born at $t_b = t - x$ and is aged x at time t and occupies state i at that age, may expect to spend in state j between ages x and x+h if the transition intensities apply. It is ${}^{h}_{kx}L_{ij}(x,t) = \int_0^h {}^{\tau}_{k}p_{ij}(x,t)\, d\tau$. The sojourn time in j during the interval from x+τ to x+τ+dτ is the product of the probability that individual k (who is in i at x) is in state j at τ ($0 \le \tau < h$) and the length of the interval. It is ${}^{\tau}_{k}p_{ij}(x,t)\, d\tau$. The exposure function measures the area under and between curves of state occupancies. Note that ${}^{h}_{k}\overline{\mathbf{L}}(x,t) = {}^{h}_{kx}\mathbf{L}(x,t)\ {}_{k}\mathbf{k}(x,t)$ The sojourn time ${}^{h}_{kx}\mathbf{L}(x,t)$ is conditional on (1) being alive at x and (2) occupying state i at x.

Exposure functions differ in the degree of conditionality imposed. To relax the conditionality, state probabilities are introduced. The time spent in state j between exact ages x and x+h by k, conditional on being alive at x but irrespective of the state occupied at x, is the product of ${}^{h}_{k}L_{ij}(x,t)$ and the state probability at x, ${}_{k}\pi_i(x,t)$, summed over all possible states at age x:

$$
{}^{h}_{k}\overline{\overline{\mathbf{L}}}(x,t) = \begin{bmatrix} {}^{h}_{k}\overline{\overline{L}}_1(x,t) \\ {}^{h}_{k}\overline{\overline{L}}_2(x,t) \\ . \\ . \\ {}^{h}_{k}\overline{\overline{L}}_I(x,t) \end{bmatrix} = \begin{bmatrix} {}^{h}_{k}L_{11}(x,t) & {}^{h}_{k}L_{21}(x,t) & . & . & {}^{h}_{k}L_{I1}(x,t) \\ {}^{h}_{k}L_{12}(x,t) & {}^{h}_{k}L_{22}(x,t) & . & . & {}^{h}_{k}L_{I2}(x,t) \\ . & . & . & . & . \\ . & . & . & . & . \\ {}^{h}_{k}L_{1I}(x,t) & {}^{h}_{k}L_{2I}(x,t) & . & . & {}^{h}_{k}L_{II}(x,t) \end{bmatrix} \begin{bmatrix} {}_{k}\pi_1(x,t) \\ {}_{k}\pi_2(x,t) \\ . \\ . \\ {}_{k}\pi_I(x,t) \end{bmatrix}
$$

The expected sojourn times in the different states between ages x and x+h, irrespective of whether or not individual k is alive at x and irrespective of the state occupied at x, is obtained by replacing the conditional state probabilities $_k\pi_i(x,t)$ by the unconditional state probabilities $_k\ell_i(x,t)$. Then ${}^h_k\overline{\mathbf{L}}(x,t) = {}^h_{k\,x}\mathbf{L}(x,t)\ {}_k\mathbf{k}(x,t)$. Note that ${}_k\mathbf{k}(x,t) = {}^x_k\mathbf{P}(0,t_b)\ {}_k\mathbf{k}(0,t_b)$ with ${}^x_k\mathbf{P}(0,t_b)$ the matrix of transition probabilities between birth and age x, and ${}_k\mathbf{k}(0,t_b)$ the vector of state probabilities at birth. An element ${}^x_k p_{ij}(0,t_b)$ of ${}^x_k\mathbf{P}(0,t_b)$ is the probability that individual k born in state i at t_b occupies state j at exact age x (and $t = t_b+x$). An element $_k\ell_i(0,t_b)$ of ${}_k\mathbf{k}(0,t_b)$ is the probability than a child born at t-x occupies state i at birth.

In case of piecewise constant transition intensities, ${}_k\boldsymbol{\mu}(x+\tau,t+\tau) = {}_k\boldsymbol{\mu}(x,t) = {}^h_k\mathbf{m}(x,t)$ for $(0 \leq \tau < h)$ and

$${}^h_k\mathbf{L}(x,t) = \int_0^h {}^\tau_k\mathbf{P}(x,t)\,d\tau = \int_0^h \exp\left[-\tau\ {}^h_k\mathbf{m}(x,t)\right]d\tau$$

Integration yields

$$-\left[{}^h_k\mathbf{m}(x,t)\right]^{-1} \left|\exp\left[-\tau\ {}^h_k\mathbf{m}(x,t)\right]\right|_0^h$$ which is equal to

$$\left[{}^h_k\mathbf{m}(x,t)\right]^{-1}\left[\mathbf{I} - \exp\left[-h\ {}^h_k\mathbf{m}(x,t)\right]\right]$$

Hence the expected sojourn times in the various states during the (x,x+h)-interval, by state occupied at x, are given by:

$${}^h_k\mathbf{L}(x,t) = \left[{}^h_k\mathbf{m}(x,t)\right]^{-1}\left[\mathbf{I} - \exp[-h\ {}^h_k\mathbf{m}(x,t)]\right]$$

where ${}^h_k\mathbf{L}(x,t)$ is a matrix of expected sojourn times in the different states during the interval from x to x+h by individual k born at t_b and currently of exact age x, by state occupied at x. The expression is also shown by Namboodiri and Suchindran (1987, p. 145), Schoen (1988, p. 101) and van Imhoff (1990). The exponential of the matrix of transition rates ${}^h_k\mathbf{m}(x,t)$ can be obtained by the Taylor series expansion (Annex A). Combining the above equation and the infinite power series leads to

$${}^h_k\mathbf{L}(x,t) = h\left[\mathbf{I} - \tfrac{1}{2!}h\mathbf{M} + \tfrac{1}{3!}(h\mathbf{M})^2 - \tfrac{1}{4!}(h\mathbf{M})^3 + \ldots\right]$$

The expected sojourn times spent in the different states between x and x+h, irrespective of survival and state occupancy at x, is

$${}^h_k\overline{\mathbf{L}}(x,t) = {}^h_k\mathbf{L}(x,t)\ {}_k\mathbf{k}(x,t) = \left[{}^h_k\mathbf{m}(x,t)\right]^{-1}\left[\mathbf{I} - \exp[-h\ {}^h_k\mathbf{m}(x,t)]\right]{}_k\mathbf{k}(x,t)$$

The equation gives the person-years (or person-months) individual k spends in the various states during the age interval from x to x+h. The sojourn time depends on the state individual k is occupying on his x-th birthday (at time t). The equation is similar to the equation derived by Van Imhoff (1990) and Van Imhoff and Keilman (1991, p. 27), but it applies to a closed population, i.e. a population in the absence of transitions from and to outside of the population system (external events). Open populations are considered in the next section.

The number of years individual k, who is born at t_b, may expect to spend in the different states during the lifetime is ${}^{\omega}_{k}\overline{\mathbf{L}}(0,t_b)$ where ω is the highest age,

$$
\begin{aligned}
{}^{\omega}_{k}\overline{\mathbf{L}}(0,t_b) &= \sum_{x=0}^{\omega} {}^{h}_{k\,x}\mathbf{L}(x,t_b+x)\; {}_{k}\mathbf{k}(x,t_b+x) \\
&= \sum_{x=0}^{\omega} \left[{}^{h}_{k}\mathbf{m}(x,t_b+x)\right]^{-1} \left[\mathbf{I} - \exp[-h\; {}^{h}_{k}\mathbf{m}(x,t_b+x)]\right] {}_{k}\mathbf{k}(x,t_b+x)
\end{aligned}
$$

The total life expectancy is $\mathbf{1}'\, {}^{\omega}_{k}\overline{\mathbf{L}}(0,t_b)$ where $\mathbf{1}$ is a vector of ones and ` denotes transpose. The proportion of the life span spent in the different states is ${}^{\omega}_{k}\overline{\mathbf{L}}(0,t_b) / \left[\mathbf{1}'\, {}^{\omega}_{k}\overline{\mathbf{L}}(0,t_b)\right]$. For instance the proportion spent in state i is ${}_{k}\theta_i = {}^{\omega}_{k}\overline{L}_i(0,t_b) / {}^{\omega}_{k}\overline{L}_+(0,t_b)$. The multistate demography, the indicator ${}_{k}\theta_i$ is known as the mobility level (Willekens and Rogers, 1978, p. 99).

The expected sojourn time ${}^{h}_{k\,x}\mathbf{L}(x,t)$ consists of two parts. The first is the time spent in the state occupied at exact age x if individual k does not leave that state during the interval. The second part is the expected sojourn times in the different states conditional on leaving the origin state. The probability that individual k who occupies state i at x stays in i for the entire interval (x,x+h) is ${}^{h}_{k}p_i^{cont}(x,t) = \exp\left[-\int_0^h {}_{k}\mu_{i+}(x+\tau,t+\tau)\,d\tau\right]$. If the transition intensities are constant during the (x,x+h)-interval then ${}^{h}_{k}p_i^{cont}(x,t) = \exp\left[-h\; {}^{h}_{k}m_{i+}(x,t)\right]$. The expected sojourn times in the different states between x and x+h, provides k leaves the state occupied at x is ${}^{h}_{k}\mathbf{a}(x,t) = {}^{h}_{k}\mathbf{L}(x,t) - h\, diag\left[{}^{h}_{k}\mathbf{P}^{cont}(x,t)\right]$ where ${}^{h}_{k}\mathbf{P}^{cont}(x,t)$ is a diagonal matrix with elements ${}^{h}_{k}p_i^{cont}(x,t)$.

The expected sojourn time in i *on a continuous basis*, i.e. if during the interval a return to i is not considered, is ${}^{h}_{k}L_i^{cont}(x,t) = \dfrac{1 - {}^{h}_{k}p_i^{cont}(x,t)}{{}^{h}_{k}m_{i+}(x,t)} = \dfrac{1-\exp[-h\, {}^{h}_{k}m_{i+}(x,t)]}{{}^{h}_{k}m_{i+}(x,t)}$. It consists of two parts: the sojourn time in i if individual k stays in i for the entire interval (= duration h) and the waiting time to leaving state i, provided k moves on to another state during the interval.

4.5. The flow equation and direct transitions

Multistate analysis is about transitions individuals make in life. Transitions may be recorded in continuous time or discrete time. Transitions are registered in continuous time if every change in attribute, i.e. every move between functional states, are recorded. Transitions that are recorded in discrete time measure transitions by comparing state occupancies at two points in time. Transitions in continuous time are referred to as *direct* *transitions* to distinguish them from discrete-time transitions. The discrete-time transitions individual k may expect to make during the interval (x,x+h) are predicted by the discrete-time transition probabilities ${}^h_k\mathbf{P}(x,t)$. Direct transitions depend on the rates of transition and the expected duration at risk estimated by the exposure function. The exposure function is an important determinant of the number of direct transitions individual k makes during the age interval from x to x+h.

Recall that ${}^h_kY_{ij}(x,t)$ denotes an indicator variable which takes on the value of 1 if individual k occupies state i at age x and time t and state j at x+h and t+h. It is zero otherwise. The indicator variable ${}_kY_{ij}(x,t)$ takes on the value of 1 if individual k makes a direct transition from i to j at age x and time t. Consider the interval (x,x+h). If individual k experiences a discrete-time transition, i.e. if he is in i at x and in j at x+h, then ${}^h_kY_{ij}(x,t)=1$. The probability of a transition is the expected value ${}_k\ell_{ij}(x,t)=E\lfloor {}_kY_{ij}(x,t)\rfloor$. The expected value of ${}_kY_{ij}(x,t)$, provided k is in state i at exact age x, is the transition intensity or instantaneous rate of transition: $E\lfloor {}_kY_{j|i}(x,t)\rfloor = {}_k\mu_{ij}(x,t)$. The number of direct (i,j)-transitions individual k experiences during the interval is $\int_0^h {}_kY_{ij}(x+\tau,t+\tau)\,d\tau$. The expected number of direct transitions between i and j during the interval (x,x+h) is ${}^h_kn_{ij}(x,t)=\int_0^h {}_k\ell_{ij}(x+\tau,t+\tau)\,d\tau=\int_0^h {}_k\mu_{ij}(x+\tau,t+\tau)\ {}_k\ell_i(x+\tau,t+\tau)\,d\tau$. The matrix of numbers of direct transitions during the interval (x,x+h) is

$${}^h_k\mathbf{n}(x,t)=\int_0^h {}_k\boldsymbol{\mu}(x+\tau,t+\tau)[diag\,{}_k\mathbf{k}(x+\tau,t+\tau)]\,d\tau$$

where $diag\,{}_k\mathbf{k}(x+\tau,t+\tau)$ is a matrix with the elements ${}_k\ell_i(x,t)$ in the diagonal. The number of times individual k leaves the different states during the interval is $\mathbf{1}'\,{}^h_k\mathbf{n}(x,t)$ and the number of times he enters the different states is ${}^h_k\mathbf{n}(x,t)\mathbf{1}$ where $\mathbf{1}$ is a column vector of ones and $\mathbf{1}'$ is a row vector of ones.

The expected number of direct transitions k makes during the interval (x,x+h) are used in the flow equation:

$${}_k\mathbf{k}(x+h,t+h)={}_k\mathbf{k}(x,t)-\int_0^h {}_k\boldsymbol{\mu}(x+\tau,t+\tau)\ {}_k\mathbf{k}(x+\tau,t+\tau)\,d\tau={}_k\mathbf{k}(x,t)-\mathbf{1}'\,{}^h_k\mathbf{n}(x,t)+{}^h_k\mathbf{n}(x,t)\mathbf{1}$$

The probability that individual k occupies state j at age x+h, irrespective of the state occupied at x, is the probability that he occupies state j at age x, minus the expected number of times individual k leaves state j during the interval, plus the expected number of entries into state j during the interval. It is the flow equation

$${}_k\ell_j(x+h,t+h)={}_k\ell_j(x,t)-\sum_{r\neq j}{}^h_kn_{jr}(x,t)+\sum_{r\neq j}{}^h_kn_{rj}(x,t)$$

which may be written as

$$_k\ell_j(x+h,t+h) = {}_k\ell_j(x,t) - \sum_{r\neq j}\int_0^h {}_k\mu_{jr}(x+\tau,t+\tau)\ {}_k\ell_j(x+\tau,t+\tau)\,d\tau$$
$$+\sum_{r\neq j}\int_0^h {}_k\mu_{rj}(x+\tau,t+\tau)\ {}_k\ell_r(x+\tau,t+\tau)\,d\tau$$

An alternative formulation is
$$_k\mathbf{k}(x+h,t+h) = {}_k\mathbf{k}(x,t) - \left[\int_0^h {}_k\boldsymbol{\mu}(x+\tau,t+\tau)\ {}_k^{\tau}\mathbf{P}(x,t)\,d\tau\right] {}_k\mathbf{k}(x,t)$$
$$= {}_k\mathbf{k}(x,t) - \left[\int_0^h {}_k\boldsymbol{\mu}(x+\tau,t+\tau)\exp[-\int_0^{\tau} {}_k\boldsymbol{\mu}(x+\xi,t+\xi)\,d\xi]\,d\tau\right] {}_k\mathbf{k}(x,t)$$
The term in brackets is the matrix of flows $_{k\,x}^{\ h}\mathbf{n}(x,t)$ with elements $_{k\,x}^{\ h}n_{ij}(x,t)$ the number of direct transitions from i to j individual k, who is born at t-x, during the interval (x,x+h), by state occupied at x. Hence
$$_k\mathbf{k}(x+h,t+h) = {}_k\mathbf{k}(x,t) - {}_k^h\mathbf{m}(x,t)\int_0^h {}_k\mathbf{k}(x+\tau,t+\tau)\,d\tau = \left[\mathbf{I} - {}_k^h\mathbf{m}(x,t)\ {}_{k\,x}^{\ h}\mathbf{L}(x,t)\right] {}_k\mathbf{k}(x,t)$$
This flow equation, in a somewhat different form, is used by Schoen (1988, p. 68 and p. 70) as a starting point to derive the multistate life table.

Since $_k\mathbf{k}(x+h,t+h) = {}_k^h\mathbf{P}(x,t)\ {}_k\mathbf{k}(x,t)$, the following system of integral equations arises:
$$_k^h\mathbf{P}(x,t) = \mathbf{I} - \int_0^h {}_k\boldsymbol{\mu}(x+\tau,t+\tau)\ {}_k^{\tau}\mathbf{P}(x,t)\,d\tau$$
where $_k\boldsymbol{\mu}(x+\tau,t+\tau)$ is the matrix of transition intensities at age x+τ (and time t+τ) for individual k and $_k^{\tau}\mathbf{P}(x,t)$ [=$\mathbf{P}$(x,x+τ,t)] is the matrix of discrete-time transition probabilities during the (x,x+τ)- interval. The integral equation is a flow equation. An element $_k^{\tau}p_{ij}(x,t)$ of $_k^{\tau}\mathbf{P}(x,t)$ is the probability that individual k, who is born at t-x and occupies state i at exact age x, is in state j at exact age x+τ. The probability that individual k is in state j at x+h depends on the entries in j and exits from i:
$$_k^h p_{ij}(x,t) = \sum_{r\neq j}\int_0^h {}_k\mu_{rj}(x+\tau,t+\tau)\ {}_k^{\tau}p_{ir}(x,t)\,d\tau - \int_0^h {}_k\mu_{jr}(x+\tau,t+\tau)\ {}_k^{\tau}p_{ij}(x+\tau,t+\tau)\,d\tau$$

where $_k\mu_{jj}(x+\tau,t+\tau) = \sum_{r\neq j} {}_k\mu_{jr}(x+\tau,t+\tau)$

The term $\int_0^h {}_k\mu_{rj}(x+\tau,t+\tau)\ {}_k^{\tau}p_{ir}(x,t)\,d\tau = {}_k^h n_{irj}(x,t)$ is the total number of direct transitions from r to j experienced during the interval (x,x+h) and (t,t+h) by individual k who is born at t-x and occupies i at x and t. Hence the flow equation is:

$$_k^h p_{ij}(x,t) = \sum_{r\neq j} {}_k^h n_{irj}(x,t) - \sum_{r\neq j} {}_k^h n_{ijr}(x,t)$$
The first term on the right-hand side is the number of moves from t to j by individual k who is in state i at exact age x. The second term is the number of time individual k, who is in i at x, moves from j to r during the interval (x,x+h).

To derive an expression involving transition rates, we write

$$ {}_k^h\mathbf{P}(x,t) = \mathbf{I} - \left[\int_0^h {}_k\boldsymbol{\mu}(x+\tau,t+\tau)\ {}_k^\tau\mathbf{P}(x,t)\,d\tau\right]\left[\int_0^h {}_k^\tau\mathbf{P}(x,t)\,d\tau\right]^{-1}\left[\int_0^h {}_k^\tau\mathbf{P}(x,t)\,d\tau\right] $$

$$ {}_k^h\mathbf{P}(x,t) = \mathbf{I} - {}_k^h\mathbf{m}(x,t)\ {}_k^h\mathbf{L}(x,t) \qquad (4.2) $$

where ${}_k^h\mathbf{m}(x,t) = \left[\int_0^h {}_k\boldsymbol{\mu}(x+\tau,t+\tau)\ {}_k^\tau\mathbf{P}(x,t)\,d\tau\right]\left[\int_0^h {}_k^\tau\mathbf{P}(x,t)\,d\tau\right]^{-1}$ is the matrix, with elements ${}_k^h m_{ij}(x,t)$, of average transition rates during the interval from x to x+h for individual k born at instantaneous time t-x (age-cohort observation interval). To solve the system of equations (4.2) the exposure function ${}_k^h\mathbf{L}(x,t) = \int_0^h {}_k^\tau\mathbf{P}(x,t)\,d\tau$ must be determined.

If the intensities are constant during the interval (x,x+h), then ${}_k^h\mathbf{L}(x,t) = \left[{}_k^h\mathbf{m}(x,t)\right]^{-1}\left[\mathbf{I} - \exp[-h\ {}_k^h\mathbf{m}(x,t)\right]$ and the flow equation leads to the same expression of ${}_k^h\mathbf{P}(x,t)$ in terms of ${}_k^h\mathbf{m}(x,t)$: ${}_k^h\mathbf{P}(x,t) = \mathbf{I} - {}_k^h\mathbf{m}(x,t)\ {}_k^h\mathbf{L}(x,t)$. The linear model is shown in Annex B.

4.6. Biographic indicators

The transition probabilities, the sojourn times and the numbers of transitions in discrete time or continuous time are the basis for several useful biographic indicators. Since the probability measures on which the indicators are based are derived unambiguously from the transition intensities, all biographic indicators can be expressed in terms of the transition intensities and changes in values of the indicators can be traced to changes in values of the transition intensities. The values of the indicators are expected values that are consistent with a given set of transition intensities. If the transition intensities are estimated from data and the estimation method accounts for censoring (such as the Kaplan-Meier method, the Nelson-Aalen method and the life-table method) then the values of the biographic indicators account for censoring. In case of parameterization of the age profile and/or time dependence of the transition intensities, the biographic indicators are consistent with the parametric model used. If the transition intensities at a given age and time are related to covariates at that age and background variables at the calendar time considered, the biographic indicators are specific for the values of the covariates and the background variables.

Consider individual k born at t_b = t-x. The states occupied by k are denotes by i, j, r and s. The subscript i generally refers to the state of origin and j to the state of destination. The indices r and s refer to intermediate states, with r generally the state of origin and s the state of destination. For instance, we may be interested in the probability that k, who is born in state i, occupies state j at age x_2, provided he is in state r at an earlier age x_1 (x_1 <

x_2). We may also wonder how many time k moves from state r to state s between ages x_2 and x_3, provided he occupies state i at age x_1 ($x_1 \le x_2 < x_3$).

A distinction is made between unconditional indicators and conditional indicators. Unconditional indicators are indicators that are not conditioned on survival and/or the state occupied at a previous age. The position of the observer is at birth. Conditional indicators are obtained when the position of the observer is at a different age.

4.6.1. Probabilities

1. The probability that k occupies state j at age x: ${}_k\ell_j(x,t)$ where $t = t_b+x$.

It is the state probability ${}_k\ell_j(x,t) = \Pr\{{}_kY(x,t) = j\} = \Pr\{{}_kY_j(x,t) = 1\}$. Since the state occupied at x depends on the state occupied at birth and the transitions made between birth and age x, the state probability is given the matrix expression

${}_k\mathbf{k}(x,t) = \exp\left[-\int_0^x {}_k\boldsymbol{\mu}(\xi, t_b+\xi)\,d\xi\right] {}_k\mathbf{k}(0,t_b)$ where ${}_k\mathbf{k}(x,t)$ is the state vector at x (vector of state probabilities), which has as j-th element ${}_k\ell_j(x,t)$, and ${}_k\mathbf{k}(0,t_b)$ is the state vector at birth.

2. The probability that k, who is born in i, does not leave i before age x. It is the probability of a continuous stay in i between birth and age x.

${}^{x}_{k}p_i^{cont}(0,t_b) = \exp\left[-\int_0^x {}_k\mu_{i+}(\xi,t_b+\xi)\,d\xi\right]$ where ${}_k\mu_{i+}(\xi,t_b+\xi)$ is the escape rate from i at age ξ. If k occupies state i at age x_1, the probability that he remains in i at least until age x_2 is ${}^{x_2-x_1}_{\;\;k}p_i^{cont}(x_1,t_b+x_1) = \exp\left[-\int_{x_1}^{x_2} {}_k\mu_{i+}(\xi,t_b+\xi)\,d\xi\right]$

The probability that k, who is born in i, is in i on his x-th birthday and will stay in i for at least 10 years is

$${}^{10}_{\;k}p_i^{cont}(x,t_b+x) = \exp\left[-\int_0^{10} {}_k\mu_{i+}(x+\xi,t_b+x+\xi)\,d\xi\right] {}_k\ell_i(x_1,t_b+x_1)$$

An example is the probability that a person is on disability insurance at age 55 and will remain on disability insurance until he retires at age 65. The indicator depends on the age-specific rates of leaving the disability insurance between ages 55 and 65. Another example is the probability that a woman develops breast cancer before the age of 45 and will be successfully treated at 45, has no relapse within a fixed period (5 or 10 years). In this example, the rate of relapse depends on age only and is independent of the duration since treatment or, equivalently, the age at treatment. If the current age of the woman and the age at treatment both influence the rate of relapse, then the probability that a woman who is treated at 45, has no relapse within 5 years, i.e. before she turns 50, is

${}^{5}_{k}p_i^{cont}(x,t_b+x) = \exp\left[-\int_0^{5} {}_k\mu_{i+}(x+\xi,\xi)\,d\xi\right]$ where x is the age at treatment (x = 45) and ξ is the duration since treatment. Note that calendar time has been replaced by another time scale; namely, duration since treatment.

The probability that k, who occupies state i at birth, ever leaves state i:
$1-{}_{k}^{\omega}p_i^{cont}(0,t_b)$ where ω is the highest age. An example is the probability of ever marrying or ever emigrating from your country of birth.
The expected age at leaving i is $\int_0^{\omega} {}_{k}^{\tau}p_i^{cont}(0,t_b)\,d\tau$ or, equivalently,

$$\frac{\int_0^{\omega} x\ {}_{k}\mu_{i+}(x,t_b+x)\exp\left[-\int_0^x {}_{k}\mu_{i+}(\xi,t_b+\xi)\,d\xi\right]dx}{\int_0^{\omega} {}_{k}\mu_{i+}(x,t_b+x)\exp\left[-\int_0^x {}_{k}\mu_{i+}(\xi,t_b+\xi)\,d\xi\right]dx}$$

3. The probability that k, who occupies state i at birth, leaves i for the first time at exact age x (i.e. between x and x+dx): ${}_{k}\mu_{i+}(x,t_b+x)\exp\left[-\int_0^x {}_{k}\mu_{i+}(\xi,t_b+\xi)\,d\xi\right]dx$. It is the density of leaving i at age x. The probability that k leaves i at age x, provided k ever leaves state i, is $\dfrac{{}_{k}\mu_{i+}(x,t_b+x)\exp\left[-\int_0^x {}_{k}\mu_{i+}(\xi,t_b+\xi)\,d\xi\right]dx}{1-\exp\left[-\int_0^{\omega} {}_{k}\mu_{i+}(\xi,t_b+\xi)\,d\xi\right]}$. In demography analysis, this indicator is known as the location. An example is the probability that a woman has her first child at age x, provided she does not remain childless.

4. The probability that k, who occupies state i at birth, leaves i for the first time between ages x and x+h is $\int_0^h {}_{k}\mu_{i+}(x+\tau,t+\tau)\exp[-\int_0^{x+\tau} {}_{k}\mu_{i+}(\xi,t_b+\xi)\,d\xi]\,d\tau$ where t = t_b + x.

If the transition rate is constant then the equation is
${}_{k}^{h}m_{i+}(x,t)\int_x^{x+h} {}_{k}^{\tau}p_i^{cont}(0,t_b)\,d\tau={}_{k}^{h}m_{i+}(x,t)\ {}_{k}^{h}L_i^{cont}(x,t)\ {}_{k}^{x}p_i^{cont}(0,t_b)$ where ${}_{k}^{h}L_i^{cont}(x,t)$ is the expected sojourn time in state i on a continuous basis, i.e. if during the interval a return to state i is not permitted (see also Schoen, 1988, p. 84). The term ${}_{k}^{x}p_i^{cont}(0,t_b)$ is the probability that k remains in the state i at least till the x-th birthday, ${}_{k}^{h}L^{cont}{}_{i}(x,t)$ is the expected sojourn time in i on a continuous basis between x and x+h and ${}_{k}^{h}m_{i+}(x,t)$ is the escape rate from i between ages x and x+h.

The probability that k, who occupies state i at age x_1, leaves i before age x_2 is
$\int_{x_1}^{x_2} {}_{k}\mu_{i+}(\tau,t_b+\tau)\exp[-\int_{x_1}^{\tau} {}_{k}\mu_{i+}(\xi,t_b+\xi)\,d\xi]\,d\tau$

5. The probability that k, who is born in state i, ever leaves i for j:

$\int_0^\infty {}_k\mu_{ij}(x,t_b+x)\, {}_k^x p_i^{cont}(0,t_b)\,dx$. It is the probability of at least one direct transition from i to j during the life span (from birth to highest age ω). The expected age at transfer to j (for those transferring) is $\dfrac{\int_0^\infty x\ {}_k\mu_{ij}(x,t_b+x)\, {}_k^x p_i^{cont}(0,t_b)\,dx}{\int_0^\infty {}_k\mu_{ij}(x,t_b+x)\, {}_k^x p_i^{cont}(0,t_b)\,dx}$

The probability of advancing for the first time between ages x and x+h is

$$\int_0^h {}_k\mu_{ij}(x+\tau,t+\tau)\ {}_k^{x+\tau}p_i^{cont}(0,t_b)\,d\tau = \int_0^h {}_k\mu_{ij}(x+\tau,t+\tau)\exp[-\int_0^{x+\tau} {}_k\mu_{i+}(\xi,t_b+\xi)\,d\xi]\,d\tau$$

The expected age at the direct transition from i to j is $\dfrac{\int_0^\infty x\ {}_k\mu_{ij}(x,t_b+x)\ {}_k\ell_i(x,t_b+x)\,dx}{\int_0^\infty {}_k\mu_{ij}(x,t_b+x)\ {}_k\ell_i(x,t_b+x)\,dx}$.

The age is the expected age at first passage from i to j.

An example is the probability that k ever divorces, or ever experiences unemployment. The probability that k ever reaches retirement is another example. Individual k may experience more than one divorce or episodes of unemployment. The expected number of divorces or episodes of unemployment is different from the probability of experiencing at least one divorce or episode of unemployment. The expected number of direct transitions is considered later in this section.

If the probability is conditioned on the state occupied at birth, then the equation becomes

$$\int_0^\infty {}_k\mu_{ij}(x,t_b+x)\, {}_k^x p_{ii}(0,t_b)\,dx$$

Note that ${}_k\ell_i(x,t_b+x) = \sum_{r=1}^{I} {}_k^x p_{ri}(0,t_b)\ {}_k\ell_r(0,t_b)$. The equation states that the state probability at exact age x is a weighted sum of transition probabilities from birth to age x, weighted by the state probabilities at birth.

6. The probability that k ever advances to j from state i. This indicator does not require k to stay continuously in i before the advancement to j. Several transitions may be experienced before k return to his state of birth i and makes a direct transition to j. A candidate expression is

$$\int_0^\infty {}_k\mu_{ij}(x,t_b+x)\ {}_k\ell_i(x,t_b+x)\,dx$$

where ${}_k\ell_i(x,t_b+x)$ is the probability that individual k is alive and in state i at exact age x (state probability). The expression does not give the probability if ever advancing to j from i but the expected number of direct transitions from i to j. To obtain the probability of ever advancing to j from i, j is considered an absorbing state rather than a transient state. If j is an absorbing state, than the probability of entering j from state i is the integral of (i,j)-transition intensities over all ages weighted by the probabilities of occupying state i. The probability that k ever advances to j from state i. This indicator does not require k to stay continuously in i before the advancement to j. Several transitions may be experienced before k return to his state of birth i and makes a direct transition to j. A candidate expression is

$$\int_0^\infty {}_k\mu_{ij}(x,t_b+x)\, {}_k\ell_i(x,t_b+x)\,dx$$

where ${}_k\ell_i(x,t_b+x)$ is the probability that individual k is alive and in state i at exact age x (state probability). The expression does not give the probability if ever advancing to j from i but the expected number of direct transitions from i to j. To obtain the probability of ever advancing to j from i, j is considered an absorbing state rather than a transient state. If j is an absorbing state, than the probability of entering j from state i is the integral of (i,j)-transition intensities over all ages weighted by the probabilities of occupying state i. The transient state j is transformed into an absorbing state by substituting a zero for the transition rates out of j.

7. The probability of ever advancing to j is $\sum_{i=1}^{I}\int_0^\infty {}_k\mu_{ij}(x,t_b+x)\,{}_k\ell_i(x,t_b+x)\,dx$. It is the probability of advancing to j from any state.

4.6.2. Counts

The number of direct transitions is denoted by n and the state at the reference age, the state of origin and the state of destination are indicated by subscripts of n. The number of direct transitions from r to s experienced by k, irrespective of the state occupied at x, is ${}^h_k n_{rs}(x,t)$. The number of times individual k may expect to enter state s before he reaches age x+h is ${}^h_k n_{+s}(x,t)$. The number of direct transitions from r to s during the interval (x,x+h) provided k occupies state i at age x is ${}^{\;\;h}_{k\,ix} n_{rs}(x,t)$. The number of direct transitions from r to s during the interval (x,x+h) provided k is alive at x but irrespective of the state occupied at x is ${}^{\;\;h}_{k\,x} n_{ij}(x,t)$.

8. The expected number of times k makes a direct transition from r to s during the interval (x,x+h)

${}^h_k n_{rs}(x,t) = \int_0^h {}_k\mu_{rs}(x+\tau,t+\tau)\,{}_k\ell_r(x+\tau,t+\tau)\,d\tau$ where ${}_k\ell_r(x,t)$ is the state probability, i.e. the probability that individual k is in state r at exact age x at time t. Note the absence of subscript i. The indicator is an unconditional indicator. The position of the observer is at birth. One conditional indicator is the number of times k makes a direct transition from r to s during the interval (x,x+h), provided k is in r at age x. It is

$${}^{\;\;h}_{k\,rx} n_{rs}(x,t) = \left[\int_0^h {}_k\mu_{rs}(x+\tau,t+\tau)\,{}_k\ell_r(x+\tau,t+\tau)\,d\tau\right] / \,{}_k\ell_r(x,t) = \int_0^h {}_k\mu_{rs}(x+\tau,t+\tau)\,{}^\tau_k p_{rr}(x,t)\,d\tau$$

If the transition intensities are constant in the interval then

$${}^h_k n_{rs}(x,t) = {}^h_k m_{rs}(x,t)\int_0^h {}_k\ell_r(x+\tau,t+\tau)\,d\tau = {}^h_k m_{rs}(x,t)\;{}^h_k\overline{L}_r(x,t)$$

and

${}^{\;\;h}_{k\,rx} n_{rs}(x,t) = {}^h_k m_{rs}(x,t)\;{}^{\;\;h}_{k\,rx} L_r(x,t)$ with ${}^{\;\;h}_{k\,rx} L_r(x,t)$ the number of years k may expect to spend in state r, provided he occupies state r on his x-th birthday.

The number of direct transitions from r to s individual k may expect to make during a lifetime is

$${}^{\omega}_{k}n_{rs}(0,t_b) = \int_0^{\omega} {}_{k}\mu_{rs}(x,t_b+x)\ {}_{k}\ell_{r}(x,t_b+x)\,dx$$

The matrix equation for all states is ${}^{\omega}_{k}\mathbf{n}(0,t_b) = \int_0^{\omega} {}_{k}\boldsymbol{\mu}(x,t_b+x)\,diag\left[{}_{k}\mathbf{k}(x,t_b+x)\right]dx$ where $diag\left[{}_{k}\mathbf{k}(x,t_b+x)\right]$ is the diagonal matrix with as elements ${}_{k}\ell_{r}(x,t_b+x)$

The discrete analogue is:

${}^{\omega}_{k}n_{rs}(0,t_b) = \sum_{x=0}^{\omega-h} {}^{h}_{k}m_{rs}(x,t)\ {}^{h}_{k}\overline{L}_{r}(x,t)$ if the number is not conditioned on the state occupied at birth and ${}^{\omega}_{k\,i0}n_{rs}(0,t_b) = \sum_{x=0}^{\omega-h} {}^{h}_{k}m_{rs}(x,t)\ {}^{h}_{k\,i0}L_{r}(x,t)$ if the number is conditioned on the state occupied at birth. The first equation in matrix terms is

${}^{\omega}_{k}\mathbf{n}(0,t_b) = \sum_{x=0}^{\omega-h} {}^{h}_{k}\mathbf{m}(x,t)\ diag\left[{}^{h}_{k}\overline{\mathbf{L}}(x,t)\right]$ and the latter equation

is ${}^{\omega}_{k\,0}\mathbf{n}(0,t_b) = \sum_{x=0}^{\omega-h} {}^{h}_{k}\mathbf{m}(x,t)\ {}^{h}_{k\,0}\mathbf{L}(x,t)$

$diag\left[{}^{h}_{k}\overline{\mathbf{L}}(x,t)\right]$ is a matrix with in the diagonal the sojourn times spent by k in the different states during the (x,x+h) interval.

In multistate demography, the lifetime number of direct transitions from r to s is known as the *net mobility rate* and is denoted by ${}_{k}NMR_{rs}$ (Willekens, 1987, p. 140). It is the number of times individual k enters state s from state r in the presence of competing transitions, including death. The indicator is also used frequently in the health sciences, in particular in the prognosis of a life event such as a chronic disease. Note that the lifetime number of events is conditional on competing events such as death and transition to other states.

9. The expected number of direct transitions from r to s k may expect to make during the interval (x,x+h), provided k is born in i is

${}^{\omega}_{k\,i0}n_{rs}(0,t_b) = \int_0^{h} {}_{k}\mu_{rs}(x+\tau,t+\tau)\ {}^{x+\tau}_{k}p_{ir}(0,t_b)\ d\tau$. If the transition intensities are constant during the interval, the number of direct transitions

is ${}^{h}_{k}m_{rs}(x,t)\sum_{j=1}^{I} {}^{h}_{k}L_{jr}(x,t)\ {}^{x}_{k}\ell_{ij}(0,t_b)$ where ${}^{x}_{k}\ell_{ji}(0,t_b)$ is the probability that k, born in state i, occupies state j at age x and ${}^{h}_{k}L_{ir}(x,t)$ is the time k, who is in j at x, may expect to spend in r between x and x+h

The number direct transitions from r to s during the interval (x,x+h), provided k occupies state i at age x, is

$${}^{h}_{k\,ix}n_{rs}(x,t) = \int_0^{h} {}_{k}\mu_{rs}(x+\tau,t+\tau)\ {}^{\tau}_{k}p_{ir}(x,t)\,d\tau$$

and in case of piecewise constant transition intensities

$${}^{h}_{k\,ix}n_{rs}(x,t) = {}^{h}_{k}m_{rs}(x,t)\ {}^{h}_{k\,ix}L_{ir}(x,t)$$

10. The expected number of times k enters state s during the interval (x,x+h)

The expected number is ${}^{h}_{k}n_{+s}(x,t)=\sum_{r\neq s}^{I}{}^{h}_{k}n_{+s}(x,t)=\sum_{r\neq s}^{I}{}^{h}_{k}m_{rs}(x,t)\ {}^{h}_{k}\overline{L}_{r}(x,t)$ where ${}^{h}_{k}\overline{L}_{r}(x,t)$ is the (unconditional) number of years k spends in state r between ages x and x+h. The observer is positioned at birth. Note that ${}^{h}_{k}\overline{L}_{r}(x,t)={}_{k\,ix}^{\ \ h}L_{r}(x,t)\ {}^{x}_{k}\ell_{i}(0,t_{b})$.

The number of times k enters state s during the interval (x,x+h), provided k is in state i at x is ${}_{k\,ix}^{\ \ h}n_{+s}(x,t)=\sum_{r\neq s}^{I}{}_{k\,ix}^{\ \ h}n_{+s}(x,t)=\sum_{r\neq s}^{I}{}^{h}_{k}m_{rs}(x,t)\ {}_{k\,ix}^{\ \ h}\overline{L}_{r}(x,t)$. The expected entries into s, irrespective of the state occupied at x, is ${}^{h}_{k}n_{+s}(x,t)=\sum_{i=1}^{I}\sum_{r\neq s}^{I}{}_{k\,ix}^{\ \ h}n_{rs}(x,t)\ {}_{k}\ell_{i}(x,t)$. The observer is positioned at k's birth.

The number of times individual k may expect to enter state s during a lifetime is

$${}^{\omega}_{k}n_{+s}(0,t_{b})=\sum_{r\neq s}^{I}{}^{\omega}_{k}n_{rs}(0,t_{b})=\sum_{x=0}^{\omega}\sum_{i=1}^{I}\sum_{r\neq s}^{I}{}^{h}_{k}m_{rs}(x,t)\ {}_{k\,ix}^{\ \ h}L_{r}(x,t)\ {}_{k}\ell_{i}(x,t)$$

where ω is the highest age and ${}^{h}_{k}\ell_{i}(x,t)$ is the state probability at age x and time t. It is the expected lifetime number of episodes of s. The number of direct transitions from s to j during a lifetime is ${}^{\omega}_{k}n_{sj}(0,t_{b})$. Hence the proportion of episodes in s that end in a direct transition to j is ${}^{\omega}_{k}n_{sj}(0,t_{b})\,/\,{}^{\omega}_{k}n_{+s}(0,t_{b})$, which may also be written as ${}_{k}NMR_{sj}\ /\ {}_{k}NMR_{+s}$. Examples are the proportion of marriages that end in a divorce and the proportion of job spells that end in a period of unemployment.

The number of entries into state s, *by state occupied at exact age x*, is ${}^{h}_{k}\mathbf{n}(x,t)={}^{h}_{k}\mathbf{m}_{0}(x,t)\ {}_{k\,x}^{\ h}\mathbf{L}(x,t)$ where ${}^{h}_{k}\mathbf{m}_{0}(x,t)$ is the matrix of transition rates *with the diagonal equal to zero* and ${}_{k\,x}^{\ h}\mathbf{L}(x,t)$ is the exposure matrix denoting the expected sojourn time in the different states between x and x+h by state occupied at birth. To illustrate the matrix equation, suppose k is characterised by three attributes, i.e. the state space consists of three states. The matrix product is

$$\begin{bmatrix} {}_{k\,1x}^{\ \ h}n_{+1}(x,t) & {}_{k\,2x}^{\ \ h}n_{+1}(x,t) & {}_{k\,3x}^{\ \ h}n_{+1}(x,t) \\ {}_{k\,1x}^{\ \ h}n_{+2}(x,t) & {}_{k\,2x}^{\ \ h}n_{+2}(x,t) & {}_{k\,3x}^{\ \ h}n_{+2}(x,t) \\ {}_{k\,1x}^{\ \ h}n_{+3}(x,t) & {}_{k\,2x}^{\ \ h}n_{+3}(x,t) & {}_{k\,3x}^{\ \ h}n_{+3}(x,t) \end{bmatrix} =$$

$$\begin{bmatrix} 0 & {}^{h}_{k}m_{21}(x,t) & {}^{h}_{k}m_{31}(x,t) \\ {}^{h}_{k}m_{12}(x,t) & 0 & {}^{h}_{k}m_{32}(x,t) \\ {}^{h}_{k}m_{13}(x,t) & {}^{h}_{k}m_{23}(x,t) & 0 \end{bmatrix} \begin{bmatrix} {}_{k\,1x}^{\ \ h}L_{1}(x,t) & {}_{k\,2x}^{\ \ h}L_{1}(x,t) & {}_{k\,3x}^{\ \ h}L_{1}(x,t) \\ {}_{k\,1x}^{\ \ h}L_{2}(x,t) & {}_{k\,2x}^{\ \ h}L_{2}(x,t) & {}_{k\,3x}^{\ \ h}L_{2}(x,t) \\ {}_{k\,1x}^{\ \ h}L_{3}(x,t) & {}_{k\,2x}^{\ \ h}L_{3}(x,t) & {}_{k\,3x}^{\ \ h}L_{3}(x,t) \end{bmatrix}$$

The product ${}^{h}_{k}m_{23}(x,t)$ and ${}_{k\,1x}^{\ \ h}L_{2}(x,t)$ gives the number of direct transitions (moves) from state 2 to state 3 between ages x and x+h, provided individual k occupies state 1 at exact age x. ${}_{k\,1x}^{\ \ h}n_{+3}(x,t)$ is the number of times individual k, who occupies state 1 at x and t, may expect to enter state 3 during the interval from x to x+h if the transition intensities apply during the interval: ${}_{k\,1x}^{\;h}n_{+3}(x,t) = {}_{k\,1x}^{\;h}L_1(x,t)\;{}_{k\,1x}^{\;h}m_{13}(x,t) + {}_{k\,1x}^{\;h}L_2(x,t)\;{}_{k}^{h}m_{23}(x,t)$

An element ${}_{k\,ix}^{\;h}n_{+s}(x,t)$ of ${}_{k\,x}^{\;h}\mathbf{n}(x,t)$ denotes the number of times individual k, who occupies state i at x, may expect to enter state s during the (x,x+h)-interval. Any state of the state space except state s may be intermediate between the state occupied at x and x+h. By way of illustration, consider employment status and let i denote 'being employed' and s 'being unemployed'. Then ${}_{k\,ix}^{\;h}n_{+s}(x,t)$ is the number of times individual k who on his x-th birthday is employed enters unemployment during a period of h years following his birthday. In poverty research, ${}_{k\,ix}^{\;h}n_{+s}(x,t)$ may represent the number of times individual k, who at age x is above (below) the poverty line falls into poverty within the next h years.

The second measure of direct transition is the number of direct transitions into a state irrespective of the state occupied at the reference age. The vector ${}_{k}^{h}\overline{\mathbf{n}}(x,t)$ gives the number of direct transition into the various states during the (x,x+h)-interval irrespective of the state occupied at x. It is

$$\begin{bmatrix} {}_{k}^{h}\overline{n}_{+1}(x,t) \\ {}_{k}^{h}\overline{n}_{+2}(x,t) \\ {}_{k}^{h}\overline{n}_{+3}(x,t) \end{bmatrix} = \begin{bmatrix} 0 & {}_{k}^{h}m_{21}(x,t) & {}_{k}^{h}m_{31}(x,t) \\ {}_{k}^{h}m_{12}(x,t) & 0 & {}_{k}^{h}m_{32}(x,t) \\ {}_{k}^{h}m_{13}(x,t) & {}_{k}^{h}m_{23}(x,t) & 0 \end{bmatrix} \begin{bmatrix} {}_{k}^{h}\overline{L}_1(x,t) \\ {}_{k}^{h}\overline{L}_2(x,t) \\ {}_{k}^{h}\overline{L}_3(x,t) \end{bmatrix}$$

The third is the number of direct transitions between two given states, irrespective of the state occupancy at the reference age. It is

$${}_{k}^{h}\mathbf{n}^{*}(x,t) = {}_{k}^{h}\mathbf{m}_0(x,t)\,diag\left[{}_{k}^{h}\overline{\mathbf{L}}(x,t)\right]$$

or

$$\begin{bmatrix} {}_{k}^{h}n_{11}(x,t) & {}_{k}^{h}n_{21}(x,t) & {}_{k}^{h}n_{31}(x,t) \\ {}_{k}^{h}n_{12}(x,t) & {}_{k}^{h}n_{22}(x,t) & {}_{k}^{h}n_{32}(x,t) \\ {}_{k}^{h}n_{13}(x,t) & {}_{k}^{h}n_{23}(x,t) & {}_{k}^{h}n_{33}(x,t) \end{bmatrix} =$$

$$\begin{bmatrix} 0 & {}_{k}^{h}m_{21}(x,t) & {}_{k}^{h}m_{31}(x,t) \\ {}_{k}^{h}m_{12}(x,t) & 0 & {}_{k}^{h}m_{32}(x,t) \\ {}_{k}^{h}m_{13}(x,t) & {}_{k}^{h}m_{23}(x,t) & 0 \end{bmatrix} \begin{bmatrix} {}_{k}^{h}\overline{L}_1(x,t) & 0 & 0 \\ 0 & {}_{k}^{h}\overline{L}_2(x,t) & 0 \\ 0 & 0 & {}_{k}^{h}\overline{L}_3(x,t) \end{bmatrix}$$

4.6.3. Illustration and additional useful indicators

By way of illustration, we consider several biographic indicators to describe the life of individual k in the area of employment and disability. All summary measures are expressed in terms of state probabilities and transition intensities. Consider a multistate model with seven states: never-employed (1), presently employed (2), unemployed (3), disabled (4), out of the labour force (5), retired (6) and dead (7).

a. The probability that k is never-employed at age 25. It is

$$ {}^{25}_{k}p_1^{cont}(0,t_b) = \exp\left[-\int_0^{25} {}_k\mu_{1+}(x,t_b+x)\,dx\right] $$

b. The probability that at age 25 k has not yet entered the labour market. Individual k is in the labour market if he is employed or unemployed. If he has not yet entered the labour market by age 25, he is either never-employed or dead. The probability is being never-employed at 25 is given in (a). The probability that k dies before 25 and before he enters the labour market is $\int_0^{25} {}_k\mu_{17}(x,t_b+x)\exp\left[-\int_0^x {}_k\mu_{1+}(\xi,t_b+\xi)\,d\xi\right]dx$ since dead is an absorbing state. The second term is the probability of being in state 1 at age x and the first term is the instantaneous death rate at age x. The second term differs from the probability of being alive at age 25 which is $\int_0^{25} {}_k\mu_{17}(x,t_b+x)\left[\sum_{j=1}^{6} {}_k\ell_j(x,t_b+x)\right]dx$ with ${}_k\ell_j(x,t_b+x)$ the state probability which is obtained by solving the state equation which is a matrix equation or by solving the flow equation. The second expression accounts for the probability of dying after transitions out of the state of never-employed. The first expression accounts for the events that compete with death to be event that results in an exit from the state of never-employed. In the absence of competing events, i.e. if death would be the only event, then individual k would be in state 1 during the entire lifetime and the probability of dying would be the cumulative risk of death (cumulative incidence) $\int_0^{25} {}_k\mu_{17}(x,t_b+x)\;dx$. In the presence of competing risks, the cumulative incidence of the event of death is larger than the probability of dying. The difference between the cumulative incidence of an event and the probability of the event in the presence of competing risks is a main subject in the theory of competing risks (e.g. David and Moeschberger, 1978) and is discussed quite extensively in the medical literature (see e.g. Lloyd-Jones et al., 1999).

c. The probability that k will ever get a job. It is the probability of ever moving from state 1 to state 2: $\int_0^{\omega} {}_k\mu_{12}(x,t)\exp[-\int_0^x {}_k\mu_{1+}(\xi,t_b+\xi)\,d\xi]\,dx$. Since state 1 can only be left once, the equation is correct. If re-entry into state 1 would be possible, the equation would give the expected number of times k would enter a job from state 1, unless state 2 is converted into an absorbing state.

d. The mean age at first job: $\dfrac{\int_0^{\omega} x\; {}_k\mu_{12}(x,t_b+x)\exp[-\int_0^x {}_k\mu_{1+}(\xi,t_b+\xi)\,d\xi]\,dx}{\int_0^{\omega} {}_k\mu_{12}(x,t_b+x)\exp[-\int_0^x {}_k\mu_{1+}(\xi,t_b+\xi)\,d\xi]\,dx}$. The equation accounts for the effect of mortality and transition from never-employed to unemployment.

e. The probability that k will experience at least one episode of disability is the probability of a job ends in disability, provided disability is converted into an absorbing state. It is $\int_0^{\omega} {}_k\mu_{24}(x,t_b+x)\,{}_k\ell_2(x,t_b+x)\,dx$

f. The probability that a job ends in disability:

$$\frac{{}^{\omega}_{k}n_{24}(0,t_b)}{{}^{\omega}_{k}n_{+2}(0,t_b)} = \frac{\int_0^{\omega} {}_k\mu_{24}(x,t_b+x)\, {}_k\ell_2(x,t_b+x)\,dx}{\int_0^{\omega}\sum_{r\neq 2}^{I} {}_k\mu_{r2}(x,t_b+x)\, {}_k\ell_r(x,t_b+x)\,dx}$$

The numerator is the number of jobs that end in disability and the denominator is the number of jobs individual k may expect during the life span. Since k may experience several job episodes during his lifetime, one or several may end in disability.

g. The probability that individual k will be disabled at age 55: ${}_k\ell_4(55,t_b+55)$. To compare different birth cohorts, different values of t_b may be selected.

h. The number of times k may expect to return to a job after an episode of disability: $\int_0^{\omega} {}_k\mu_{42}(x,t_b+x)\, {}_k\ell_4(x,t_b+x)\,dx$. If the rate of returning to a job after a period of disability is constant, then the expected number of returns is ${}_k\mu_{42}\ {}^{\omega}_{k}\overline{L}_4(0,t_b)$.

i. The rate at which k will return to a job during an episode of disability (exit rate of disability for a job): $\frac{\int_0^{\omega} {}_k\mu_{42}(x,t_b+x)\, {}_k\ell_4(x,t_b+x)\,dx}{\int_0^{\omega} {}_k\ell_4(x,t_b+x)\,dx}$ where the denominator is the total duration of disability, aggregated over all spells of disability. It is ${}^{\omega}_{k}\overline{L}_4(0,t_b)$.

j. The number of years individual k may expect to spend in disability: ${}^{\omega}_{k}\overline{L}_4(0,t_b) = \int_0^{\omega} {}_k\ell_4(x,t_b+x)\,dx$

k. The average duration of a disability episode: $\frac{\int_0^{\omega} {}_k\ell_4(x,t_b+x)\,dx}{\int_0^{\omega} {}_k\mu_{24}(x,t_b+x)\, {}_k\ell_2(x,t_b+x)\,dx}$. The numerator is the total duration of disability and the denominator is the number of disability episodes.

l. The expected age at entry into disability (from employment): $\frac{\int_0^{\omega} x\ {}_k\mu_{24}(x,t_b+x)\, {}_k\ell_2(x,t_b+x)\,dx}{\int_0^{\omega} {}_k\mu_{24}(x,t_b+x)\, {}_k\ell_2(x,t_b+x)\,dx}$. The denominator is the number of times individual k leaves a job because of disability. It is the number of disability episodes. Note the difference with the expected age at *first* entry into disability.

m. The probability that k will die during an episode of disability. It is the probability of moving from state 4 to state 7: $\int_0^{\omega} {}_k\mu_{47}(x,t)\, {}_k\ell_4(x,t_b+x)\,dx$

n. The probability that k will die while employed: $\int_0^{\omega} {}_k\mu_{27}(x,t)\, {}_k\ell_2(x,t_b+x)\,dx$. The expected age at death is $\frac{\int_0^{\omega} x\ {}_k\mu_{27}(x,t)\, {}_k\ell_2(x,t_b+x)\,dx}{\int_0^{\omega} {}_k\mu_{27}(x,t)\, {}_k\ell_2(x,t_b+x)\,dx}$

o. The probability that individual k will not reach retirement, i.e. dies before retirement. It is the probability that k enters state 7 from state 1, 2, 3, 4 or 5:

$\sum_{r=1}^{5} \int_0^{\omega} {}_k\mu_{r7}(x,t)\, {}_k\ell_r(x,t_b+x)\, dx$

Note that the retirement age is not fixed at 65 but is determined by the age profile of the instantaneous rates of retiring from any of the states that may proceed retirement.

p. The proportion of life spent in retirement: $\dfrac{\int_0^{\omega} {}_k\ell_6(x,t_b+x)\, dx}{\int_0^{\omega} {}_k\ell_+(x,t_b+x)\, dx} = \dfrac{{}^{\omega}_{k}\overline{L}_6(0,t_b)}{{}^{\omega}_{k}\overline{L}_+(0,t_b)}$ where the numerator is the expected number of years in retirement, ${}_k\ell_+(x,t_b+x) = \sum_{i=1}^{I} {}_k\ell_i(x,t_b+x)$ and ${}^{\omega}_{k}\overline{L}_+(0,t_b) = \sum_{i=1}^{I} {}^{\omega}_{k}\overline{L}_i(0,t_b)$

Analogous summary measures are used in studies of marriage and divorce. Schoen (1988, p. 95) considers several indicators for a multistate model with five states: never-married (1), presently married (2), divorced (3), widowed (4) and dead (5).

4.7. From biographic indicators to economic and actuarial indicators: a brief note

The biographic model generates expected state occupancies, transitions and sojourn times. A state occupancy may involve an activity, such as consumption and saving. A state occupancy and a state transition may involve a payment. By associating a particular level of consumption and/or saving to state occupancies, a micro-economic model emerges that may be used to determine optimal consumption and savings patterns over the life cycle. By associating payments to state occupancies and state transitions, a biographic actuarial model emerges that is generic and encompasses most insurance schemes that exist today. A *payment function* indicates what needs to be paid, who should pay and who should receive the payment.

4.7.1. A lifetime consumption model

By way of illustration consider the consumption module of the life-cycle overlapping generation model proposed by Heijdra and Romp (2005). The authors present Blanchard-Yaari type overlapping-generations model incorporating a realistic description of demography (mortality). The other part, the budget constraint, is not considered. Let ${}_kc(\xi,t_b+\xi)$ be the consumption level of individual k at age ξ at time $t_b+\xi$, where t_b is the date of birth. To most individuals, consumption in a distant future is worth less than the same consumption in the immediate future, meaning that with the consumption is associated a time preference. If the time preference is independent of age, it may be represented by the constant rate of time preference θ. At time of birth the consumption level of ${}_kc(\xi,t_b+\xi)$ at instantaneous age ξ is worth $\exp[-\theta\, \xi]\ {}_kc(\xi,t_b+\xi)$ since the rate of time preference is independent of age. The expected utility at birth of a lifetime of consumption is the utility function $\Lambda(0,t_b) = \int_0^{\omega} \exp[-{}_k\theta\, x]\ {}_kc(x,t_b+x)\ {}_k\ell(x)\, dx$ where

${}_k\ell(x)$ is the probability that k survives to age x: ${}_k\ell(x) = \exp\left[-\int_0^x {}_k\mu(\xi, t_b + \xi)\, d\xi\right]$ where ${}_k\mu(\xi, t_b + \xi)$ is the instantaneous rate of death at age ξ at time t = t_b+ ξ. The integral is the cumulative hazard: ${}_k\mathrm{M}(x,t) = \int_0^x {}_k\mu(\xi, t_b + \xi)\, d\xi$. The utility function may be written as $\Lambda(0,t_b) = \int_0^\omega {}_kc(x, t_b + x)\ \exp[-({}_k\theta\, x + {}_kM\,(x, t_b + x))]\,dx$ which is equation (2.5) in the paper by Heijdra and Romp (2005, p. 5). The instantaneous rate of death represents the mortality risk. Because of the uncertainty of survival, individuals discount future felicity more heavily than in the absence of mortality risk (Yaari, 1965, quoted by Heijdra and Romp, 2005, p. 2).

In addition to the lifetime utility, the utility may be determined for any age x of individual k. The utility at x depends on the expected consumption during the *remaining* lifetime. The utility at age x, provided k reaches age x, is

$$\Lambda(x, t_b + x) = \frac{1}{{}_k\ell(x)}\int_x^\omega \exp[-{}_k\theta\tau]\ {}_kc(\tau, t_b + \tau)\ {}_k\ell(\tau)\, d\tau$$
$$= \int_x^\omega {}_kc(\tau, t_b + \tau)\ \exp\left[-[{}_k\theta\tau + \int_\tau^\omega {}_k\mu(\xi, t_b + \xi)\, d\xi\right] d\tau$$

The utility at age x depends on the consumption and the mortality risks at ages beyond x and on the time preference. Given the biographic model presented above, the utility function may easily be extended to cases in which consumption, mortality and time preference at a given age depend on the state occupied at that age.

4.7.2. An actuarial model

To derive a simplified version of an actuarial model, suppose individual k pays a premium if he occupies some states and receives a benefit if he occupies some other states. The *instantaneous premium rate* is ${}_kp_i(x,t)$ and the amount paid during the interval (x, x+dx) is ${}_kp_i(x,t)$ dx. The *instantaneous benefit rate* is ${}_kb_j(x,t)$ and the amount the beneficiary receives (from e.g. an insurer of the government) during the interval (x, x+dx) is ${}_kb_i(x,t)$ dx. Suppose individual k is born at t_b and is aged x at time t = t_b+x. The present value (PV) at birth of that benefit received at x is exp(-δx) ${}_kb_j(x,t_b+x)$, where δ is the instantaneous rate of interest, which is assumed to be constant. The expression exp(-δx) is the discount function and exp(-δ) the annual discount factor. The PV of the benefit individual k receives at age x because of the state occupancy j, is ${}_k\mathrm{B}_j(x,t) = \exp[-\delta\, x]\,{}_kb_j(x,t)\ {}_kY_j(x,t)\,dx$ where ${}_kY_j(x,t)$ is indicator variable which is 1 if individual k occupies state j at age x at time t. The expected value of being in state j at x is the state probability ${}_k\ell_j(x,t) = E\lfloor {}_kY_j(x,t) \rfloor$. The actuarial value of an annuity benefit received during the life span is the expected present value

$$E\left[{}_k^\omega\mathrm{B}(0,t_b)\right] = \int_0^\omega \exp[-\delta\, x]\left[\sum_{j=1}^I {}_kb_j(x, t_b + x)\ {}_k\ell_j(x, t_b + x)\right]dx$$

The benefit depends on the expected sojourn time is each state, the amount paid in each state and the discount factor.

The premium paid while in state i during the infinitesimal interval [x, x+dx) is ${}_kp_i(x)$. The PV of the premium paid is determined in a way analogous to the benefit. Let ${}^{\omega}_{k}\Pi_i(0,t_b)$ denote the PV of the premium paid during the different episodes when individual k is in state i between birth and death. It is obtained by the following expression:

$${}^{\omega}_{k}\Pi_i(0,t_b) = \int_0^{\omega} \exp[-\delta\, x]\; {}_kp_i(x,t_b+x)\; {}_kY_i(x,t_b+x)\,dx$$

The actuarial value of the premium paid is the expected present value

$${}^{\omega}_{k}\Pi_i(0,t_b) = \int_0^{\omega} \exp[-\delta\, x]\; {}_kp_i(x,t_b+x)\; {}_k\ell_i(x,t_b+x)\,dx$$

The expected PV at birth of all premiums paid by individual k during his lifetime is

$$E\left[{}^{\omega}_{k}\Pi_i(0,t_b)\right] = \int_0^{\omega} \exp[-\delta\, x]\left[\sum_{i=1}^{I} {}_k\, p_i(x,t_b+x)\; {}_k\ell_i(x,t_b+x)\right]dx$$

The difference between premiums paid and benefits received varies over the life course. A premium reserve exists when the actuarial value of future benefits exceeds the actuarial value of future premiums. The premium reserve may be defined prospectively or retrospectively. The *prospective premium reserve* at age x is the actuarial value of the future benefits less the actuarial value of the future premiums: $E\left[{}^{\omega-x}_{k}\mathrm{B}(x,t_b+x)\right] - E\left[{}^{\omega-x}_{k}\Pi(x,t_b+x)\right]$. It is a summary of benefit minus premium payment streams between x and the highest age ω and is aggregated over all states. The *retrospective premium reserve* is the actuarial accumulated value of the past premiums minus past benefits. It is a summary of the premium minus benefit payment streams leading up to x. The retrospective reserve at age x is defined over the interval (0, x) while the prospective reserve is defined over the interval (x, ω). If the prospective reserve is positive, the expected future benefits exceed the expected premiums. In the absence of a retrospective reserve, the insurer or the government has a funding requirement. To restore the balance, premiums may be raised or benefits may be reduced.

The *equivalence principl*e is fulfilled if at a given age or point in time the expected value of future benefits is equal to the expected value of future premiums. If the premium individual k pays is determined by the equivalence principle, the premiums paid cover the expected benefit payments. The insurance scheme is *actuarially fair* at the individual level. Actuarial fairness is usually determined at the group level. If for a group of Q individuals the net premium is determined such that $\sum_{k=1}^{Q} E\left[{}^{\omega}_{k}\mathrm{B}(0,t_b)\right] = \sum_{k=1}^{Q} E\left[{}^{\omega}_{k}\Pi(0,t_b)\right]$, then the insurance organization can meet its obligations (net of administrative expenses) and the insurance system is actuarially fair for the group although some members of the group may benefit more than others. If the equivalence principle is the basis for setting the premium, the insurance scheme is actuarially fair. For further details, see Haberman and Pitacco (1999) and Willekens (2007).

5. The cohort biography

5.1. Introduction

Individual k, who is born at instantaneous time t_b, is a member of a birth cohort. A cohort is an aggregate of individuals (within some population definition) who experienced the same event within the same time interval (Ryder, 1965). If birth is the *cohort-defining event,* the group of individuals is born during the same period. They constitute a birth cohort. They differ from contemporaries, defined as people alive at the same time. Individuals born during a same period share characteristics that give each cohort a distinctive character. The character reflects the historical context. "In an epoch of change, each person is dominated by his birth date. He derives his philosophy from his historical world, the subcultures of his cohort. The community of dates equips each cohort with its own expanse of time, its own style and its own truth. The ideas, sentiments and values of members of the same cohort converge; their actions become quasi-organized." (Ryder, 1985, p. 29). The terms cohort biography and cohort life cycle are from Ryder (1965). In his view, the cohort biography, as macro-biography, is the aggregate analogue of the individual life history. In that view, the cohort is the temporal unit in the analysis of social change since historical events modify people of different ages in different ways, and the effects of these transformations are persistent. Ryder concludes that two broad orientations for theory and research flow from this position: first, the study of intra-cohort temporal development throughout the life cycle; second, the study of comparative cohort careers, i.e. inter-cohort temporal differentiation in the various parameters that may be used to characterize the aggregate histories. This section and the entire paper are written in the spirit that follows from Ryder's vision. The individual biography was the subject of the previous section and the cohort biography is the subject of the current section. Both the individual and the cohort biography are characterized by the same set of fundamental parameters; namely, the transition intensities. Although members of a same cohort have life histories that characterized by different transition, the fundamental parameters of the life course are likely to be similar if compared to the transition intensities of members of a different cohort. Individual differences can be attributed to differences in personal attributes, experiences and circumstances, and the effect of chance. If chance would be the only factor, all members of a same cohort would have the same expected value of a transition intensity. That perspective is used in this section to address inter-cohort differentiation of transition intensities in the next section which approaches a population as being composed of successive birth cohorts. Since inter-cohort differentiation will be expressed in terms of transition intensities, individual transition intensities must be aggregated into cohort transition intensities. The aggregation is carried out in an age-time framework visualized by the Lexis diagram. Two major schemes of aggregation are considered: the age-cohort scheme, which is most appropriate for cohort analysis exemplified by life-table analysis, and the period-cohort scheme, which is most appropriate for period analysis, exemplified by population projections.

The operationalization of the cohort biography in this paper is a simplification. The historical context affects every single cohort member, but members respond differently to the historical factors. Consequently, the cohort biography is not merely a summation of individual biographies. Recall the distinction between individual state variables (i-state variables) and population state variables (p-state variables), introduced in Section 1. Population state variables describe population characteristics, such as the number of individuals in a given state at a given age or time, and the total sojourn time in that state. The state variables relate to state occupancies and state transitions. The aggregation of individual state variables (biographies) into cohort state variables (biographies) is dependent on real or assumed patterns of interaction between individuals (Metz and Diekmann, 1986). Interactions may involve both competition and cooperation. It may result in connectedness of individuals and the emergence of social institutions that facilitate and constrain the behaviour of individuals. In this paper interactions between individuals are assumed to be absent. The cohort biography is operationalized as the expected value of the individual biography: the cohort state variables are expected values of individual state variables. Individual state variables have been denoted by ${}_kY(x,t)$ with the indicator variable ${}_kY_i(x,t)$ denoting the state occupied and ${}_kY_{ij}(x,t)$ the state transition. In the previous section the expected values $E[{}_kY(x,t)]$ have been used to predict individual biographies. In this section, the expected values are used to predict cohort biographies.

The period of birth is assumed to start at t_c and to be of length h. Hence the birth cohort consists of the group of individuals for which the date of birth t_b is between t_c to t_c+h ($t_c \le t_b < t_c+h$) . The number of births in an interval of unit length following t_b is denoted by $b(t_b)$. It is referred to as the arrival rate of newborns at t_b. If the number of newborns is observed, $b(t_b)$ is the empirical arrival rate. If the number of newborns is predicted by a model, e.g. the Poisson process model, the arrival rate is the expected number of arrivals. For an introduction to the arrival rate in the context of the Poisson process, see e.g. Çinlar (1975, p. 75). The number of children born during the infinitesimally small period (dt) following time t_b is $b(t_b)dt$. In this section, the date of birth will also be written as $t_b= t_c+\tau$ with $0 \le \tau < h$. The number of children born during the interval ($t_c+\tau$, $t_c+\tau+d\tau$) is $b(t_c+\tau)d\tau$. The number of children born during the interval of length h following t_c is ${}^hB(t_c) = \int_0^h b(t_c+\tau)\,d\tau$. If births are uniformly distributed during the interval (t_c, t_c+h), i.e. if the arrival rate of newborns is constant, then the number of children born during the interval $d\tau$ is independent of τ, hence $b(t_c+\tau)\,d\tau = b(t_c)\,d\tau$ for all τ between 0 and h. In case the birth are uniformly distributed, the number of births during the interval is ${}^hB(t_c) = b(t_c)\int_0^h d\tau = h\,b(t_c)$ where $b(t_c)$ is the constant arrival rate of newborns during the interval (t_c, t_c+h).

The newborns occupy one of several states. Newborns in a same state may be viewed as a subcohort. The state occupied by a child born at $t_b = t_c+\tau$ is given by the random variable $Y(0,t_b)$ or the indicator variable $Y_i(0,t_b)$ (for all $i \in I$). The expected value of $Y_i(0,t_b)$ is $E[Y_i(0,t_b)] = \ell_i(0,t_b) = \Pr\{Y_i(0,t_b)=1\} = \Pr\{Y(0,t_b)=i\}$. The states occupied by the children born at t_b are given by the vector of state probabilities at birth, $\mathbf{k}(0,t_b)$. The number of children born during a unit time interval following t_c that occupy state i at birth is $\ell_i(0,t_c)b(t_c)$. The number of children born during the infinitesimal interval ($t_c+\tau$, $t_c+\tau+d\tau$) by state occupied at birth is $\mathbf{k}(0,t_c+\tau)\, b(t_c+\tau)d\tau$.The number of newborns during the interval (t_c, t_c+h) by state occupied at birth is

${}_c^h\mathbf{K}(0,t_c) = \int_0^h b(t_c+\tau)\mathbf{k}(0,t_c+\tau)\,d\tau$ where the left subscript c denotes cohort and ${}_c^h\mathbf{K}(0,t_c)$ is the vector of numbers of newborns during the interval (t_c, t_c+h) by state occupied at birth. An element ${}_c^hK_i(0,t_c)$ of ${}_c^h\mathbf{K}(0,t_c)$ denotes the number of newborns during the interval (t_c, t_c+h) occupying state i at birth. If the state occupied at birth is independent of when the child is born during the interval (t_c, t_c+h) then $\mathbf{k}(0,t_c+\tau) = \mathbf{k}(0,t_c)$ for $0 \leq \tau < h$ and ${}_c^h\mathbf{K}(0,t_c) = \mathbf{k}(0,t_c)\int_0^h b(t_c+\tau)\,d\tau = \mathbf{k}(0,t_c)\,{}^hB(t_c)$.

Members of a birth cohort celebrate their x-th birthday during a same period. Children born in the period (t_c, t_c+h) celebrate their x-th birthday in the period (t_c+x, t_c+x+h), i.e. between time $t = t_c+x$ and $t+h = t_c+x+h$ (see Lexis diagram). At a given exact time t the members of a birth cohort belong to the same age group. For instance, at instantaneous time t children born between t_c and t_c+h are between age x-h and x. At time t children born between t_c-h and t_c are between exact ages x and x+h. The experience of a birth cohort may be registered at *instantaneous ages* x, in which case the registration extends over an interval of time, or at *instantaneous points in time* t, in which case the registration extends over an age interval. The first type of registration is consistent with the age-cohort observation plan: age is a continuous variable and time is a discrete variable. The second case is consistent with the period-cohort observation plan: time is a continuous variable and age is a discrete variable. The age-cohort perspective is central in life-table and life-history analysis and the period-cohort perspective is central in demographic projections. The two perspectives are considered in this section. The age-period perspective is indicated by the subscript c and the period-cohort perspective by the subscript p. We assume that all cohort members experience the same age-specific transition intensities.

5.2. Age-cohort perspective

Let ${}_kY_i(0,t_c+\tau)$ be an indicator variable that is one if individual k is born at time $t_c+\tau$ and zero otherwise. The number of children born between $t_c+\tau$ and $t_c+\tau+d\tau$ occupying state i at birth is ${}_cN_i(0,t_c+\tau)\,d\tau = [\sum_{k=1}^{Q} {}_kY_i(0,t_c+\tau)]\,d\tau$. The number of children born between t_c and t_c+h, irrespective of the state occupied at birth, is

${}_c^hN_+(0,t_c) = \int_0^h \left[\sum_{k=1}^{Q}\sum_{i=1}^{I} {}_kY_i(0,t_c+\tau)\right]d\tau$. The *expected* number of newborns between t_c and t_c+h is ${}^hB(t_c) = E\left[{}_c^hN_i(0,t_c)\right]$.

Let N(x,t) denote the density of individuals aged x at time t, with $t = t_b+x$. In the age-cohort perspective, the cohort experience is measured at instantaneous ages. The measurement extends over time intervals of length h. The state occupancies at age x is expressed in terms of the state occupancies of the individual cohort members. The number of individuals who reach their x-th birthday during the period from t to t+h and who occupy state i at that age is

$${}_c^hN_i(x,t) = \int_0^h N_i(x,t+\tau)\,d\tau = \int_0^h \left[\sum_{k=1}^{Q} {}_kY_i(x,t+\tau)\right] d\tau$$

where the subscript c denotes cohort and h denotes the length of the interval. The indicator variable ${}_kY_i(x,t+\tau)$ is one if individual k reaches his x-th birthday at time t+ τ ($0 \leq \tau \leq h$) and occupies state i at that time. It is zero otherwise. In the Lexis diagram, the number of individuals who reach their x-th birthday during the period (t, t+h) is represented by the number of lifelines crossing the segment VW. The individuals are born between t-x and t-x+h.

The *expected* number of cohort members in state i at exact age x is ${}_c^hK_i(x,t_c+x) = E\left[{}_c^hN_i(x,t_c+x)\right] = k_i(x,t_c+x)\ {}^hB(t_c)$. The expected number of cohort members at age x by state occupied at x is given by the following equation:

$${}_c^h\mathbf{K}(x,t_c+x) = \mathbf{k}(x,t_c+x)\ {}^hB(t_c)$$

where $\mathbf{k}(x,t_c+x)$ are state probabilities at instantaneous age x (the same for all individuals born during the interval from t_c to t_c+h) and ${}_c^h\mathbf{K}(x,t_c+x)$ is the vector with as elements the number of individuals born during the period from t_c-x-h and t_c-x by state occupied at exact age x (age-cohort perspective). An element ${}_c^hK_i(x,t_c+x)$ of ${}_c^h\mathbf{K}(x,t_c+x)$ denotes the number of cohort members occupying state i on their x-th birthday (instantaneous age x) sometime during the interval from t_c+x to t_c+x+h. In the age-cohort perspective, x denotes the exact age and t_c+x denotes the time interval from t_c+x to t_c+x+h.

The vector of state occupancies may be expressed in terms of the transition probabilities:

$${}_c^h\mathbf{K}(x,t_c+x) = {}_c^x\mathbf{P}(0,t_c)\,\mathbf{k}(0,t_c)\ {}^hB(t_c) = \exp\left[-\int_0^x \boldsymbol{\mu}(\tau,t_c+\tau)\,d\tau\right]\mathbf{k}(0,t_c)\ {}^hB(t_c)$$

where $\mathbf{k}(0,t_c)$ is the vector of state probabilities at exact age 0, i.e. at birth, for individuals born in period (t_c, t_c+h) and ${}_c^h\mathbf{P}(0,t_c)$ is the matrix of discrete-time transition probabilities between exact ages 0 and x for individuals born in year t_c. ${}_c^x\mathbf{P}(0,t_c)$ is a matrix of age-cohort transition probabilities. The age-cohort projection model is

$${}_c^h\mathbf{K}(x+h,t+h) = {}_c^h\mathbf{P}(x,t)\ {}_c^h\mathbf{K}(x,t) = \exp\left[-\int_0^h \boldsymbol{\mu}(x+\tau,t+\tau)\,d\tau\right] {}_c^h\mathbf{K}(x,t)$$

where t = t_c+x. ${}_c^h\mathbf{P}(x,t)$ is a matrix of discrete-time *age-cohort transition probabilities*. If the transition intensities are constant during the age-cohort interval from x to x+h and t to t+h (parallelogram PVSQ in the Lexis diagram), then

$${}_c^h\mathbf{K}(x+h,t+h) = {}_c^h\mathbf{P}(x,t)\ {}_c^h\mathbf{K}(x,t) = \exp\left[-{}_c^h\mathbf{m}(x,t)\int_0^h d\tau\right] {}_c^h\mathbf{K}(x,t) = \exp\left[-h\ {}_c^h\mathbf{m}(x,t)\right] {}_c^h\mathbf{K}(x,t)$$

Note that ${}_c^h\mathbf{m}(x,t)$ is a matrix of *age-cohort transition rates*, i.e. transition rates that apply to the age-cohort interval.

The time spent in the different states during the age-cohort interval from x to x+h and t to t+h (parallelogram PVSQ in the Lexis diagram) by cohort members who survive at exact age x, by state occupied at the x-th birthday, is

$$ {}_{c}^{h}\mathbf{L}(x,t) = \int_0^h {}_{c}^{h-\tau}\mathbf{P}(x+\tau,t+\tau)d\tau = \int_0^h \exp\left[-\tau\, {}_{c}^{h}\mathbf{m}(x,t)\right]d\tau $$

$$ = \left[{}_{c}^{h}\mathbf{m}(x,t)\right]^{-1}\left[\mathbf{I} - \exp[-h\, {}_{c}^{h}\mathbf{m}(x,t)]\right] $$

${}_{c}^{h}\mathbf{L}(x,t)$ is the age-cohort exposure function by state occupied at the reference age, which is age x. The sojourn times in the different states between exact ages x and x+h by all members of the birth cohort (t-x-h, t-x), irrespective of the state occupied at x, is ${}_{c}^{h}\overline{\mathbf{L}}(x,t)\ {}^{h}B(t_c) = {}_{c}^{h}\mathbf{L}(x,t)\,\mathbf{k}(x,t)\ {}^{h}B(t_c)$ where $t = t_c + x$.

The number of events experienced by the cohort members between ages x and x+h is ${}_{c}^{h}\mathbf{n}^{*}(x,t) = {}_{c}^{h}\mathbf{m}(x,t)\, diag\left[{}_{c}^{h}\overline{\mathbf{L}}(x,t\right]\ {}^{h}B(t_c)$

5.3. Period-cohort perspective

Let ${}_{k}Y_i(x,t)$ be an indicator variable which is one if individual k occupies state i at age x at time t. The number of individuals aged x+ξ to x+ξ+dξ at instantaneous time t and occupying state i at t is ${}_{p}N_i(x+\xi,t)\,d\xi = [\sum_{k=1}^{Q} {}_{k}Y_i(x+\xi,t)]d\xi$ where p denotes period. The number of cohort members aged x to x+h at exact time t and in state i at that time is

$$ {}_{p}^{h}N_i(x,t) = \int_0^h N_i(x+\xi,t)\,d\xi = \int_0^h \left[\sum_{k=1}^{Q} {}_{k}Y_i(x+\xi,t)\right]d\xi $$

${}_{k}Y_i(x+\xi,t)$ is one if individual k occupies state i at exact time t and if at that time he is aged x+ ξ $(0 \le \xi \le h)$. It is zero otherwise. In the Lexis diagram, that number of individuals is represented by the number of lifelines crossing the segment VQ (for h = 1). These individuals are born between t-x-h and t-x. The total population at time t is the number of individuals at t irrespective of age:

$$ {}_{p}^{\omega}N_{+}(0,t) = \int_0^{\omega} N_{+}(\xi,t)\,d\xi = \int_0^{\omega}\left[\sum_{k=1}^{Q}\sum_{i=1}^{I} {}_{k}Y_i(\xi,t)\right]d\xi $$

where ω is the highest age. The population at t is the number of cohort members that is surviving at exact time t. The proportion of the population in state i at time t aged x to x+h is

$$ \frac{{}_{p}^{h}N_i(x,t)}{{}_{p}^{\omega}N_i(0,t)} = \frac{\int_x^{x+h} N_i(\xi,t)\,d\xi}{\int_0^{\omega} N_i(\xi,t)\,d\xi} $$

The mean age of the population in i is[11]

[11] Starting from that measure, Vaupel and Canudas-Romo (2002) and Canudas-Romo (2003) develop formulae for decomposing change in a population average (in this case the mean age) over time into two components, an effect of changes in the characteristics of interest and an effect of changes in population composition.

$$\bar{x}_i = \frac{\int_0^{\omega} \xi\, N_i(\xi,t)\, d\xi}{\int_0^{\omega} N_i(\xi,t)\, d\xi}$$

a. Population at time t = t_c+h

Consider a child born at time t_b ($t_c \leq t_b < t_c$+h) and let $t_b = t_c + \tau$. At the end of the interval (t_c, t_c+h), i.e. at instantaneous time t_c+h, the child is h-τ years old. At instantaneous time t = t_c+x+h, the child is x+h-τ years old. The state probabilities at t_c+h are

$$\mathbf{k}(h-\tau, t_c+h) = \exp\left[-\int_0^{h-\tau} \boldsymbol{\mu}(\xi, t_c+\tau+\xi)\, d\xi\right]\mathbf{k}(0,t_c+\tau) = {}^{h-\tau}\mathbf{P}(0,t_c+\tau)\,\mathbf{k}(0,t_c+\tau)$$

where an element $\mu_{ij}(\xi, t_c+\tau+\xi)$ of $\boldsymbol{\mu}(\xi, t_c+\tau+\xi)$ is the instantaneous rate of transition from i to j at instantaneous age ξ experienced by an individual born at instantaneous time t_c+τ. The probability of a transition from i to j during the interval dξ following exact age ξ is $\mu_{ij}(\xi, t_c+\tau+\xi)\, d\xi$.

Next we consider all children born during the small interval dτ following t_c+τ with 0 ≤ τ < h. The number of children born during that small interval is $b(t_c+\tau)\, d\tau$. The state probabilities at birth are given by the state vector $\mathbf{k}(0,t_c+\tau)$. The time between birth and the end of the interval (t_c, t_c+h) is h-τ and the probability that a child born at t_c+τ and occupying state i at birth, survives and occupies state j at the end of the interval at instantaneous time t_c+h is ${}^{h-\tau}_{p}p_{ij}(0,t_c+\tau)$. It is an element of the matrix of transition probabilities ${}^{h-\tau}_{p}\mathbf{P}(0,t_c+\tau)$. The number of children born during the interval (t_c+τ, t_c+τ+dτ) surviving at instantaneous time t_c+h, by state occupied at t_c+h is

$$\begin{aligned}{}^{d\tau}_{p}\mathbf{K}(h-\tau, t_c+h) &= {}^{h-\tau}_{p}\mathbf{P}(0,t_c+\tau)\ \mathbf{k}(0,t_c+\tau)\, b(t_c+\tau)\, d\tau \\ &= \exp\left[-\int_0^{h-\tau} \boldsymbol{\mu}(\xi, t_c+\tau+\xi)\, d\xi\right]\mathbf{k}(0,t_c+\tau)\, b(t_c+\tau)\, d\tau\end{aligned}$$

At t_c+h these children are (h-τ-dτ, h-τ) years old.

Now consider all children born during the interval (t_c, t_c+h). At instantaneous time t_c+h, these children are between 0 and h years old. The number of children at instantaneous time t_c+h, by state occupied is

$$\begin{aligned}{}^{h}_{p}\mathbf{K}(0, t_c+h) &= \int_0^{h} {}^{h-\tau}_{p}\mathbf{P}(0,t_c+\tau)\ \mathbf{k}(0,t_c+\tau)\, b(t_c+\tau)\, d\tau \\ &= \int_0^{h}\exp\left[-\int_0^{h-\tau} \boldsymbol{\mu}(\xi, t_c+\tau+\xi)\, d\xi\right]\mathbf{k}(0,t_c+\tau)\, b(t_c+\tau)\, d\tau\end{aligned}$$

where an element ${}^{h}_{p}K_i(0,t_c+h)$ of ${}^{h}_{p}\mathbf{K}(0,t_c+h)$ denotes the number of children born in the interval (t_c, t_c+h) that occupies state i at instantaneous time t_c+h. At time t_c+h, the children are between age 0 and h. That age group is denoted by 0 in ${}^{h}_{p}\mathbf{K}(0,t_c+h)$. Hence in period-cohort perspective, 0 denotes an age interval and t_c+h denotes an exact time.

${}^{h-\tau}_{p}\mathbf{P}(0,t_c+\tau)$ is the probability that a child born at time $t_c+\tau$ in the interval (t_c, t_c+h) survives to exact time t_c+h by state occupied at birth and state occupied at t_c+h. ${}^{h-\tau}_{p}\mathbf{P}(0,t_c+\tau)$ is a matrix of *period-cohort transition probabilities*. The equation assumes that the age-specific and time-varying transition rates are the same for all cohort members born during the interval (t_c, t_c+h).

If the births are uniformly distributed during the interval (t_c, t_c+h), then

$${}^{h}_{p}\mathbf{K}(0,t_c+h)=\int_0^h \exp\left[-\int_0^{h-\tau}\boldsymbol{\mu}(\xi,t_c+\tau+\xi)\,d\xi\right]\mathbf{k}(0,t_c+\tau)\,b(t_c)\,d\tau$$

If in addition, the state probabilities are independent of the date of birth provided the birth is during the interval (t_c, t_c+h) then

$${}^{h}_{p}\mathbf{K}(0,t_c+h)=\left[\int_0^h \exp[-\int_0^{h-\tau}\boldsymbol{\mu}(\xi,t_c+\tau+\xi)\,d\xi]\,d\tau\right]\mathbf{k}(0,t_c)\,b(t_c)$$

If the transition intensities do not vary with age and time during the age-period-cohort interval between birth in the interval (t_c, t_c+h) and instantaneous time t_c+h (visualized by the triangle MNT in the Lexis diagram), then $\boldsymbol{\mu}(\xi,t_c+\tau+\xi)={}^{h}_{p}\mathbf{m}(00,t_c)$ for $0 \le \tau < h$ and $0 \le \xi < h-\tau$. The 'age' 00 refers to children born during the interval (t_c, t_c+h) with t_c the starting date of the interval. At the end of the interval, i.e. at t_c+h, these children are between 0 and h years old. An element ${}^{h}_{p}m_{ij}(00,t_c)$ of ${}^{h}_{p}\mathbf{m}(00,t_c)$ is the rate of transition from i to j experienced by a child born during the interval (t_c, t_c+h) before the end of the interval. At time t_c+h, the distribution of the birth cohort between the functional states, is ${}^{h}_{p}\mathbf{K}(0,t_c+h)=\left[\int_0^h \exp[-(h-\tau)\ {}^{h}_{p}\mathbf{m}(00,t_c)]\,d\tau\right]\mathbf{k}(0,t_c)\,b(t_c)$ provided the births are uniformly distributed during the interval (t_c, t_c+h) and the state probabilities for birth during that interval are time-invariant.

Solution of the integral yields

$${}^{h}_{p}\mathbf{K}(0,t_c+h)=\left[{}^{h}_{p}\mathbf{m}(00,t_c)\right]^{-1}\left[\mathbf{I}-\exp[-h\ {}^{h}_{p}\mathbf{m}(00,t_c)]\right]\ \mathbf{k}(0,t_c)\,b(t_c)$$

$$={}^{h}_{p}\mathbf{P}(00,t_c)\,\mathbf{k}(0,t_c)\,b(t_c)$$

where ${}^{h}_{p}\mathbf{P}(00,t_c)$ denotes the matrix of transition probabilities between birth during the interval (t_c, t_c+h) and the end of the interval. This equation was first obtained by Namboodiri and Suchindran (1987, p. 145) and van Imhoff (1990, p. 177, equation 16).

Now we consider the sojourn time. The expected time spent in the various states between birth and t_c+h by cohort member k born at $t_c+\tau$, by state occupied at birth, is (provided k experiences the transition rates that apply during the age-period-cohort interval)

$${}^{h-\tau}_{k}\mathbf{L}(0,t_c+\tau)=\int_0^{h-\tau}{}^{h-\tau}_{k}\mathbf{P}(0,t_c+\tau+\xi)\,d\xi=\int_0^{h-\tau}\exp\left[-(h-\tau-\xi)\ {}^{h}_{p}\mathbf{m}(00,t_c)\right]d\xi$$

Integration yields

$\left[{}^{h}_{p}\mathbf{m}(00,t_c)\right]^{-1}\left|\exp[-(h-\tau-\xi)\ {}^{h}_{p}\mathbf{m}(00,t_c)\right|_0^{h-\tau}$ which is equal to

$$\left[{}_p^h\mathbf{m}(00,t_c)\right]^{-1}\left[\mathbf{I}-\exp[-(h-\tau)\;{}_p^h\mathbf{m}(00,t_c)\right]$$

The time spent in the various states irrespective of the state occupied at birth is

$${}_k^{h-\tau}\overline{\mathbf{L}}(0,t_c+\tau)={}_k^{h-\tau}\mathbf{L}(0,t_c+\tau)\;{}_k\mathbf{k}(0,t_c)$$

The sojourn times in the different states between birth and t_c+h by the $b(t_c+\tau)d\tau$ children born between $t_c+\tau$ and $t_c+\tau+d\tau$ is

$${}_p^{h-\tau}\overline{\mathbf{L}}(00,t_c+\tau)={}_p^{h-\tau}\mathbf{L}(00,t_c+\tau)\;\;\mathbf{k}(0,t_c+\tau)b(t_c+\tau)\,d\tau$$

and the sojourn times by all members of the birth cohort, provided the state probabilities at birth are time-independent and the arrival rate of newborns is constant, is

$$\int_0^h {}_p^h\mathbf{K}(0,t_c+\tau)\,d\tau=\left[\int_0^h {}_p^{h-\tau}\mathbf{P}(00,t_c+\tau)\;d\tau\right]\mathbf{k}(0,t_c)b(t_c)$$

$$=\left[\int_0^h \left[{}_p^h\mathbf{m}(00,t_c)\right]^{-1}\left[\mathbf{I}-\exp[-(h-\tau)\;{}_m^h\mathbf{m}(00,t_c)]\right]d\tau\right]\mathbf{k}(0,t_c)b(t_c)$$

$$=\left[{}_p^h\mathbf{m}(00,t_c)\right]^{-1}\left[\mathbf{I}\int_0^h d\tau\;-\int_0^h \exp\left[-(h-\tau)\;{}_m^h\mathbf{m}(00,t_c)\right]d\tau\right]\mathbf{k}(0,t_c)b(t_c)$$

$$=\left[{}_p^h\mathbf{m}(00,t_c)\right]^{-1}\left[h\mathbf{I}-\left|\exp[-(h-\tau)\;{}_m^h\mathbf{m}(00,t_c)]\right|_0^h\right]\mathbf{k}(0,t_c)b(t_c)$$

$$=\left[{}_p^h\mathbf{m}(00,t_c)\right]^{-1}\left[h\mathbf{I}-\;{}_p^h\mathbf{P}(00,t_c)\right]\mathbf{k}(0,t_c)b(t_c)$$

$$=\left[{}_p^h\mathbf{m}(00,t_c)\right]^{-1}\left[\mathbf{k}(0,t_c)[hb(t_c)]-{}_p^h\mathbf{K}(0,t_c+h)\right]$$

This equation is equivalent to equation 16 of van Imhoff (1990, p. 177).

The sojourn time an average child born between t_c and t_c+h may expect to spend in the different states before the end of the interval at t_c+h, by state occupied at birth, will be denoted by ${}_p^h\mathbf{L}(00,t_c)$. Note that the interval between birth and t_c+h is an age-period-cohort interval. The sojourn times spent in the different states during the age-period-cohort interval by the total of all children born during the interval (t_c, t_c+h), by state occupied at birth, are given by the matrix ${}_p^h\mathbf{L}(00,t_c)\mathbf{k}(0,t_c)\;{}^hB(t_c)$.

The assumptions of (1) constant arrival rates of newborns implying a uniform distribution of births and (2) state occupancies at birth independent of time at birth during the interval (t_c, t_c+h) lead to the equation

$${}_p^h\mathbf{L}(00,t_c)=\left[{}_p^h\mathbf{m}(00,t_c)\right]^{-1}\left[h\mathbf{I}-\;{}_p^h\mathbf{P}(00,t_c)\right]\mathbf{k}(0,t_c)b(t_c)$$

b. Population at time t = t_c+x+h

Next we follow the cohort born during the period (t_c, t_c+h) as the members age. At instantaneous time t = t_c+x+h, the members of the cohort are aged x to x+h. Let ${}_p^hK_i(x,t)$ denote the number of individuals aged x to x+h at exact (instantaneous) time t. At time t, those born at t_c are x+h years old and those born at t_c+h are x years old. The *vector of state occupancies* is

$$
{}_p^h\mathbf{K}(x,t) = \begin{bmatrix} {}_p^h K_1(x,t) \\ {}_p^h K_2(x,t) \\ \ldots \\ {}_p^h K_I(x,t) \end{bmatrix}
$$

The relation between the vector of occupancies ${}_p^h\mathbf{K}(x,t)$ and the vector of state probabilties **k**(x,t) is

$$
{}_p^h\mathbf{K}(x,t) = \left[\int_0^h \mathbf{k}(x+\tau,t)\, b(t-x-\tau)\, d\tau \right]
$$

where $b(t\text{-}x\text{-}\tau)$ is the arrival rate of newborns at t-x-τ, $b(t\text{-}x\text{-}\tau)d\tau$ is the number of children born during the interval from t-x-τ-dτ and t-x-τ, and **k**(x+τ,t) is the vector of state probabilities for individuals aged from x+τ to x+τ+dτ at exact time t, and ${}_p^h\mathbf{K}(x,t)$ denotes the distribution among the states of the cohort members aged x to x+h at exact time t. The cohort members aged x to x+h at t are born between t-x-h and t-x. The probability that an individual born at t-x occupies state i at exact age x and exact time t is $k_i(x,t)$. The number of newborns during the interval from t-x-h (t_c) to t-x (t_c+h) is $\int_0^h b(t-x-\tau)\, d\tau$. If the newborns are uniformly distributed during the interval, then the number may be written as ${}^hB(t_c)$. Hence

$$
{}_p^h\mathbf{K}(x,t) = \left[\int_0^h \mathbf{k}(x+\tau,t)\ d\tau \right] {}^hB(t_c)
$$

Assume that the transition rates are piecewise constant in the period-cohort interval. It means that a cohort member who is aged x at exact time t experiences the same transition rates during the interval (t, t+h) as a cohort member who is aged x+τ at exact time t (0 ≤ τ < h). As a consequence, ${}_p^h\mathbf{P}(x+\tau,t) = {}_p^h\mathbf{P}(x,t)$ for (0 ≤ τ < h). The number of cohort members aged x+h to x+2h at exact time t+h, by state occupied at t+h, is

$$
{}_p^h\mathbf{K}(x+h,t+h) = \left[\int_0^h \mathbf{k}(x+h+\tau,t+h)\ d\tau \right] {}^hB(t_c)
$$

$$
{}_p^h\mathbf{K}(x+h,t+h) = \left[\int_0^h {}_p^h\mathbf{P}(x+\tau,t)\mathbf{k}(x+\tau,t)\ d\tau \right] {}^hB(t_c)
$$

$$
{}_p^h\mathbf{K}(x+h,t+h) = {}_p^h\mathbf{P}(x,t)\ {}_p^h\mathbf{K}(x,t)
$$

where ${}_p^h\mathbf{P}(x,t)$ is the matrix of period-cohort transition probabilities. An element ${}_p^h p_{ij}(x,t)$ is the probability that an individual aged x to x+h at time t and occupying state i, occupies state j exactly h years later. The period-cohort transition probability may be visualized using the Lexis diagram. The period-cohort observational plan is represented by the interval VQRS.

For the exponential model, the relation between transition probabilities and the underlying transition rates is

$${}_p^h\mathbf{K}(x+h,t+h)=\exp\left[-h\ {}_p^h\mathbf{m}(x,t)\right]{}_p^h\mathbf{K}(x,t) \text{ and } {}_p^h\mathbf{P}(x,t)=\exp\left[-h\ {}_p^h\mathbf{m}(x,t)\right]$$

where ${}_p^h\mathbf{m}(x,t)$ is the matrix of period-cohort transition rates for individuals aged x to x+h at exact time t.

The age interval and the time interval may be of different length. Let h denote the age interval and Δt the time interval. Individuals aged x to x+h at t are aged x+Δt to x+h+Δt at time t+Δt. Assume that all individuals aged x to x+h at exact time t have the same transition rates during the time interval from t to t+Δt, i.e. for an age interval of length Δt following the age at exact time t. If h > Δt, an individual experiences different transition rates during the same age interval. To introduce intervals of different lengths in the projection model, calendar time may be viewed as a time-varying covariate that changes its value only at discrete points in time. When the value of calendar time changes (after intervals of length Δt) the transition rates change. At all points in time, when the covariate 'calendar time' changes its value, the age interval is split into subintervals. Each subinterval is treated as a different age group. In case of five-year age groups (h = 5) and one-year time intervals (Δt = 1; h =5Δt), the following subintervals may be distinguished: (x, x+Δt), (x+Δt ,x+2Δt), (x+2Δt ,x+3Δt), (x+3Δt ,x+4Δt), (x+4Δt ,x+h). The method is analogous to the method of episode splitting applied in event history analysis to estimate the effect of time-varying covariates on transition rates (see Blossfeld and Rohwer, 2002, pp. 140ff).

The expected sojourn times in the various states during the interval (t, t+h) by an individual of exact age x+ξ (0 ≤ ξ < h) at exact time t are

$${}_p^h\mathbf{L}(x+\xi,t)=\int_0^h {}_p^h\mathbf{P}(x+\xi+\tau,t+\tau)\,d\tau=\int_0^h \exp\left[-\tau\ {}_p^h\mathbf{m}(x,t)\right]d\tau$$

Integration yields

$$\left[{}_p^h\mathbf{m}(x,t)\right]^{-1}\left|\exp[-\tau\ {}_p^h\mathbf{m}(x,t)\right|_0^h \text{ which is equal to}$$

$${}_p^h\mathbf{L}(x+\xi,t)=\left[{}_p^h\mathbf{m}(x,t)\right]^{-1}\left[\mathbf{I}-\exp[-h\ {}_p^h\mathbf{m}(x,t)]\right]=\left[{}_p^h\mathbf{m}(x,t)\right]^{-1}\left[\mathbf{I}-{}_p^h\mathbf{P}(x,t)\right]$$

The expected sojourn times during the interval (t, t+h) is the same for all individuals aged between x and x+h at instantaneous time t. The total number of years spent in the various states between t and t+h by all individuals aged x to x+h at t is ${}_p^h\mathbf{L}(x,t)\ {}_p^h\mathbf{K}(x,t)$.

The sojourn times in the various states irrespective of the state occupied at time t is ${}_p^h\overline{\mathbf{L}}(x,t)={}_p^h\mathbf{L}(x,t)\ {}_p\boldsymbol{\pi}(x,t)$ where ${}_p\boldsymbol{\pi}(x,t)$ is the vector of state probabilities at time t for individuals alive and aged (x,x+h) at t. The sojourn times a child born during the interval (t-x-h,t-x) may expect to spend in the various states between t and t+h, is ${}_p^h\overline{\mathbf{L}}(x,t)={}_p^h\mathbf{L}(x,t)\ {}_p\mathbf{k}(x,t)$ where ${}_p\mathbf{k}(x,t)$ is the vector of state probabilities at time t for member of the birth cohort (t-x-h,t-x).

6. The population

A population consists of different birth cohorts. As members of a birth cohort age, a growing number leaves the populations through death. They are replaced by new cohorts. That demographic metabolism is a vehicle for social change (Ryder, 1965). Mortality and fertility make flexibility and innovation possible and at the same time make stability problematic because 'each new cohort makes fresh contact with the contemporary social heritage and carries the impress of the encounter through life' (Ryder, 1985, p. 11). As a result, new cohorts provide the opportunity for social change to occur. Cohort replacement is a vehicle for change, i.e. it permits change. If change does occur, it differentiates cohorts from one another and the comparison of the careers of successive cohorts is a way to study social change. The successive birth cohorts represent a sequence of overlapping generations. The demographic metabolism that results in the replacement of old cohorts by new cohorts is governed by transition intensities. Transition intensities are the fundamental parameters of demographic change.

As cohort members age, they eventually die but before they do, they transit between functional states. Demographic projection models that distinguish birth cohorts and that consider births, deaths and transitions as the forces that drive demographic change are known as *cohort-component models*. The model presented in this section is a cohort-component model. The model relates the states successive birth cohorts occupy at an instantaneous time t to state occupancies of the same cohorts at a previous point in time and the transitions during the interval. We consider age and time intervals of length h. Since the projection model adopts a period-cohort perspective, this chapter is limited to that perspective, in which calendar time is a continuous variable and age is a discrete variable.

Although most demographic projection models that are in existence today look similar, they differ in several respects. A review of projection models is beyond the scope of this paper. The reader is referred to recent reviews by Booth (2006) and Wilson and Rees (2005). For a review of earlier developments and a discussion of the development path of forecasting methods, see Willekens (1990). A recent introductory text on demographic forecasting from a statistical perspective, in the language of the statistician, is Alho and Spencer (2005). Recent innovations in demographic forecasting models include Girosi and King (2006) who introduce the Bayesian perspective in demographic (mortality) forecasting and show how to incorporate demographic knowledge into a model in a statistically appropriate way. Bijak (2007) introduces Bayesian methods in international migration forecasting. In this paper, we do not adopt a Bayesian perspective but the method can be extended relatively easily to incorporate the Bayesian perspective because the method has strong roots in probability theory.

One of the differences in projection models relates to the distribution of events that occur during discrete age or time intervals. Traditionally it is assumed that the distribution of events is uniform, i.e. that the number of events in each subinterval is the same and is independent of the number of survivors and/or state occupancies at the beginning of the subinterval. The projection model is sometimes referred to as the *linear model* (see e.g. van Imhoff, 1990) although piecewise linear would be a better characterization. In the language of the statistician, the model assumes that the probability density functions of the different transitions are piecewise constant. The linear model emerges when the projection model is derived from demographic accounting equations. The model is presented in Annex B. An alternative assumption is that the distribution of the events that occur during an age or time interval is exponential. That assumption is directly linked to the assumption that the intensity functions of the different transitions are piecewise constant. Note that the intensity is a conditional probability density. It is the density of events, conditional on surviving and/or the state occupancy at the beginning of the subinterval. The model, which is known as the *exponential model*, emerges when the demographic projection model is derived from probability theory with its emphasis on transition intensities as the basic parameters governing change. We consider the exponential version of the projection model. The exponential version is selected because the assumption of piecewise constant intensities it implies is consistent with an assumption frequently made in the implementation of probability models and in statistical analysis of duration data. The linear version was presented by Willekens and Drewe (1984) and has been implemented in MUDEA and LIPRO (Van Imhoff and Keilman, 1991, p 28). The exponential version of the multistate demographic projection model was first derived by Namboodiri and Suchindran (1987, p. 145) and, in more detail, by Van Imhoff (1990). The model is implemented in LIPRO. The linear model is an approximation of the linear model.

The projection model is specified in continuous time, in terms of transition intensities. The events considered are birth, death, interstate transition, emigration and immigration. By considering events that cross the boundary of the population system (so-called external events) the system is an open system. Exits are deaths and emigrations; entries are births and immigrations. Emigration may be replaced by an event that results in a departure beyond the boundaries of the population system and immigration may be an event that implies an arrival from outside the population system. This section consists of four subsections. The first and the second discuss the contribution of immigrants to the population. Section 8.1 addresses immigrant children and Section 8.2 covers other immigrants. Section 8.3 combines the contributions to demographic change of the population present in the country at the beginning of the interval (t,t+h) and of immigrants. Section 8.4 discusses the contribution of children born during the interval (t,t+h).

6.1. Immigrant children (x < h)

Children born abroad during the period (t_c, t_c+h) may immigrate before the end of the interval, i.e. before t_c+h. They contribute therefore to the population at t_c+h. In this section, the contribution is determined. Consider child k born outside of the country being studied. The date of birth $t_b = t_c + \tau$. The child migrates at instantaneous age x. Hence the date of immigration is $t = t_b + x = t_c + \tau + x$. The data of immigration t is also denoted as t_{im}. If the date of birth is not known, but the date of immigration (t) and the exact age at immigration (x) are known, the date of birth can be determined: $t_b = t - x$. The age of k at t_c+h is $h-(t_b- t_c) = h - \tau$ and the duration of residence is $h-(t- t_c) = h - \tau - x$, provided k does not emigrate during the interval. These time measures will be used in this section.

A newborn immigrant during the interval from t_c to t_c+h is defined as an immigrant whose time of birth and time of immigration are in the interval. Consider a newborn immigrant born at $t_{c+}\tau$ and aged ξ at the time of immigration. The calendar time at immigration is $t = t_c+\tau+x$. Let ${}_kY^{im}(x,t)$ be an indicator variable that is one if at instantaneous time t child k, who is born abroad and currently of instantaneous age x, immigrates. The number of immigrants during the infinitesimal interval (t, t+dt) of the same age x is $N^{im}(x,t)\,dt = \sum_k^{Q_{im}} {}_kY^{im}(x,t)\,dt$ where k denotes an individual and Q_{im} is the total number of children born abroad at instantaneous time t-x. The obtain the expected value of the number of immigrants, the indicator variable ${}_kY^{im}(x,t)$ and the number of immigrants are not related to the number of children born in the rest of the world (Q_{im}). Instead, the number of immigrants is expressed per unit of time. The expected value of the number of immigrants **per unit time interval** is the *arrival rate of immigrants* $i_{im}(x+\tau,t+\tau) = E\left[N^{im}(x+\tau,t+\tau)\right]/\,d\tau$. Note that $i_{im}()$ does not depend on the number of children born in the rest of the world that constitutes in fact the population at risk of migrating to the country being studied. The arrival rate is determined exogenously, comparable to the arrival rate of newborns in the previous section.

In this section, we consider the immigration during a time interval of length h of children born during the same interval. These children are less than h years of age at the time of immigration. The arrival rate of newborn immigrants of instantaneous age x at instantaneous time t is $i_{im}(x,t)$ where $t = t_b+x$ with $t_c \le t_b < t_c+h$ and $t_b \le t < t_c+h$. A newborn child that migrates at instantaneous age x may migrate at instantaneous times from t_c+x to t_c+h. A child that migrates at t_c+x is born at t_c. A child that migrates at age x at t_c+h is born at t_c+h-x (=t_b). Let $i_{im}(x,t)$ denote the arrival rate of immigrants of instantaneous age x at instantaneous time t, where $t = t_b+x = t_c+\tau+x$. The number of immigrant children aged x to x+dx during the infinitesimally small interval from $t_c+\tau+x$ to $t_c+\tau+x+d\tau$ is $i_{im}(x,t_c+\tau+x)\,d\tau\,dx$. The number of newborn immigrants during the interval (t_c,t_c+h) is ${}_p^hI_{im}(00,t_c) = \int_0^h \int_0^{h-x} i_{im}(\xi,t_c+x+\tau)\,d\tau\,dx$. 00 denotes that the children are born during the interval (t_c,t_c+h) and p refers to the period-cohort observation plan (age-period-cohort in this case)

The constant *age-specific* arrival rate of immigrants implies that the number of immigrants during the interval ($t_c+\tau+x$, $t_c+\tau+x+d\tau$) increases with the maximum age of immigrants. It implies that the number of immigrants, aggregated over age, is not uniformly distributed over the interval (t_c,t_c+h). The number of immigrants *of a given age* are uniformly distributed, but the age variation of immigrant children is larger near the end of the interval than at the beginning of the interval. Note that van Imhoff (1990, p.

176) and van Imhoff and Keilman (1991, p. 25) assume that newborn immigrants are uniformly distributed. On the other hand, it is assumed that the newborn immigrants are born between the start of the projection interval and the moment of immigration (van Imhoff and Keilman, 1991, p. 27). The combination of the two assumptions requires that the age structure of newborn immigrants changes considerably from t_c to t_c+h, which is not realistic.

At time of immigration, the newborn immigrant occupies one of the states of the state space. The vector of state probabilities at time of immigration is $\mathbf{k}_{im}(x,t_c+x+\tau)$. The number of newborn immigrants aged x to x+dx during the interval (t_c+τ+x, t_c+τ+x+dτ), by state occupied at immigration, is $\mathbf{k}_{im}(x,t_c+x+\tau)\, i_{im}(x,t_c+x+\tau)\, d\tau\, dx$. The probability that a child that was born at time $t_b = t_c+\tau$ during the period (t_c,t_c+h) and migrated at time t = t_c+x+τ at age x survives to t_c+h, by state occupied at immigration and state at t_c+h, is ${}^{h-\tau-x}\mathbf{P}_{im}(x,t_c+x+\tau)=\exp\left[-\int_0^{h-\tau-x}\mu_{im}(x+\theta,t_c+x+\tau+\theta)\,d\theta\right]$, where the transition intensities are defined for immigrant children. The age of the child at t_c+h is h-(t-t_c) = h-x-τ. The expected number of children at t_c+h by state occupied is

$${}^{h}_{p}\mathbf{K}_{im}(0,t_c+h)=\int_0^h\int_0^{h-\xi} {}^{h-\tau-x}\mathbf{P}_{im}(x,t_c+x+\tau)\ \mathbf{k}_{im}(x,t_c+x+\tau)\, i_{im}(x,t_c+x+\tau)\, d\tau\, dx$$

If the arrival rates of newborn immigrants is constant during the interval (t_c,t_c+h) and the state occupied at immigration is independent of the date of migration, then

$${}^{h}_{p}\mathbf{K}_{im}(0,t_c+h)=\left[\int_0^h\int_0^{h-x} {}^{h-\tau-x}\mathbf{P}_{im}(x,t_c+x+\tau)\ \ d\tau\, dx\right]\mathbf{k}_{im}(00,t_c)\ {}^{h}_{p}I_{im}(00,t_c)$$

where $\mathbf{k}_{im}(00,t_c)$ denotes the vector of state occupancies at immigration for children born and migrating in the interval (t_c,t_c+h).

If the transition rates are constant in the age-period-cohort interval[12], with
$\boldsymbol{\mu}_{im}(x,t_c+x+\tau)={}^{h}_{p}\mathbf{m}_{im}(00,t_c)$ **for $0 \le \tau < h$ and $0 \le x < \tau$, then**

$${}^{h-\tau-x}\mathbf{P}_{im}(x,t_c+x+\tau)=\exp[-(h-\tau-x)\ {}^{h}_{p}\mathbf{m}_{im}(00,t_c)]$$

and

$${}^{h}_{p}\mathbf{K}_{im}(0,t_c+h)=\left[\int_0^h\int_0^{h-x} \exp[-(h-\tau-x)\ {}^{h}_{p}\mathbf{m}_{im}(00,t_c)\, d\tau\, dx\right]\mathbf{k}_{im}(00,t_c)\ {}^{h}_{p}I_{im}(00,t_c)$$

Note that

$$\int_0^{h-x}\exp[-(h-\tau-x)\ {}^{h}_{p}\mathbf{m}_{im}(00,t_c)\, d\tau=\left[{}^{h}_{p}\mathbf{m}(00,t_c)\right]^{-1}\left|\exp[-(h-\tau-x)\ {}^{h}_{p}\mathbf{m}(00,t_c)\right|_0^{h-x}$$

$$=\left[{}^{h}_{p}\mathbf{m}_{im}(00,t_c)\right]^{-1}\left[\mathbf{I}-\exp[-(h-x)\ {}^{h}_{p}\mathbf{m}_{im}(00,t_c)\right]$$

Hence

[12] The occurrence of demographic events in an age-period-cohort observation plan was also studied by Calot and Franco (2002). They assume that $\mu(x,t_b+x)$ varies linearly with x (see p. 39).

$$ {}_p^h\mathbf{K}_{im}(0,t_c+h) = \left[\int_0^h \left[{}_p^h\mathbf{m}_{im}(00,t_c)\right]^{-1}\left[\mathbf{I}-\exp[-(h-x)\ {}_p^h\mathbf{m}_{im}(00,t_c)]\right]dx\right]\mathbf{k}_{im}(00,t_c)\ {}_p^hI_{im}(00,t_c) $$

where $\exp[-(h-x)\ {}_p^h\mathbf{m}(00,t_c)] = {}_p^{h-x}\mathbf{P}_{im}(x,t_c)$ is the probability that on average, a child that immigrates during the interval (t_c,t_c+h) at age x, survives till t_c+h, by state occupied at immigration and at t_c+h. The average is over all dates of birth. An element ${}_p^{h-x}p_{im\ ij}(x,t_c)$ of ${}_p^{h-x}\mathbf{P}_{im}(x,t_c)$ is the average probability that a child born during the interval (t_c, t_c+h) that immigrates at age x (<h) and occupies state i at immigration, survives to the end of the interval and occupies state j at that time. The solution of the integral is

$$ {}_p^h\mathbf{K}_{im}(0,t_c+h) = \left[{}_p^h\mathbf{m}_{im}(00,t_c)\right]^{-1}\left[h\,\mathbf{I}-\left[{}_p^h\mathbf{m}_{im}(00,t_c)\right]^{-1}\left[\mathbf{I}-\exp(-h\,{}_p^h\mathbf{m}_{im}(00,t_c))\right]\right]\mathbf{k}_{im}(00,t_c)\ {}_p^hI_{im}(00,t_c) $$

It is the number of persons that the immigrants, born during the period (t_c,t_c+h), contribute to the population aged 0 to h at time t_c+h if the arrival rate of immigrants of a given age is constant (immigrants of a given age are uniformly distributed over the interval), the state occupancy at immigration is independent of age and date of immigration, and all children born and migrating during (t_c,t_c+h) experience the same transition rates.

A child that is born at time $t_b = t_c+\tau$ during the period (t_c,t_c+h) and migrates at time $t = t_c+x+\tau$ and age x may expect to live a maximum of $h-x-\tau$ years before the end of the interval. The expected sojourn time in each of the states is

$$ {}_p^{h-\tau-x}\mathbf{L}_{im}(x,t_c+x+\tau) = \int_0^{h-\tau-x} {}_p^{h-\tau-x}\mathbf{P}_{im}(x+\theta,t_c+x+\tau+\theta)\,d\theta $$

and the expected sojourn time by a newborn immigrant irrespective of the state occupied at immigration is

$$ {}_p^{h-\tau-x}\overline{\mathbf{L}}_{im}(x,t_c+x+\tau) = \int_0^{h-\tau-x} {}_p^{h-\tau-x}\mathbf{P}_{im}(x+\theta,t_c+x+\tau+\theta)\,\mathbf{k}_{im}(x+\theta,t_c+x+\tau+\theta)\,d\theta $$

If the state occupancy at immigration does not depend on the time of immigration and the age at immigration, then

$$ {}_p^{h-\tau-x}\overline{\mathbf{L}}_{im}(x,t_c+x+\tau) = {}_p^{h-\tau-x}\mathbf{L}_{im}(x,t_c+x+\tau)\,\mathbf{k}_{im}(00,t_c) $$

where an element $k_{im\ i}(00,t_c)$ of $\mathbf{k}_{im}(00,t_c)$ denotes the probability that a child that is born and immigrates during the interval (t_c,t_c+h) occupies state i at immigration. The state occupancy does not depend on the date of birth and the age at immigration as long as the two events occur during the time interval (t_c,t_c+h).

If the transition intensities are constant during the age-period-cohort interval, then

$$ {}_p^{h-\tau-x}\mathbf{L}_{im}(x,t_c+x+\tau) = \int_0^{h-\tau-x} \exp[-(h-\tau-x-\theta)\ {}_p^h\mathbf{m}_{im}(00,t_c)]\,d\theta $$

since ${}^{h-\tau-x}\mathbf{P}_{im}(x+\theta,t_c+x+\tau+\theta) = \exp[-(h-\tau-x-\theta)\ {}_p^h\mathbf{m}_{im}(00,t_c)]$

$$ {}^{h-\tau-x}_{p}\mathbf{L}_{im}(x,t_c+x+\tau) = \left[{}^{h}_{p}\mathbf{m}_{im}(00,t_c)\right]^{-1}\left[\mathbf{I}-\exp[-(h-\tau-x)\ {}^{h}_{p}\mathbf{m}_{im}(00,t_c)\right] $$

6.2. Immigrants aged x ≥ h

Recall that ${}^{h}_{p}\mathbf{K}(x,t)$ is the number of individuals born during the interval (t-x-h, t-x) and aged x to x+h at instantaneous time t, by state occupied at t. The arrival rate of immigrants of age x at time t is $i_{im}(x,t)$. The total number of immigrants during the period (t, t+h) who were born during the interval (t_c, t_c+h) with t_c = t-x-h is

${}^{h}_{p}I_{im}(x,t) = \int_0^h\int_\tau^{\tau+h} i_{im}(x+\xi,t+\tau)\,d\xi\,d\tau$ The contribution of immigrants during the period (t,t+h) to the population at the end of the interval is

$${}^{h}_{p}\mathbf{K}_{im}(x+h,t+h) = {}^{h}_{p}\mathbf{P}_{im}(x,t)\ {}^{h}_{p}I_{im}(x,t)$$

where ${}^{h}_{p}\mathbf{P}_{im}(x,t)$ is the matrix of period-cohort transition probabilities. An element ${}^{h}_{p}p_{im\,ij}(x,t)$ is the probability that an immigrant during the interval (t, t+h) who is aged x to x+h at time t (i.e. is born during the interval from t-x-h to t-x) and occupies state i at immigration, occupies state j exactly at t+h.

In the transition intensities are piecewise constant during the period-cohort interval, i.e. if all immigrants born during the interval (t-x-h,t-x) and immigrating during the interval (t,t+h) experience the same transition intensities and if the transition intensities are constant in the period-cohort interval defined by the time interval (t, t+h) and the birth cohort (t-x-h,t-x), then

$$\begin{aligned}
{}^{h}_{p}\mathbf{K}_{im}(x+h,t+h) &= \int_0^h \exp[-(h-\tau)\ {}^{h}_{p}\mathbf{m}_{im}(x,t)]\mathbf{k}(x,t)\ {}^{h}_{p}I_{im}(x,t)\,d\tau \\
&= \left[{}^{h}_{p}\mathbf{m}_{im}(x,t)\right]^{-1}\left[\mathbf{I}-\exp[-h\ {}^{h}_{p}\mathbf{m}_{im}(x,t)]\right]\ \mathbf{k}(x,t)\ {}^{h}_{p}I_{im}(x,t) \\
&= {}^{h}_{p}\mathbf{P}_{im}(x,t)\mathbf{k}(x,t)\ {}^{h}_{p}I_{im}(x,t)
\end{aligned}$$

where ${}^{h}_{p}\mathbf{m}_{im}(x,t)$ is the matrix of period-cohort transition rates for immigrants born during the interval (t-x-h,t-x) and migrating during the interval (t,t+h). An element ${}^{h}_{p}m_{im\,ij}(x,t)$ of ${}^{h}_{p}\mathbf{m}_{im}(x,t)$ is a ***period-cohort transition rate*** (denoted by the subscript p) that applies to recent immigrants.

The arrival rate at instantaneous time t of immigrants of instantaneous age x+ξ is $i_{im}(x+\xi,t)$. The number of immigrants of exact age x during the period (t, t+h) is $\int_0^h i_{im}(x,t+\tau)\,d\tau$ and the number of immigrants at exact time t aged x to x+h, is $\int_0^h i_{im}(x+\xi,t)\,d\xi$. Note that at time t+τ, immigrants born of the birth cohort (t-x-h,t-x) are between x+τ and x+τ+h years old. The number of immigrants during the interval (t,t+h) who are born between t-x-h and t-x is

${}_p^h I_{im}(x,t) = \int_0^h \left[\int_0^h i_{im}(x+\xi+\tau, t+\tau)\, d\xi \right] d\tau$ which may also be written as

${}_p^h I_{im}(x,t) = \int_0^h \int_\tau^{\tau+h} i_{im}(x+\xi, t+\tau)\, d\xi\, d\tau$ (see above).

The total number of immigrants in the cohort born between t-x-h and t-x, irrespective of their survival status and state occupied, is

${}_p^h I_{im}^0(x,t) = \int_0^h \left[\int_0^\omega i_{im}(\xi+\tau, t_c+\tau)\, d\xi \right] d\tau$

where ω is the highest age. Let $\mathbf{k}_{im}(x+\xi+\tau,t+\tau)$ denote the state vector for immigrants of instantaneous age x+ξ+τ at instantaneous time t+τ. An element $k_{im\,i}(x+\xi+\tau,t+\tau)$ denotes the probability that an immigrant who is age x+ξ+τ at time of immigration t+τ, occupies state i at the time of immigration. The arrival rate of immigrants by state occupied upon immigration is $\mathbf{k}_{im}(x+\xi+\tau, t+\tau)\; i_{im}(x+\xi+\tau, t+\tau)$. If the destination states of immigrants do not differ among migrants born in the same period who enter the country during a same period, then the state occupancies of migrants aged x to x+h during the interval t and t+h is $\mathbf{k}_{im}(x,t)\; {}_p^h I_{im}^0(x,t)$.

6.3. Population at t+h

The population at t+h depends on the population at time t and immigration during the interval (t,t+h). The vector ${}_p^h \mathbf{I}_{im}(x,t)$ denotes the number of immigrants during the unit interval aged x to x+h at t by state occupied at immigration. These migrants are born during the interval from t-x-h to t-x. The population aged x+h,x+2h at time t+h is

$${}_p^h \mathbf{K}(x+h, t+h) = \exp\left[-h\; {}_p^h \mathbf{m}(x,t)\right] {}_p^h \mathbf{K}(x,t) + \int_0^h \exp\left[-(h-\tau)\; {}_p^h \mathbf{m}(x,t)\, d\tau\right] {}_p^h \mathbf{I}_{im}(x,t)$$

where the transition rates are the same for natives and immigrants. The first term is the contribution to the population at t+h by the population at time t and the second term is the contribution of immigrants during the interval (t,t+h).

The equation may be written as

$$\begin{aligned} {}_p^h \mathbf{K}(x+h, t+h) \\ &= \exp\left[-h\; {}_p^h \mathbf{m}(x,t)\right] {}_p^h \mathbf{K}(x,t) + \left[{}_p^h \mathbf{m}(x,t) \right]^{-1} \left[\mathbf{I} - \exp[-h\; {}_p^h \mathbf{m}(x,t)] \right] {}_p^h \mathbf{I}_{im}(x,t) \\ &= {}_p^h \mathbf{P}(x,t)\, {}_p^h \mathbf{K}(x,t) + {}_p^h \mathbf{F}(x,t)\; {}_p^h \mathbf{I}_{im}(x,t) \end{aligned}$$

${}_p^h \mathbf{P}(x,t) = \exp\left[-h\; {}_p^h \mathbf{m}(x,t)\right]$ represents the contribution of cohort members aged x to x+h present at time t to the population aged x+h to x+2h at t+h, and

${}_{p}^{h}\mathbf{F}(x,t) = \left[{}_{p}^{h}\mathbf{m}(x,t)\right]^{-1}\left[\mathbf{I} - \exp[-h\ {}_{p}^{h}\mathbf{m}(x,t)]\right]$ represents the contribution of persons who immigrated during the interval from t to t+h and who are aged x to x+h at time t, to the population at t+h.

The exponential model of multistate population growth is derived from a system of heterogeneous differential equations

$$\frac{d\,\mathbf{K}(x+\tau,t+\tau)}{d\,\tau} = -\boldsymbol{\mu}(x+\tau,t+\tau)\,\mathbf{K}(x,t) + \mathbf{k}(x+\tau,t+\tau)\,i_{im}(x+\tau,t+\tau)$$

where $\mathbf{K}(x,t)$ is a vector of state occupancies for individuals of age x at time t and $i_m(x,t)$ is the arrival rate at time t of immigrants of exact age x. The vector $\mathbf{k}(x,t)$ is defined above. In open systems, the diagonal elements of the matrix of transition intensities $\mu(x,t)$ include the instantaneous rates of transition to outside the population system:

$$\mu_{ii}(x,t) = \mu_{id}(x,t) + \sum_{j \neq i}^{I} \mu_{ij}(x,t) + \mu_{io}(x,t)$$

where $\mu_{id}(x,t)$ is the instantaneous death rate in state i for individuals aged x and born at t-x and $\mu_{io}(x,t)$ the instantaneous rate at which individuals aged x in state i leave the population system during the small interval from t to t+dτ (emigration rate).

The solution to the above system of differential equations is

$${}_{p}^{h}\mathbf{K}(x+h,t+h) = \exp\left[-h\ {}_{p}^{h}\mathbf{m}(x,t)\right]{}_{p}^{h}\mathbf{K}(x,t) + \left[\int_{0}^{h}\exp[-(h-\tau)\ {}_{p}^{h}\mathbf{m}(x,t)]\,d\tau\right]{}_{p}^{h}\mathbf{I}_{im}(x,t)$$

(see e.g. Director and Rohrer, 1972, p. 431; Van de Vegte, 1994, p. 331)

6.4. Births and first age group

Consider the birth cohort (t,t+h). The children born during that interval are aged less than h at time t+h. Some children were born in the country and some were born abroad and immigrated. Children born in the country are referred to as native-born children. Children born abroad are foreign-born. Newborns and immigrant children may change attributes and they emigrate or die before reaching the end of the time interval, i.e. t+h.

a. Native-born children

The number of births during the interval (t,t+h), ${}_{p}^{h}B(x,t)$, depends on the number of women, their age structure and the age-specific rates of birth. Consider a woman born at instantaneous time t_c = t-x. At time t she is exactly x years old. Let $\mu_{ib}(x,t)$ denote the instantaneous rate at which the woman gives birth to a child provided she occupies state i at age x at time t. The rate applies to the small interval from x to x+dx.

The matrix of instantaneous fertility rates is a diagonal matrix if children are born in the state occupied by the woman at time of birth. If newborns may occupy states different from the state occupied by the mother, the off-diagonal elements of $\mu_b(x,t_c+x)$ denote the state occupied by the newborn by state of the mother at time of birth. For instance, if the states are marital states, then the child born to a married woman occupies the never-married state. The product $\boldsymbol{\mu}_b(x,t_c+x)$ $\mathbf{k}(x,t_c+x)$ is the *maternity function* for the woman born at time t_c and currently of age x, by state occupied at x and state at birth of the child.

The number of children born during the period (t,t+h) to women born themselves between $t_c = t - x - h$ and $t_c + h = t - x$ is:

$$_p^hB(x,t) = \mathbf{1}'\left[\int_0^h\int_0^h \boldsymbol{\mu}_b(x+\xi+\tau,t+\tau)\ {}_f\mathbf{k}(x+\xi+\tau,t+\tau)\ {}_fb(t-x-\tau)\,d\tau\,d\xi\right]$$

where $\mathbf{1}'$ is a row vector of ones, the subscript f refers to the female population, ${}_fb(t\text{-}x\text{-}\tau)$ is the arrival rate of girls at t-x-τ (the potential mothers), ${}_fb(t\text{-}x\text{-}\tau)d\tau$ is the number of girls born during the interval from t-x-τ-dτ and t-x-τ, and ${}_f\mathbf{k}(x+\xi+\tau,t)$ is the vector of state probabilities for women aged from x+ξ+τ to x+ξ+τ+ dξ at exact time t. The interval is a period-cohort observation plan, which in the Lexis diagram is visualized by the parallelogram VQRS. The women aged x+ξ to x+ξ+dξ at time t may expect to have $^hB(x+\xi,t) = \mathbf{1}'\left[\int_0^h \boldsymbol{\mu}_b(x+\xi+\tau,t+\tau)\ {}_f\mathbf{k}(x+\xi+\tau,t+\tau)\ {}_fb(t-x-\tau)\,d\tau\right]d\xi$ children between t and t+h.

Let α denote the lowest age at childbearing and β the highest age. The total number of children born during the interval (t,t+h) to women aged between α-h and β at t (i.e. born between t-β and t+h-α) is

$$_p^hB(t) = \mathbf{1}'\left[\int_\alpha^\beta\int_0^h \boldsymbol{\mu}_b(x+\tau,t+\tau)\ {}_f\mathbf{k}(x+\tau,t+\tau)\ {}_fb(t-x-\tau)\,d\tau\,dx\right]$$

where $_p^hB(t)$ denotes the total number of births during the period (t,t+h). It is a period-cohort measure, denoted by the subscript p.

The vector aggregates the numbers of births by state. It is a period measure of fertility. Compare this with the cohort measure, which is the number of children born throughout the reproductive period to the birth cohort (t_c,t_c+h)

$$_c^hB(t_c) = \mathbf{1}'\left[\int_0^h\left[\int_\alpha^\beta \boldsymbol{\mu}_b(x,t_c+x+\tau)\ {}_f\mathbf{k}(x,t_c+x+\tau)\ dx\right]{}_fb(t_c+\tau)\ d\tau\right]$$

where $_c^hB(t_c)$ denotes the total number of births to women born during the interval (t_c, t_c+h). It is an age-cohort measure, denoted by the subscript c.

The quantity $\mathbf{1}'\left[\int_\alpha^\beta \boldsymbol{\mu}_b(x,t_c+x+\tau)\ {}_f\mathbf{k}(x,t_c+x+\tau)\,dx\ \right]{}_fb(t_c+\tau)\,d\tau$ is the number of children born throughout the reproductive life to the ${}_fb(t_c+\tau)\,d\tau$ women born during the period (t_c+τ, t_c+τ+dτ). Note that a girl born at t_c+τ reaches her x-th birthday at time

t_c+x+τ. For a woman who is born at instantaneous time $t_c+\tau$, the cumulative fertility ($_k$CF) is:

$$CF = \mathbf{1}'\left[\int_{\alpha}^{\beta} \boldsymbol{\mu}_b(x, t_c + x + \tau)\,\mathbf{k}(x, t_c + x + \tau)\,dx\right]$$

$$= \sum_{i=1}^{I}\int_{\alpha}^{\beta} \mu_{ib}(x, t_c + x + \tau)\,k_i(x, t_c + x + \tau)\,dx$$

where $k_i(x,t_c+x)$ is the probability that a woman born at t_c occupies state i at exact age x. α is the youngest age of childbearing and β the oldest.

Suppose all women born during the period from t-x-h to t-x, and hence aged x to x+h at time t, experience the same state-specific instantaneous fertility rates $\boldsymbol{\mu}_b(x+\xi, t+\xi)$. If in addition the rates remain constant during the interval (t,t+h), then $\boldsymbol{\mu}_b(x+\xi, t+\xi) = {}_p^h\mathbf{m}_b(x,t)$ for all ξ satisfying 0 ≤ ξ <h and for all k. The fertility rates ${}_p^h\mathbf{m}_b(x,t)$ are period-cohort rates. If in addition the births during the period (t_c,t_c+h) are uniformly distributed, if the state probabilities at time t are the same for women aged (x,x+h) and remain the same throughout the interval (t,t+h), then the number of newborns during the interval (t,t+h) to women aged (x,x+h) at t is

$${}_p^hB(x,t) = \mathbf{1}'\int_0^h {}_p^h\mathbf{m}_b(x,t)\,{}_{pf}^h\mathbf{K}(x+\tau, t+\tau)\,d\tau = \mathbf{1}'\left[\left[{}_p^h\mathbf{m}_b(x,t)\int_0^h {}_{pf}^h\mathbf{P}(x+\tau, t+\tau)\,d\tau\right]{}_{pf}\mathbf{k}(x,t)\,h_f\,b(t-x)\right]$$

$${}_p^hB(x,t) = \mathbf{1}'\left[{}_p^h\mathbf{m}_b(x,t)\ {}_{pf}^h\mathbf{L}(x,t)\,{}_{pf}\mathbf{k}(x,t)\,h_f\,b(t-x)\right]$$

where

$${}_{pf}^h\mathbf{L}(x,t) = \int_0^h {}_{pf}^h\mathbf{P}(x+\tau, t+\tau)\,d\tau = \int_0^h \exp[-\tau\ {}_p^h\mathbf{m}(x,t)]\,d\tau = \left[{}_{pf}^h\mathbf{m}(x,t)\right]^{-1}\left[\mathbf{I} - \exp[-h\ {}_{pf}^h\mathbf{m}(x,t)]\right]$$

Hence

$${}_p^hB(x,t) = \mathbf{1}'\left[{}_p^h\mathbf{m}_b(x,t)\left[{}_{pf}^h\mathbf{m}(x,t)\right]^{-1}\left[\mathbf{I} - \exp[-h\ {}_{pf}^h\mathbf{m}(x,t)]\right]{}_{pf}\mathbf{k}(x,t)\,h_f\,b(t-x)\right]$$

${}_{pf}^h\mathbf{L}(x,t)$ is the number of time units, years say, a woman aged (x,x+h) at time t is exposured to the risk of childbearing during the period (t, t+h) by state occupied at **t.** ${}_{pf}^h\mathbf{L}(x,t)\ {}_f\mathbf{k}(x,t)$ is the exposure time or sojourn time irrespective of the state occupied at time t. Since ${}_{pf}^h\mathbf{K}(x,t) = {}_{pf}\mathbf{k}(x,t)\ h\ {}_f b(t-x)$

$${}_p^hB(x,t) = \mathbf{1}'\left[{}_p^h\mathbf{m}_b(x,t)\ {}_{fp}^h\mathbf{L}(x,t)\ {}_{pf}^h\mathbf{K}(x,t)\right] = \mathbf{1}'\left[{}_p^h\mathbf{m}_b(x,t)\left[{}_{pf}^h\mathbf{m}(x,t)\right]^{-1}\left[\mathbf{I} - \exp[-h\ {}_{pf}^h\mathbf{m}(x,t)]\right]{}_{pf}^h\mathbf{K}(x,t)\right]$$

The number of children born between t and t+h to an average woman aged x to x+h at t, by state occupied at t and state occupied at birth, is $\mathbf{m}_b(x,x+h,t)\ {}_f\mathbf{L}(x,x+h,t)$.

Note that the number of children born during the interval (t,t+h) to a woman aged (x,x+h) at t, by state occupied at time of birth, irrespective of the state occupied at t, is ${}_p^h\mathbf{m}_b(x,t)\ {}_{pf}^h\overline{\mathbf{L}}(x,t)$.

The total number of children born during the interval (t,t+h) is

$${}_p^h B(\alpha,\beta,t) = \sum_{x=\alpha}^{\beta} {}_p^h B(x,t)$$

The number of children per unit interval is the arrival rate of children. It is $b(t) = {}_p^h B(\alpha,\beta,t)$.

The contribution of children born in the country during the interval (t,t+h) to the population at t+h is (see Section 5):

$$\begin{aligned} {}_p^h\mathbf{K}(0,t+h) &= \left[{}_p^h\mathbf{m}(00,t)\right]^{-1}\left[\mathbf{I} - \exp\left[-h\ {}_p^h\mathbf{m}(00,t)\right]\right]\ \mathbf{k}(0,t)\,h\ b(t) \\ &= {}_p^h\mathbf{P}(00,t)\,\mathbf{k}(0,t)\,h\ b(t) \end{aligned}$$

where b(t) is the arrival rate of newborns, i.e. the number of newborns during the period (t,t+h) **per unit time interval**, and ${}_p^h\mathbf{P}(00,t)$ denotes the matrix of transition probabilities between birth during the interval (t, t+h) and the end of the interval.

b. Foreign-born children

The contribution of immigrant children is (see above)

$${}_p^h\mathbf{K}_{im}(0,t_c+h) = \left[{}_p^h\mathbf{m}_{im}(00,t_c)\right]^{-1}\left\lfloor h\,\mathbf{I} - \left[{}_p^h\mathbf{m}_{im}(00,t_c)\right]^{-1}\left[\mathbf{I} - \exp(-h\,{}_p^h\mathbf{m}_{im}(00,t_c)\right]\right\rfloor \mathbf{k}_{im}(00,t_c)\ {}_p^h I_{im}(00,t_c)$$

The combined contribution of newborns and immigrant children is

$${}_p^h\mathbf{K}(0,t+h) = \left[{}_p^h\mathbf{m}(00,t)\right]^{-1}\left[\mathbf{I} - \exp\left[-h\ {}_p^h\mathbf{m}(00,t)\right]\right]\ \left[\mathbf{k}(00,t)\,b(t) + \mathbf{k}_{im}(00,t)\ {}_p^h I_{im}(00,t)\right]$$

where it is assumed that newborns and immigrant children experience the same transition rates during the interval (t,t+h).

7. Multistate models with covariates

7.1. Introduction

The fundamental parameters of multistate models are the transition intensities (continuous time) and transition rates (discrete time). The transition intensity ${}_k\mu_{ij}(x,t)$ is the rate at which individual k who is born at t_c=t-x and who occupies state i at instantaneous age x at time t experiences a transition to state j during the infinitesimally small interval from x to x+dx. In Sections 5 and 6, it was assumed that individuals born at the same time experience the same age-specific transition intensities: ${}_k\mu_{ij}(x,t) = \mu_{ij}(x,t)$ for all individuals k born at exact time t_c=t-x or during the interval t. Individuals born at $t_c+\tau$ with $0 \le \tau < h$ belong to the same birth cohort. They celebrate their x-th birthday between time t = t_c + x and t +h = t_c + x + h. It corresponds to the interval VW on the Lexis diagram. At instantaneous time t, they are between age x-h and x year old and at t+h, they are aged x to x+h. It is generally assumed that members of the same birth cohort experience the same transition intensities. In cohort analysis, the transition intensities at instantaneous age x are assumed to apply to all individuals born between t_c and t_c+h: ${}_k\mu_{ij}(x, t_c+x+\tau) = \mu_{ij}(x, t)$ for $0 \le \tau < h$ and with $t = t_c+x$. In period analysis and demographic projections, the transition intensities at instantaneous time t are assumed to apply to all individuals aged between x and x+h, i.e. to all individuals born between t-x-h and t-x: ${}_k\mu_{ij}(x+\tau, t) = \mu_{ij}(x, t)$ for $0 \le \tau < h$ where $t = t_c+x+h$ (an individual born at instantaneous time $t_c = t - x - h$ is x+h years old at time t). It corresponds to the interval VQ on the Lexis diagram.

For practical purposes, it is further assumed that the transition intensities are constant during particular intervals of age and/or time. In cohort analysis, it is generally assumed that members of the same birth cohort from t_c to t_c+h and of the same age x experience the same transition intensities and that the transition intensities remain constant between two consecutive birthdays. Let t_c be t-x-h. The birth cohort consists of children born between t-x-h and t-x. They reach their x-th birthday between t-h and t and the x+h-birthday between t and t+h. Ifhe transition intensities are constant between these two birthdays, then $\mu(x+\xi, t-h+\xi+\tau) = {}_c^h m_{ij}(x, t-h)$ for $0 \le \tau < h$ and $0 \le \xi < h$. The observation interval is the age-cohort interval. The interval corresponds to PVSQ on the Lexis diagram. In period analysis and demographic projections, it is generally assumed that all members of the birth cohort from t_c to t_c+h and currently at time t experience the same transition intensities and that the transition intensities remain constant for h years. Consider the birth cohort from t_c = t–x-h to t_c+h = t-x. They are between x and x+h at time t. Constant transition intensities during the interval imply that $\mu(x+\xi+\tau, t+\tau) = {}_p^h m_{ij}(x, t)$ for $0 \le \xi < h$ and $0 \le \tau < h$. The observation interval is the period-cohort interval. The interval corresponds to QVSR on the Lexis diagram.

Transition intensities and transition rates may be converted into transition probabilities ${}_k^h\mathbf{P}(x,t)$, ${}_c^h\mathbf{P}(x,t)$ and ${}_p^h\mathbf{P}(x,t)$. An element ${}_k^h p_{ij}(x,t)$ of ${}_k^h\mathbf{P}(x,t)$ denotes the probability that individual k, who is born at t-x and currently at age x in state i, is in state j at age x+h.

Transition intensities, transition rates and transition probabilities vary with age and in time. They also vary between members of a same cohort. Members are distinguished on the basis of observed characteristics, known as covariates. The models that link transition rates or probabilities to covariates are regression models with the transition rates or probabilities as the dependent variable to be predicted by the covariates (predictors). *Transition rate models and transition probability models are fundamentally different.* The difference is related to the different constraints imposed on rates and probabilities. Link functions are used to operationalize the constraints. Transition rates are constrained to be nonnegative. They may be larger than one. To assure a nonnegative transition rate, the logarithmic transformation is used. The regression model that results is a log-linear model, i.e. the log transformation of the rates is linear in the parameters (regression coefficients). In the 1980s, Tuma and Hannan (1984) and others introduced transition models in the social sciences. For a more recent treatment, see Blossfeld and Rohwer (2002). Transition probabilities are constrained to have a value between zero and one. To assure a value between zero and one, whatever the values of the covariates, a cumulative distribution is used that transforms a real number between $-\infty$ and $+\infty$ into a figure between zero and one. In social sciences the logistic distribution is often used and the link function is the logit. The logit transformation of the probability is linear in the parameters (regression coefficients). In economics the normal distribution is popular. The link function is the inverse cumulative normal distribution. The model that results is the probit model. In health sciences, extreme value distributions are often used. The transition rate model and the probability model are sometimes combined. Recall that $\mu_{ij}(x,t)$ may be written as the product of an escape rate from i, $\mu_{i+}(x,t)$, and a probability of destination, conditional on leaving state i, $\xi_{ij}(x,t) = p_{ij}(x,t) / p_{i+}(x,t)$.

Covariates may be introduced in two ways: by stratifying the population by the relevant covariate(s), provided the covariates are discrete variables, or by a regression equation. In population projections the population is usually stratified by sex and birth cohort (age) and demographic parameters are sex- and birth cohort-specific. Regression models are more economic at a cost of precision. Stratification involves as many parameters as there are cells in the cross-classification of covariates. The same is true only in a saturated regression model, which is a model that has as many independent parameters as there are unknowns (cells in the cross-classification). A saturated regression model involves several interaction effects, which may be redundant in most practical applications.

In the regression model, the covariates of individual k are denoted by the vector ${}_k\mathbf{Z}$ (${}_k\mathbf{Z} = \{{}_kZ_1, {}_kZ_2, {}_kZ_3, \ldots\}$). A covariate ${}_kZ_p$ may represent a single attribute or a combination of attributes (to denote interaction effects). The covariates may vary with age. The state occupied at x is always one of the covariates. The vector ${}_k\mathbf{Z}(x,t)$ denotes the covariates of individual k at age x at time t. The state occupied at x may be included in the list of p covariates or may be treated separately in which case an additional subscript is added to ${}_k\mathbf{Z}(x,t)$. In this paper, the state occupied is one of the covariates. The expected value of a transition rate or transition probability predicted by the regression model is determined by the covariates. Individuals with the same covariates have the same expected transition rates and transition probabilities. In addition to covariates, one may include time (when the rates are time-varying) and/or cohort (when rates are estimated for different cohorts) among the explanatory variables or predictors. The covariates may also include personal attributes at a prior age, such as childhood characteristics. In fact, the transition rates and the transition probabilities may depend on the entire life history. The life course from birth to age x at time t is denoted by ${}_k\mathbf{\Phi}(x,t)$ (see Section 4). If a transitions depends on the entire life course or on a selection of characteristics at earlier points in time (including states occupied at previous ages), the Markov assumption no longer holds. In statistical inference, we assume that individuals are independent and that the observation of life histories is independent across individuals (see also Hoem and Funck Jensen, 1982, p. 243).

We first consider transition probabilities. Transition rates are considered next. Transition probabilities are related to covariates using a logit model or logistic regression. Transition rates are related to covariates using transition rate models that are related to the family of Poisson regression models.

7.2. Transition probabilities

In Section 4, the indicator variable ${}_kY_i(x,t)$ was introduced. It takes on a value of 1 if at age x and time t individual k is in state i. It is zero otherwise. Now we add covariates. Let ${}_kY_{ij}[x,t,{}_k\mathbf{Z}(x,t)]$ be a time-varying indicator variable which takes on the value 1 if individual k who has characteristics ${}_k\mathbf{Z}(x,t)$ at current age x and who occupies state i at x, occupies state j at age x+h. It is zero otherwise. An observed value of the random variable Y is denoted by y. The expected value of ${}_kY_{ij}[x,t,{}_k\mathbf{Z}(x,t)]$ is the transition probability ${}^h_k p_{ij}[x,t,{}_k\mathbf{Z}(x,t)]$. It is the probability that individual k, who is born at t-x, who is currently aged x, *who occupies state i* and who has other characteristics ${}_k\mathbf{Z}(x,t)$, occupies state j at age x+h. The logit equation relates state probabilities to covariates.

$$\log it\left[{}^h_k p_{ij}[x,t,{}_k\mathbf{Z}(x,t)]\right] = \ln \frac{{}^h_k p_{ij}\left[x,t,{}_k\mathbf{Z}(x,t)\right]}{{}^h_k p_{ir}\left[x,t,{}_k\mathbf{Z}(x,t)\right]}$$

$$= \ln {}^h_k\theta_{ij}\left[x,t,{}_k\mathbf{Z}(x,t)\right] = {}^h_k\eta_{ij}\left[x,t,{}_k\mathbf{Z}(x,t)\right]$$

$$= \beta_0(x,t) + \beta_1(x,t)\,{}_kZ_1(x,t) + \beta_{i2}(x,t)\,{}_kZ_2(x,t) + \beta_{i3}(x,t)\,{}_kZ_3(x,t) + \ldots$$

where r is the reference state (category) and ${}^h_k\theta_{ij}[x,t,{}_k\mathbf{Z}(x,t)] = \dfrac{{}^h_k p_{ij}[x,t,{}_k\mathbf{Z}(x,t)]}{{}^h_k p_{ir}[x,t,{}_k\mathbf{Z}(x,t)]}$ is the odds that individual k with characteristics ${}_k\mathbf{Z}(x,t)$ and currently in state i and aged x, occupies state j at x+h rather than the reference state r. ${}_kZ_p(x,t)$ is the value of the p-th covariate of individual k at age x and time t. Note that ${}_k\mathbf{Z}(x)$ may include interaction effects, such as the interaction between the covariates at x and the state occupied at x. If that interaction is included, the regression equation may be written as

$$\log it\left[{}^h_k p_{ij}[x,t,{}_k\mathbf{Z}_i(x,t)]\right] = \beta_{i0}(x,t) + \beta_{i1}(x,t)\,{}_kZ_{i1}(x,t) + \beta_{i2}(x,t)\,{}_kZ_{i2}(x,t) + \beta_{i3}(x,t)\,{}_kZ_{i3}(x,t) + \ldots$$

If the effects of the covariates vary with age but do not vary over time, than

$$\log it\left[{}^h_k p_{ij}[x,t,{}_k\mathbf{Z}_i(x,t)]\right] = \beta_{i0}(x) + \beta_{i1}(x)\ {}_kZ_{i1}(x,t) + \beta_{i2}(x)\ {}_kZ_{i2}(x,t) + \beta_{i3}(x)\ {}_kZ_{i3}(x,t) + \ldots$$

The logit transformation assures that the state probabilities lie between 0 and 1, and that their sum is equal to one. The value of ${}_k\eta_{ij}$ may range from -∞ to +∞, but the value of p_{ij} remains within 0 and 1. To obtain the probabilities, the logit scale is converted into the probability scale:

$${}^h_k p_{ij}[x,t,{}_k\mathbf{Z}(x,t)] = \frac{\exp\left[{}^h_k\eta_{ij}[x,t,{}^k\mathbf{Z}(x,t)]\right]}{\exp\left[{}^h_k\eta_{i1}[x,t,{}^k\mathbf{Z}(x,t)]\right] + \exp\left[{}^h_k\eta_{i2}[x,t,{}^k\mathbf{Z}(x,t)]\right] + \ldots} = \frac{\exp\left[{}^h_k\eta_{ij}[x,t,{}^k\mathbf{Z}(x,t)]\right]}{\sum_{s=1}^{I}\exp\left[{}^h_k\eta_{is}[x,t,{}^k\mathbf{Z}(x,t)]\right]}$$

where $\exp\left[{}^h_k\eta_{ir}[x,t,{}_k\mathbf{Z}(x,t)]\right] = 1$ (s=r), with r the reference category. The model is the multinomial logistic regression model. The regression model is estimated for each state of origin separately. Since all individuals with the same characteristics have the same transition probabilities, k may be omitted.

The coefficients of the logit model and the logistic regression model are estimated from the data by the maximum likelihood method. The method maximizes the probability that the model predicts the data. Let ${}^{h}n_{ij}[x,t,\mathbf{Z}(x,t)]$ denote the number of individuals in i aged x at t and with characteristics $\mathbf{Z}(x,t)$ that occupy state j at x+h. The total number of individuals with characteristics $\mathbf{Z}(x,t)$ and aged x in state i is ${}^{h}n_{i+}[x,t,\mathbf{Z}(x,t)]$. The probability of observing ${}^{h}n_{i1}[x,t,\mathbf{Z}(x,t)]$ individuals in state 1 at age x+h, ${}^{h}n_{i2}[x,t,\mathbf{Z}(x,t)]$ in state 2, ${}^{h}n_{i3}[x,t,\mathbf{Z}(x,t)]$ in state 3, etc., is given by the multinomial distribution

$$\Pr\{N_{i1} = n_{i1}, N_{i2} = n_{i2}, ...\} = \frac{{}^{n}n_{i+}[x,t,\mathbf{Z}(x,t)]!}{\prod_{j=1}^{I} {}^{h}n_{ij}[x,t,\mathbf{Z}(x,t)]!} \prod_{j=1}^{I} \left[{}^{h}p_{ij}[x,t,\mathbf{Z}(x,t)]\right]^{{}^{h}n_{ij}[x,t,\mathbf{Z}(x,t)]}$$

where

${}^{h}n_{ij}[x,t,\mathbf{Z}(x,t)] = \sum_{k=1}^{n_{i+}(x,t)} {}^{h}_{k}Y_{ij}[x,t,{}_{k}\mathbf{Z}(x,t)]$ and $n_{i+}(x,t)$ is the size of the (sample) population in i and of age x at time t, and k is an individual in i at age x.

The log likelihood function for state i, age x and time t is

$$l = \sum_{j=1}^{I} {}^{h}n_{ij}[x,t,\mathbf{Z}(x,t)] \ln {}^{h}p_{ij}[x,t,\mathbf{Z}(x,t)]$$

where

$$\operatorname{logit}\left[{}^{h}_{k}p_{ij}[x,t,\mathbf{Z}_i(x,t)]\right] = \beta_{i0}(x,t) + \beta_{i1}(x,t)Z_{i1}(x,t) + \beta_{i2}(x,t)Z_{i2}(x,t) + \beta_{i3}(x,t)Z_{i3}(x,t) + ...$$

The transition probabilities are obtained by maximizing the log likelihood subject to the constraint that the destination probabilities $p_{ij}(.)$ must sum to one :

$\sum_{j=1}^{I} {}^{h}p_{ij}[x,t,\mathbf{Z}_i(x,t)] = 1$ (see also Amemiya, 1989, p. 417).

The Lagrangean function is (for convenience, we omit the age, time and covariate indices)

$$la = \sum_{j=1}^{I} n_{ij} \ln p_{ij} - \lambda_i \sum_{j=1}^{I} p_{ij} - 1$$

where λ_i is the Lagrange multiplier associated with the constraint that all destination probabilities for transitions that originate in i must sum to one.

The first-order conditions that must be satisfied are

$$\frac{\partial la}{\partial p_{ij}} = 0$$

Hence $\frac{n_{ij}}{p_{ij}} - \lambda_i = 0$ and $n_{ij} = \lambda_i \, p_{ij}$

To determine λ_i, we sum both sides over j and use the constraint $\sum_{j=1}^{I} p_{ij} = 1$. Hence $\lambda_i = \sum_{j=1}^{I} n_{ij} = n_{i+}$.

The maximum likelihood estimator of p_{ij} is

$$ {}^{h}\hat{p}_{ij}[x,t,\mathbf{Z}_i(x,t)] = \frac{{}^{h}n_{ij}[x,t,\mathbf{Z}_i(x,t)]}{{}^{h}n_{i+}[x,t,\mathbf{Z}_i(x,t)]} $$

For a more extensive discussion of the estimation of Markov chain models from data, see Amemiya (1989, pp. 412ff).

7.3. Transition rates

The instantaneous rate of transition may also depend on personal characteristics at the current age and on the entire life history. The matrix of transition intensities is $\boldsymbol{\mu}$[x,t,$\mathbf{Z}$(x,t)] and an element of the matrix is μ_{ij}[x,t,$\mathbf{Z}$(x,t)]. The dependence of transition intensities on personal attributes and other explanatory variables is described by transition rate models (Tuma and Hannan, 1984; Blossfeld and Rohwer, 2002; Lancaster, 1990; Andersen et al., 1993). In this section, the dependent variable is the transition rates or occurrence-exposure rate, which is the discrete-time counterpart of the transition intensity. It is the ratio of the number of events during an interval to the total exposure time in the interval. Compare with ${}^{h}_{k}\mathbf{M}(x,\mathrm{t}) = {}^{h}_{k}\mathbf{n}(x,\mathrm{t})\left[diag\ {}^{h}_{k}\overline{\mathbf{L}}(x,\mathrm{t})\right]^{-1}$ in Section 4.6 and ${}^{h}_{c}\mathbf{n}^{*}(x,t) = {}^{h}_{c}\mathbf{m}(x,t)\, diag\left[{}^{h}_{c}\overline{\mathbf{L}}(x,t\right]{}^{h}B(t-x)$ in Section 5.2.

The dependent variable is a ratio of two random variables: the number of events during an interval experienced by a group of people with a given set of characteristics and the exposure during the same interval by the same group of people. It is generally assumed that the exposure time (denominator) does not change with changes in numbers of events. In that case, the denominator is not a random variable, and the transition rate models become Poisson regression models with a constant, generally referred to as *offset*. The offset will be denoted by PY (person-years) since the total time a group of people is exposed to the risk of an event is usually expressed in *person-years*. A special case of a Poisson regression model is when the dependent variable of the regression model and all independent variables are discrete variables. The special case is known as the log-rate model, which is a log-linear model with offset (Laird and Olivier, 1981; Yamaguchi, 1991, Chapter 4).

Transition rate models include the basic (exponential) transition rate model, the piecewise constant rate model, and the Cox regression model. Transition rate models also include parametric model of time- (or age-)dependence such as the Gompertz model, the Weibull model, the Coale-McNeil (1972) model (for fertility and nuptiality), the Hernes (1972) model (for first marriage), the Heligman-Pollard (1980) model (for mortality) and the Rogers-Castro (1981) model (for migration). Transition rate models are estimated from empirical data. The data may be vital statistics, census data or surveys. The estimation of transition rates from survey data require recently developed theories of statistical inference.

The basic parameters of the multistate model are the transition rates ${}^{h}m_{ij}(x,t)$. As described in a previous section, a transition rate may be written as the product of an exit rate (escape rate) ${}^{h}m_{i}(x,t)$ [or ${}^{h}m_{i+}(x,t)$] and a conditional transition probability to capture the distribution among multiple destinations. The exit rate is modelled using a transition rate model for a single event (leaving the state of origin). The regression model linking an exit rate to covariates is

$$ {}^{h}m_{i}[x,t,\mathbf{Z}(x,t)] = \exp\left[\beta_{i0}(x) + \beta_{i1}(x)Z_{1}(x,t) + \beta_{i2}(x)Z_{2}(x,t) + ...\right] $$

where the regression coefficients are assumed to be constant in time. A regression model linking transition rates to covariates is

$$ {}^{h}m_{ij}[x,t,\mathbf{Z}(x,t)] = \exp\left[\beta_{ij0}(x) + \beta_{ij1}(x)Z_{1}(x,t) + \beta_{ij2}(x)Z_{2}(x,t) + ...\right] $$

The models may be written as log-linear models

$$ \ln {}^{h}m_{i}[x,t,\mathbf{Z}(x,t)] = \beta_{i0}(x) + \beta_{i1}(x)Z_{1}(x,t) + \beta_{i2}(x)Z_{2}(x,t) + ... $$

$$ \ln {}^{h}m_{ij}[x,t,\mathbf{Z}(x,t)] = \beta_{ij0}(x) + \beta_{ij1}(x)Z_{1}(x,t) + \beta_{ij2}(x)Z_{2}(x,t) + ... $$

The transition rate is the ratio of number of events over total exposure time. These two components may be studied separately, as is done in the log-rate model (see e.g. Yamaguchi, 1991, Chapter 4). In that model, it is assumed that changes in the number and timing of direct transitions (events) do not significantly affect the total exposure time. The assumption is realistic when exposure time is large compared to the number of transitions. If a variation in number or timing of transitions does not affect total exposure, the latter component may be considered fixed and may be treated as an *offset* in probability models including regression models. The problem of modeling transition rates reduces to the prediction of the expected number of events (counts) given the covariate combination. The expected number of events is the numerator of the transition rate. The number of direct transitions that occur during a unit interval is often represented by a Poisson random variable. The number of events that may occur during the interval is not restricted in any way, but it is assumed that the events are independent. Subjects in a (sample) population may experience more than one event during the unit interval. The Poisson model is (for convenience, we omit the age, time and covariate indices)

$$ \Pr\{N_{\mathrm{ij}} = \mathrm{n}_{\mathrm{ij}}\} = \frac{\lambda_{ij}^{n_{ij}}}{n_{ij}!} \exp[-\lambda_{\mathrm{ij}}] $$

where N_{ij} is a random variable denoting the number of transitions from i to j during a unit interval, n_{ij} is the observed number of transitions, and λ_{ij} is the expected number of transitions. The latter is the parameter of the Poisson model. It is assumed that the transitions are independent. The parameter may be made dependent on covariates:

$$ E\left[N_{ij}\right] = \lambda_{ij} = \exp\left[\beta_{ij0} + \beta_{ij1}Z_{1} + \beta_{ij2}Z_{2} + ...\right] $$

The model may be written as a log-linear model:

$$ \ln \lambda_{ij} = \beta_{ij0} + \beta_{ij1}Z_{1} + \beta_{ij2}Z_{2} + ... $$

If all covariates are discrete or categorical, the observations on transitions can be arranged in a contingency table (Laird and Olivier, 1981). The covariates refer to rows, columns, layers and combinations of these (to represent interaction effects).

The log-rate model is a log-linear model with an offset:

$$E\left[\frac{N_{ij}}{PY_i}\right]=\frac{\lambda_{ij}}{PY_i}=\exp\left[\beta_{ij0}+\beta_{ij1}Z_1+\beta_{ij2}Z_2+...\right]$$

where PY_i denotes exposure time in state i (origin state). Since PY_i is fixed, the equation may be rewritten as follows:

$$E\left[N_{ij}\right]=\lambda_{ij}=PY_i\exp\left[\beta_{ij0}+\beta_{ij1}Z_1+\beta_{ij2}Z_2+...\right]$$

The age dependence may be introduced in two ways: non-parametric and parametric. In the first approach, the population is stratified by age and a transition rate is estimated for each age separately. In the parametric approach, age dependence is represented by a model. A common model is the Gompertz model, which imposes onto the transition rate an exponential change with duration. The Gompertz model has two parameters and each may be made dependent on covariates (For a detailed treatment, see Blossfeld and Rohwer, 2002; Willekens ??). Other parametric models of duration dependence may be used. In studies of marriage and fertility, the Coale-McNeil model and the Hernes model are often used to describe the age dependence of the marriage or first birth rate (see e.g. Liang, 2000; Kaneko, 2003). In migration studies, the model migration schedule is a common representation of the age dependence of the migration rate (see e.g. Rogers and Castro, 1986). Each parameter of the model may be related to covariates (see e.g. Blossfeld and Rohwer, 2002, p. 181). In practice, only one or a selection of parameters is assumed to depend on covariates[13]. Parametric models of age dependence of transition rates are relatively common in multistate projections (Rogers, 1986).

In some cases, the researcher is not interested in the age dependence or duration dependence of transition rates, but in the effect of covariates on the level of transition. Rather than omitting duration altogether, the transition rate is allowed to vary with age but the effect of the covariates on the transition rate does not vary with age/duration. The transition rate model that results is a Cox proportional hazard model. It is written as

$${}^{h}m_{ij}[x,t,\mathbf{Z}(x,t)]={}^{h}m_{ijo}(x,t)\exp\left[\beta_{ij0}+\beta_{ij1}Z_1+\beta_{ij2}Z_2+...\right]$$

where ${}^{h}m_{ijo}(x,t)$ is the baseline hazard. It is the set of age-specific transition rates for the reference category. Note that if the age dependence (age structure) of transition is independent of the dependence on covariates, the baseline hazard may be represented by a parametric model and the two components may be estimated separately since the log-likelihood to maximize splits in a sum that includes two separate parts, one related to age and the other to the covariates. For a recent application, see Bender et al. (2005). Also note that in cohort analysis, $t = t_c+x$, and time basically refers to the individual's membership of a birth cohort.

---

[13] TDA (Transition Data Analysis) has a facility for user-defined rate models (Rohwer and Pötter, 1999, Section 6.17.5). The programme may be downloaded from prof. Rohwer's homepage: **http://www.stat.ruhr-uni-bochum.de/** The manual (extensive) can be downloaded from the same site. Willekens (2002) has written a brief introduction to TDA with examples.

8. Computing transition rates when data on events and exposures are missing

The estimation of transition rates requires observations on events and exposure times. When these data are missing but the continuous-time Markov chain model is expected to apply, transition rates may still be obtained. Two cases are considered. In the first case, data on events and exposures are missing but transition probabilities can be derived directly from the data. In the second case, data on events are available but information on exposures is missing.

8.1. Transition rates from transition probabilities

Transitions are frequently measured in discrete-time. Examples include the census, panel data and retrospective surveys. For instance, the census of the United States, Mexico, and several other countries, records for each respondent the place of residence at two points in time: the time of census and 5 years prior to the census. Labour force surveys often record the occupation at time of survey and the occupation one year prior to the survey. From that information, discrete-time transitions are available for individuals who are present at the two points in time. Individuals who leave the population during the observation period are lost to follow-up. Individuals who enter the population system during the observation period are included as new entries. In case discrete-time transitions are recorded, multiple transitions during the observation interval are disregarded. Consider an individual who was not in the labour force one year ago, entered the labour force in occupation A and changed to occupation B before the survey. That invidivual is registered as having experienced a discrete-time transition from 'out-of-the-labour-force' to occupation B. If state occupancies are measured at two points in time, discrete-time transition probabilities can be obtained directly from the data for individuals present at the two points in time. In retrospective surveys, the transition probabilities are conditional on survival.

Discrete-time transitions do not accurately describe the level of transitions in a population since several important transitions may escape the eye of the observer. For instance, several young adults leaving school experience a brief period of unemployment while they are searching for the job they like. Many young adults with a job at time of the survey, who experienced a brief period of unemployment during the past 12 months, were in school one year prior to the survey. Discrete-time transition data do not reveal many of these short-term unemployment spells. The measurement of transitions in continuous time (events) would reveal the unemployment spells, whereas the measurement of transitions in discrete time does not. The number of events during an interval can be inferred from the information on discrete-time transitions if the timing of the events and the directions of the transitions can be described by a continuous-time Markov Chain (CTMC). This section describes the estimation method.

If a Markov chain describes the direct transitions during the observation period, then the rates of transition may be derived from the discrete-time transition probabilities. The estimation of the rate of transition during an interval from data on states occupied at two consecutive points in time is the ***inverse problem***, which was discussed by Singer and Spilerman (1979) (see also Bartholomew, 1982, p. 94; Kalbfleisch and Lawless, 1985; van Imhoff and Keilman, 1991, pp. 76ff. The issue is also discussed by Amemiya, 1989, pp. 440ff). The problem is equivalent to one in which the discrete-time transition probabilities ${}^h\mathbf{P}(x,t)$ are given and the empirical transition rates ${}^h\mathbf{M}(x,t)$ are required. The derivation starts with the exponential expression ${}^h\mathbf{P}(x,t) = \exp\left[-h\ {}^h\mathbf{M}(x,t)\right]$. We consider three methods. The age and time indices are omitted for convenience.

a. Diagonalization of the matrix ${}^h\mathbf{P}$

Assume ${}^h\mathbf{P}$ has distinct eigenvalues. In Annex A.3, we show that ${}^h\mathbf{P}$ may be transformed into a diagonal matrix:

$${}^h\mathbf{P} = \exp\left[-h\ {}^h\mathbf{M}\right] = \mathbf{D}\exp\left[-h\,\mathbf{\Lambda}\right]\mathbf{D}^{-1}$$

where $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues λ_i of ${}^h\mathbf{M}$, and $\mathbf{D}$ is the modal matrix. ${}^h\mathbf{P}$ and ${}^h\mathbf{M}$ have the same modal matrices and the eigenvalues of ${}^h\mathbf{P}$ are $b_i = \exp[-h\ \lambda_i]$. If ${}^h\mathbf{P}$ is known, the eigenvalues of ${}^h\mathbf{P}$ are known and the eigenvalues of ${}^h\mathbf{M}$ can be derived:

$$\lambda_i = -\frac{1}{h}\ln b_i$$

The matrix of transition rates is therefore

$${}^h\mathbf{M} = \mathbf{D}\,\mathbf{\Lambda}\,\mathbf{D}^{-1}$$

b. Linear approximation

The exponential expression may be approximated by a linear model. The approximation is adequate when the transition rates are small or the interval is short.

$${}^h\mathbf{P} = \exp\left[-h\ {}^h\mathbf{M}\right] \cong \left[\mathbf{I} + \tfrac{h}{2}\ {}^h\mathbf{M}\right]^{-1}\left[\mathbf{I} - \tfrac{h}{2}\ {}^h\mathbf{M}\right]$$

$$\left[\mathbf{I} + \tfrac{h}{2}\ {}^h\mathbf{M}\right]{}^h\mathbf{P} \cong \left[\mathbf{I} - \tfrac{h}{2}\ {}^h\mathbf{M}\right]$$

$${}^h\mathbf{P} + \tfrac{h}{2}\ {}^h\mathbf{M}\ {}^h\mathbf{P} \cong \mathbf{I} - \tfrac{h}{2}\ {}^h\mathbf{M}$$

This leads to the desired result

$${}^h\mathbf{M} \cong \tfrac{2}{h}\left[\mathbf{I} - {}^h\mathbf{P}\right]\left[\mathbf{I} + {}^h\mathbf{P}\right]^{-1} = \tfrac{2}{h}\left[2(\mathbf{I} + {}^h\mathbf{P})^{-1} - \mathbf{I}\right]$$

since ${}^h\mathbf{P} + \mathbf{I} \cong \mathbf{I} + \left[\mathbf{I} + \tfrac{h}{2}\ {}^h\mathbf{M}\right]^{-1}\left[\mathbf{I} - \tfrac{h}{2}\ {}^h\mathbf{M}\right] = 2\left[\mathbf{I} + \tfrac{h}{2}\ {}^h\mathbf{M}\right]^{-1}$

The linear approximation requires that the inverse $\left[\mathbf{I} + {}^h\mathbf{P}\right]^{-1}$ exists. The expression is also shown by Schoen (1988, p. 78). Schoen derives the expression to infer movements between states form knowledge of where individuals are at two points in time.

The inverse relation may be used to infer transition probabilities for intervals that are different from the measurement intervals. For instance, if the states occupied are recorded at ages x and x+h, the inverse relation may be used to infer the average transition rates ${}^{h}\mathbf{M}(x,t)$ and to derive the transition probabilities over a one-year period. The expression is ${}^{1}\mathbf{P}(x,t)=\exp\left[-{}^{h}\mathbf{M}(x,t)\right]$ where ${}^{h}\mathbf{M}(x,t)$ is estimated from ${}^{h}\mathbf{P}(x,t)$ using the inverse method. The method assumes that transition rates are constant during the (x,x+h)-interval and that the linearity assumption is an adequate approximation of the exponential model.

c. Taylor series expansion

Annex A gives the Taylor series expansion for $\exp\left[-{}^{h}\mathbf{M}(x,t)\right]$. From that expansion, the inverse may be obtained (Namboodiri and Suchindran, 1987, p. 162):

$${}^{h}\mathbf{M}=-\tfrac{1}{h}\left[(\mathbf{I}-{}^{h}\mathbf{P})+\tfrac{1}{2}(\mathbf{I}-{}^{h}\mathbf{P})^{2}+\tfrac{1}{3}(\mathbf{I}-{}^{h}\mathbf{P})^{3}+\ldots\right]$$

8.2. Transition rates from events and initial population

In the absence of data on exposure times, rates and sojourn times may be estimated using an iterative procedure. The iterative computation scheme is due to Gill and Keilman (1990) and Gill (1986). The procedure is also described in van Imhoff and Keilman (1991, p. 31).

We derive the equations for individual k. Given an initial guess of the sojourn times, the empirical rates ${}^{h}_{k}\mathbf{M}(x,t)$ may be determined from events using the following equation (see Section 4):

$${}^{h}_{k}\mathbf{M}(x,t)=-{}^{h}_{k}\mathbf{n}(x,t)\left[diag\ {}^{h}_{k}\overline{\mathbf{L}}(x,t)\right]^{-1}$$

where an element ${}^{h}_{k}n_{ij}(x,t)$ of ${}^{h}_{k}\mathbf{n}(x,t)$ is the observed number of moves from i to j during the interval (x, x+h) by individual k and ${}^{h}_{k}\overline{L}_{i}(x,t)$ is the number of years (or other unit of time) the individual stays in state i between the ages x and x+h. The minus sign is because, in the matrix configuration adopted in this paper, the off-diagonal elements of the matrix of transition rates are negative. The diagonal of ${}^{h}_{k}\mathbf{n}(x,t)$ contains the negative of the number of direct transitions out of the state occupied at the time of leaving. The diagonal element of ${}^{h}_{k}\mathbf{n}(x,t)$ is $-\sum_{j\neq i}n_{ij}(x,t)$ (Gill and Keilman, 1990, p. 128; van Imhoff and Keilman, 1991, p. 30).

The initial guess is the solution to the linear model (see below). Improved estimates of sojourn times are obtained from the exponential model

$${}^{h}_{k}\overline{\mathbf{L}}(x,t)={}^{h}_{k\,x}\mathbf{L}(x,t)\ {}_{k}\mathbf{k}(x,t)=\left[{}^{h}_{k}\mathbf{M}(x,t)\right]^{-1}\left[\mathbf{I}-\exp[-h\ {}^{h}_{k}\mathbf{M}(x,t)]\right]{}_{k}\mathbf{k}(x,t)$$

where ${}^{h}_{k\,x}\mathbf{L}(x,t)$ is a matrix of expected sojourn times in the different states during the interval from x to x+h by individual k born at t-x and currently of exact age x, by state occupied at x and ${}^{h}_{k}\overline{\mathbf{L}}(x,t)$ is a vector with elements the expected number of years spent in the different states. The improved estimates of ${}^{h}_{k}\overline{\mathbf{L}}(x,t)$ are entered into the rate equation to obtain improved estimates of the transition rates. If members of the same cohort are assumed to experience the same transition rates, then information on events from all members can be used to determine the transition rates.

The initial estimates of the sojourn times are derived from the flow equation that is consistent with the linear model (see Annex B):

$$ {}_k\mathbf{k}(x+h,t+h) = \left[\mathbf{I} - {}^h_k\mathbf{m}(x,t)\ {}^h_{k\,x}\mathbf{L}(x,t)\right] {}_k\mathbf{k}(x,t) $$

and the assumption of uniform distribution of events, which leads to the following approximation of the sojourn times

$$ {}^h_k\overline{\mathbf{L}}(x,t) = \tfrac{h}{2}\left[{}_k\mathbf{k}(x,t) + {}_k\mathbf{k}(x+h,t+h)\right] $$

The number of events individual k may expect to experience during the interval (t,t+h) is ${}^h_k\mathbf{n}(x,t) = -{}^h_k\mathbf{M}(x,t)\left[diag\,{}^h_k\overline{\mathbf{L}}(x,t)\right]$. The sojourn times follow from ${}^h_k\mathbf{n}(x,t)$ and the state probably vector ${}^h_k\mathbf{k}(x,t)$ associated with an individual aged x at time t:

$$ {}^h_k\overline{\mathbf{L}}(x,t) = \tfrac{h}{2}\left[{}_k\mathbf{k}(x,t) + {}_k\mathbf{k}(x+h,t+h)\right] = \left[h\mathbf{I} + \tfrac{h}{2}\ {}^h_k\mathbf{n}(x,t)\right] {}_k\mathbf{k}(x,t) $$

9. A note on individual biographies and microsimulation

In Section 4, we introduced the variable ${}_kY(x,t)$ and the indicator variable ${}_kY_i(x,t)$ to denote the state occupied by individual k at exact age x and exact time t. They are discrete random variables. ${}_kY(x,t)$ is a variable that can take on as many non-zero values as there are states in the state space. The sequence $\{{}_kY(x,t),\ x \geq 0 \text{ and } t \geq t_b\}$ is a stochastic process identifying the state occupied at each age x and time t, with t_b the date of birth. The stochastic process $\{{}_kY(x,t),\ x \geq 0 \text{ and } t \geq t_b\}$ has been described by a continuous-time Markov chain (CTMC) with transition intensities ${}_k\mu_{ij}(x,t)$. In Section 4, the model specification is for a specific individual; namely, individual k. In Section 5, it was assumed that all individuals born during the same period experience the same transition intensities. As a consequence of the homogeneity assumption, the expected values of the demographic indicators predicted by the models apply to all cohort members. That restrictive assumption was relaxed in Section 7. The introduction of covariates paved the way to differentiate individual cohort members on the basis of observed attributes and to use the information in predicting behaviour. Technically, it means that the value the random variable ${}_kY(x,t)$ at x and t depends on the characteristics of individual k at x and t and at earlier ages. The life path beyond x may depend on the entire observed life history (the sample path). The life path is given by the stochastic process $\{{}_kY(x,t,{}_k\mathbf{\Phi}(x,t)),\ x \geq 0 \text{ and } t \geq t_b\}$, where ${}_k\mathbf{\Phi}(x,t)$ denotes the life history of k up to age x at time t. The transition intensities are ${}_k\mu_{ij}[x,t,{}_k\mathbf{\Phi}(x,t)]$. If the transition intensities do not depend on the entire life history but only on a limited number of factor, the transition intensities are denoted by ${}_k\mu_{ij}[x,t,{}_k\mathbf{Z}(x,t)]$, where the set of relevant current and previous characteristics of individual k at age x and time t is denoted by ${}_k\mathbf{Z}(x,t)$. The relations between transition rates and covariates are captured by regression models, known as *transition rate models.* The models include a selection of characteristics as predictor variables. Individuals with the same characteristics may expect to have the same transition intensities. If these individuals maintain the same characteristics at future ages, the models will predict a similar life course.

Individuals are not identical, even if they have the same age, occupy the same state of the state space and have the same set of characteristics ${}_k\mathbf{Z}(x,t)$. They differ in characteristics that are not recorded or observed, or that are omitted from the selection of relevant factors. These characteristics are represented by latent variables. If the latent variables are discrete, individuals with the same latent characteristics constitute a group and group membership is not observed but must be determined in another way. The discrete latent variable representing group membership defines a finite mixture model (McLachlan and Peel, 2000). Applications of finite mixture models in demography are few. Most studies of unobserved heterogeneity in demography assume continuous mixing distributions (for a discussion of unobserved heterogeneity in demographic models, see Heckman and Singer, 1982, and for a list of references, see Dias and Willekens, 2005). Let ${}_k\mathbf{Z}(x,t)$ denote the observed attributes of individual k at age x and time t and let ${}_kX(x,t)$ denote a single latent attribute of k at age x and time t. We assume that the latent attribute can take on one of S values, i.e. there are S latent states and hence S subpopulations. Suppose that individuals with different latent states differ in a single parameter θ. The value of the parameter for an individual in latent state s is θ_s. The transition intensity is

$$ {}_k\mu_{ij}\left[x,t,{}_k\mathbf{Z}(,t)\right]=\sum_{s=1}^{S} {}_k\pi_s(x,t)\ {}_k\mu_{ij}\left[x,t,{}_k\mathbf{Z}(x,t),\theta_s\right] $$

where ${}_k\pi_s(x,t)$ is the probability that at age x and time t individual k belongs to group s. Individual k, who at time t is aged x and has overt attributes ${}_k\mathbf{Z}(x,t)$ and a single latent attribute characterised by parameter θ_s, has the transition intensity ${}_k\mu_{ij}[x,t,{}_k\mathbf{Z}(x,t),\theta_s]$. The transition intensity at age x and time t depends on the latent groups individual k may be part of and the characteristics of these groups (captured by parameter θ_s). The distribution that predict the transition intensities is a mixing distribution. The latent variable describing the distribution of an unobserved characteristic may also be a continuous variable. For an overview, see e.g. Vermunt (2002).

In some applications, it is not sufficient to determine *expected* transition intensities (or its discrete-time analogue, the transition rate) for individuals of the same age that share personal attributes. One needs to know who experiences a transition and who does not. In other words, it is not sufficient to know the instantaneous rate of transition at age x and time t for individuals in state i and with attributes $\mathbf{Z}_i(x,t)$, $\mu_{ij}[x,t,\mathbf{Z}_i(x,t)]$, the rate of transition ${}^hm_{ij}[x,t,\mathbf{Z}_i(x,t)]$, or the probability of transition ${}^hp_{ij}[x,t,\mathbf{Z}_i(x,t)]$ that are expected given the data. One needs to know whether individual k with attributes ${}_k\mathbf{Z}_i(x,t)$ makes a transition from i to j during a given interval. Since the transition rate models will predict the same *expected* transition intensities for all individuals with the same attributes, chance must determine who experiences a transition and who does not. It comes down to obtaining values for the indicator variables given expected values of transition rates and transition probabilities. That will be done by drawing random numbers from a probability distribution.

In section 4, the indicator variable ${}_kY_{ij}(x,t)$ was used to indicate whether or not individual k makes a transition from state i to state j at instantaneous age x and time t (direct transition or immediate jump) and ${}^h_kY_{ij}(x,t)$ was used to indicate whether or not individual k who occupies state i at age x and time t occupies state j h years later, i.e. whether or not he experiences a discrete-time transition. Note that the first random variable is not conditional on the state occupancy at x whereas the second is conditional on occupying state i at x. The expected value of the indicator variable ${}_kY_{ij}(x,t_c+x)$ is the probability density ${}_k\ell_{ij}(x,t_c+x)$ (note that t = t_c+x) (See section 4.1). The expected value of ${}^h_kY_{ij}(x,t)$ is ${}^h_k p_{ij}(x,t)$. The expected value E[${}_kY_{ij}(x,t_c+x)$]dx, provided k is in state i at age x at time t=t_c+x, is the probability ${}_k\mu_{ij}(x,t_c+x)\,dx$. The indicator variables ${}_kY_{ij}(x,t)$ and ${}^h_kY_{ij}(x,t)$ may depend on covariates ${}_k\mathbf{Z}(x,t)$. If one wants to know which members of a group experience a transition and which members do not, the expected values of the indicator variables are not sufficient. The individual values of the indicator variables must be determined. Given the *expected* values of the random variables ${}_kY_{ij}(x,t)$ and ${}^h_kY_{ij}(x,t)$, realizations of the random variables are determined by a chance mechanism. Transitions are randomly allocated to individuals of the same age and the same characteristics ${}_k\mathbf{Z}(x,t)$. The determine whether individual k experiences a transition, in discrete time or in continuous time, a random number is generated from an appropriate probability distribution. For transitions in discrete time, the appropriate probability distribution is the uniform distribution. For transitions in continuous time, the distribution is a waiting time distribution, also known as time-to-event distribution. Waiting time distributions describe the time (or duration) dependence of the direct transitions. The selection of the appropriate distribution is guided by the time dependence of the events. We first consider discrete-time transitions based on transition probabilities. Next we consider continuous-time transitions based on transition rates.

9.1. Random number generation for transitions in discrete time: generating discrete random variables

Recall that ${}^h_k p_{ij}[x,t,{}_k\mathbf{Z}_i(x,t)]$ denotes the conditional probability that an individual born at t-x, with attributes ${}_k\mathbf{Z}_i(x,t)$ at x and t, and occupying state i at exact age x, occupies state j at x+h. It is the expected value of the indicator variable ${}^h_kY_{ij}(x,t)$**:** $E\left[{}^h_kY_{ij}(x,t)\right] = {}^h_kp_{ij}[x,t,{}_k\mathbf{Z}_i(x,t)]$**.** The argument ${}_k\mathbf{Z}_i(x,t)$ is omitted for convenience. The problem is to determine realizations ${}^h_k y_{ij}(x,t)$ of the random variable ${}^h_kY_{ij}(x,t)$ that are consistent with the expected value and the assumed distribution of ${}^h_kY_{ij}(x,t)$. The appropriate distribution is the uniform distribution. For each individual k, a random number between 0 and 1 is drawn from the uniform distribution U~U[0,1], using a random number generator. Individual k with attributes ${}_k\mathbf{Z}_i(x,t)$ occupying state i experiences a transition to j if the number generated is less than the expected value ${}^h_k p_{ij}(x,t)$. If the value drawn from the distribution exceeds the expected value, the individual remains in the state of origin. If at x and t different events may occur, each having a probability, or an event may result in multiple destinations then different types of events are distinguished. The type refers to an event or a destination. The type of event that occurs is also determined by a random draw from a uniform distribution U~U[0,1]. Consider an event with I possible outcomes. If the event is a migration, then the outcome could be the region of destination. If the event is death, the type of event could be the cause of death. The probability that first event-type occurs is $p_1(t)$, the probability that the second event-type occurs is $p_2(t)$, etc. To determine the type of event, values of a discrete random variable are generated from a uniform distribution (see e.g. Ross, 2006, p. 49). Let the draw from U~U[0,1]be denoted by u. If u is less than $p_1(t)$, the first event-type occurs, if $p_1(t) \le u < p_1(t) + p_2(t)$, the second event-type occurs, if $p_1(t) + p_2(t) \le u < p_1(t) + p_2(t) + p_3(t)$, the third event-type occurs, etc. $[\sum_{i=1}^{I} p_i(t) = 1]$.The transitions are allocated to the individuals with the same attributes by sampling from a theoretical distribution, in this case a uniform distribution. Using this sampling technique, the sample proportion of individuals [with the same attributes $_k\mathbf{Z}_i(x,t)$] and occupying state i at x and t that experiences a transition will generally differ slightly from the expected value ${}_k^h p_{ij}(x,t)$ because of sampling variation, known as Monte Carlo variation. The larger the sample selected from the theoretical distribution, the smaller the Monte Carlo variation.

9.2. Random number generation for transitions in continuous time: generating continuous random variables

In this section we address the problem to determine who experiences direct transitions and when. For individual k, direct transitions are realizations ${}_k y_{ij}(x,t)$ of the random variable ${}_kY_{ij}(x,t)$**,** which indicates the timing of the (i,j)-transition. The realizations must be consistent with the expected value ${}_k\ell_{ij}(x,t) = E[{}_kY_{ij}(x,t)]$ (see section 4)**,** and the assumed distribution of the random variable ${}_kY_{ij}(x,t)$. The appropriate distribution is a waiting time distribution. The parameter of the waiting time distribution is the transition intensity ${}_k\mu_{ij}(x,t) = {}_k\ell_{ij}(x,t) / {}_k\ell_i(x,t)$ where ${}_k\ell_i(x,t)$ is the state probability, i.e. the probability that individual k occupies state i at age x, which is reached at time t. If the transition intensity is constant, the waiting time to the direct transition (event) is exponentially distributed. If the transition intensity changes exponentially with x, the waiting time to the event follows a Gompertz distribution. If the transition intensity is a power function of the duration x, the waiting time to the event follows a Weibull distribution. Other distributions may be added. The transition intensity may follow a given pattern throughout the life span or for a given segment of the life span only. For instance, the transition intensity may remain at a constant value during a given age interval and may shift to another constant value during the next interval. In that case, the transition intensity is piecewise constant.

Let X be a continuous random variable denoting the waiting time to the event under study. The distribution of X, F(x), indicates the probability that the waiting time to the event is less than x. The real number or quantile x may vary from 0 to ∞ whereas the probability varies from 0 to 1. The waiting time distribution of X maps a realization x of X into a probability. The inverse distribution function maps a probability into a real number. Let U be a uniform random variable with values between 0 and 1. For any continuous distribution function F(x) the random variable X defined by X = F^{-1}(U) has distribution F. This proposition implies that we can generate a random variable X with distribution F by generating a random variable U with a uniform distribution in the range (0,1) and then setting X = F^{-1}(U) (for the proof, see e.g. Ross, 2006, pp. 67ff). The inverse distribution F^{-1}(u) is defined to be that value of x such that F(x) = u. The inverse distribution function (of probability u) is also known as the quantile function because it maps a probability u into a quantile x (Evans et al., 2000, p. 8). It is denoted by F^{-1}(u) and G(u). The general method for generating a random variable with a given continuous distribution is called the inverse transformation method or the quantile method. By way of illustration consider an exponentially distributed random variable X indicating the waiting time to an event. The distribution follows when the event occurs at a constant rate, μ say. The probability that the event occurs before x is given by the distribution function F(x) = 1 – exp(-μt). The procedure to generate a realization of X that is consistent with an exponential waiting time distribution, consists of two steps. In a first step, a random number is drawn from the uniform distribution U~U[0,1]. Let that number be denoted by u. In a second step, the quantile function is used to determine the value of X: x = G(u) = F^{-1}(u) = $-\frac{\ln[1-u]}{\mu}$. By drawing many realization u of U and applying the quantile function, the distribution of X is generated. This method is applicable when the inverse distribution function has a closed-form expression. When the inverse distribution function has no closed-form expression, the method to determine the quantile x associated with the probability F(x) is more complicated. For instance, draws from a standard normal distribution is complicated because there is no closed form for the inverse standard normal. Greene (1997, p. 178) discusses two ways to obtain the quantiles associated with given probabilities. Liang (2000, pp. 35-36) presents a method to obtain the quantile x associated with a given probability F(x) when the distribution F(x) is the Coale-McNeil model, which has no closed-form expression either.

The general method may be applied to determine the timing of direct transitions assuming (piecewise) constant transition intensities. Assume that the transition intensity is constant during the age interval (x,x+h) and let the constant transition intensity be ${}_k^h m_{ij}(x,t)$. The probability that individual k, who occupies state i on his x-th birthday, leaves i between x and x+ξ (0 ≤ ξ < h) is the distribution function, ${}_k^\xi F_i(x,t) = 1 - {}_k^\xi p_i^{cont}(x,t) = 1 - \exp[-{}_k^h m_{i+}(x,t)\,\xi]$ where ${}_k^\xi p_i^{cont}(x,t)$ is the survival function, defined in Section 4. The procedure to generate a value of ξ that is consistent with an exponential waiting time distribution, consists of two steps. In a first step, a random number is drawn from the uniform distribution U~U[0,1]. Let that number be denoted by u. In the second step, the waiting time to the event (leaving state i) ξ is derived from u. If u is ${}_k^\xi F_i^*(x,t)$ then ξ is given by the inverse distribution function or quantile function: $\xi = {}_k^\xi F_i^{-1}(x,t) = G(u) = -\frac{\ln[1-u]}{{}_k^h m_{ii+}(x,t)}$. For a given value of u, G(u) gives the value of ξ for which ${}_k^\xi F_i(x,t) = u$. If ξ < h, the direct transition occurs in the interval (x, x+h). If ξ ≥ h, the event does not occur in the interval. The destination state is determined by generating a discrete random variable following the method described in Section 10.1.

Most microsimulation models consider transitions in discrete time. **C**ontinuous-time microsimulation models include the SOCSIM model developed by Hammel et al. at Berkeley (Hammel et al., 1976; Hammel 1990), the demographic PopSim part of the DYNAMOD model developed at NATSEM at the University of Canberra (Antcliff 1993), MICROHUS of Uppsala University (Klevmarken and Olovsson , 1996), LifePaths of Statistics Canada (Gribble, 1997; Statistics Canada, 2001) and PENSIM of the US Department of Labor (Holmer et al., 2006). PENSIM uses the same algorithm as LifePaths (Holmer et al., 2006, p.3). The algorithm consists of drawing a sample of waiting times to event and comparing waiting times, generated by hazard models, to determine the timing and sequence of events. For a description of several of these models, including LifePaths, see Zaidi and Rake (2001). For a discussion of microsimulation in continuous time, see Willekens (2006).

9.3. A note on microsimulation and sampling

According to Wolf (2000) the essential ingredients of microsimulation are an analysis conducted at the level of the individual and the use of computer-based sampling. The sampling perspective on microsimulation is particularly interesting. Microsimulation consists of drawing a sample from a virtual population by generating realizations of random variables. The random variables considered in this study are ${}_k^hY_{ij}(x,t)$ and ${}_kY_{ij}(x,t)$. The 'observations' may pertain to other random variables. They may be a cross-section of the population at one point in time or a longitudinal study covering an extended period of time. The virtual population and its dynamics are fully described by one or several probability models. If the models are realistic, the virtual population closely resembles the real population. Suppose a follow-up study of a real population reveals that the number of individuals that experiences a given event declines exponentially, which implies a constant event rate (empirical rate). A virtual population that has waiting times to a given event generated by drawing from an exponential distribution with an event rate that is equal to the empirical rate, closely resembles the real population as far as the waiting time to the event is concerned.

10. Conclusion

Biographic forecasting is a new approach to demographic forecasting that integrates traditional forecasting of the population by age and sex and functional population forecasting. Biographic forecasting is replacing demographic forecasting for two reasons. First, demographic change is increasingly difficult to forecast because the idiosyncratic nature of demographic behaviour. Family formation, migration and attempts to grow old healthy are part of a lifestyle that also include work and other domains of life. The interest in life strategies originates from the awareness that critical decisions in life, i.e. decisions related to life events, are not taken in isolation but are based on lessons learned from past experiences (antecedents) and general conceptions about future developments in different life domains. Increasingly, demographic events are embedded in a life plan. The individual programming of life events, viewed by Légaré and Marcil-Gratton (1990) as a challenge for demographers in the twenty-first century, may be situated within the broader context of the individual design and implementation of life strategies. The second demographic transition, with its emphasis on choice biographies and changing interpersonal relationships, may also be viewed as a consequence of individualization and the emergence of individual life strategies.

The second reason is that increasingly demographic projections are required at great detail. That is the case particularly in forecasting the sustainability of pension systems and health care systems. Traditional demographic projection models are not suited for projections of a population decomposed at great detail. Biographic projections do not have that limitation.

The new demographic regime characterized by choice biographies and life planning and the growing demand for detailed population projections raise new challenges for demographic forecasters. Traditional projections by age and sex do not adequately capture the complexity of life. As a result the uncertainty increases. Probabilistic projections quantify the uncertainty but are not able to reduce the uncertainty. Reduction of uncertainty and increase of forecasting performance require more realistic projection models, i.e. models that are better able to integrate substantive knowledge and to capture the causal links that underlie childbirth, death, migration and the other events that shape the lives of people. This new generation of models consists of transition models or multistate models that capture transitions people make in life and the developmental processes and pathways that characterize individual lives. The multistate life table and the multistate projection model, combined with regression models of transition rates, are adequate candidates for the development of a new generation of demographic projection models. Cox's (1972) paper on regression models for life tables caused a revolution in survival analysis. The emerging discipline of survival analysis was provided with tools it needed for studying the effects of prognostic factors on individual survival in clinical trials designed to evaluate new cancer therapies. The model became a central research tool. The paper by Gill (1992) on regression models for multistate life tables went largely unnoticed although it showed that the Cox model and the associated techniques of statistical inference can be immediately applied to studying transitions in multistate demographic models. It is time to move away from narrow perspectives in demographic forecasting and to broaden the perspective by bridging micro- and macro-level analysis and by effectively integrating substantive knowledge and statistical perspectives and techniques into demographic modeling. The life course provides the logical framework, from a substantive and an analytical perspective. Techniques of statistical inference may be used to obtain parameters of projection models from observational data to complement vital statistics and census data and to provide for a richer empirical basis for demographic forecasts.

Bibliography

Ahlburg, D. A., W. Lutz, and J. W. Vaupel. (1999) Ways to improve population forecasting: What should be done differently in the future? Pages 191-198 in *Frontiers of Population Forecasting* (W. Lutz, J. W. Vaupel, and D. A. Ahlburg, eds.). A Supplement to Vol. 24, 1998, *Population and Development Review*. New York: The Population Council.

Akushevich, I., A. Kulminski and K.G. Manton (2005) Life tables with covariates: dynamic models for nonlinear analysis of longitudinal data. *Mathematical Population Studies*, 12:51-80.

Alho, J.M. and Spencer, B.D. (1997). The Practical Specification of the Expected Error of Population Forecasts. *Journal of Official Statistics*, 13:201-225.

Alho, J.M. and B.D. Spencer (2005) Statistical demography and forecasting. Springer, New York.

Alho, J. (n.d.) PEP – A program for error propagation. Available from internet: http://joyx.joensuu.fi/~ek/pep/userpep.htm

Amemiya, T. (1989) Advanced econometrics. Harvard University Press, Cambridge, Mass.

Andersen, P.J., O. Borgan, R.D. Gill and N. Keiding (1993) Statistical models based on counting processes. Springer Verlag, New York.

Anderson, K.N., P.M. Odell, P.W.F. Wilson and WB. Kannel (1991a) Cardiovascular disease risk profiles. *American Heart Journal*, 121:293-298.

Anderson, K.N., P.W.F. Wilson, P.M. Odell and WB. Kannel (1991b) An updated coronary risk profile. A statement for health professionals. *Circulation*, 83:356-362.

Andreev, K. (1999a) Demographic surfaces: estimation, assessement and presentation with applications to Danish mortality, 1835-1995. PhD Thesis, Faculty of Health Sciences, University of Southern Denmark, Odense (with CD-Rom).

Andreev, K. (1999b) Overview of the program Lexis 1.0. In: S. Heinzel and T. Plesser eds. Forschung und wissenschaftliches Rechnen. Published by Gesellschaft für wissenschaftliche Datenverarbeitung. Göttingen, pp. 107-121. Available at http://www.billingpreis.mpg.de/hbp98/andreev.pdf

Andreev, K. (2002) Evolution of the Danish population from 1835 to 2000. Odense Monograph on Population Aging Vol. 9 (with CD-Rom). University Press of Southern Denmark, Odense.

Antcliff, S. (1993) An Introduction to DYNAMOD: A Dynamic Microsimulation

Model, DYNAMOD Technical Paper no. 1, National Centre for Social and Economic Modelling, University of Canberra.

Aoki, M. (1976) Optimal control and system theory in dynamic economic analysis. North Holland, New York.

Aoki, M. (1996) New approaches to macroeconomic modeling. Evolutionary stochastic dynamics, multiple equilibria, and externalities as field effects. Cambridge University Press, Cambridge, UK.

Aoki, M. (2004) New frameworks for macro-economic modelings: some illustrative examples. Available at http://www.econ.ucla.edu/people/papers/Aoki/Aoki306.pdf

Arthur, B.W. and J.W. Vaupel (1984) Some general relationships in population dynamics. *Population Index*, 50(2):214-226.

Barker, D.J.P. (1998) Mothers, babies and health in later life. Churchill Livingstone, Edinburgh.

Bartholomew, D.J. (1982) Stochastic models for social processes. Third Edition. Wiley, Chichester.

Beiser, A., R.B. D'Agostino, S. Seshadri, L.M. Sullivan and P. Wolf (2000) Computing estimates of incidence, including lifetime risk: Alzheimer's diseases in the Framingham Study. The practical incidence estimators (PIE) macro. *Statistics in Medicine*, 19:1495-1522

Bender, R., T. Augustin and M. Blettner (2005) Generating survival times to simulate Cox proportional hazard models. *Statistics in Medicine*, 24:1713-1723.

Ben-Shlomo, Y. And D.L. Kuh (2002) A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology*, 31:285-293 (Editorial)

Bernardelli, H. (1941) Population waves. *Journal of Burma Research Society*, 31(1):1-18.

Bijak, J. (2007) Bayesian methods in international migration forecasting. Paper presented at the Workshop on the Estimation of International Migration in Europe: Issues, models and assessment. Co-organized at the University of Southampton by the Southampton Statistical Sciences Research Institute (S3RI) and the Netherlands Interdisciplinary Demographic Institute (NIDI), September 2005. Forthcoming in J. Raymer and F. Willekens eds. (2007) The estimation of international migration in Europe: Issues, models and assessment. Wiley, Chichester, UK.

Billari, F.C. and A. Prskawetz eds. (2003) Agent-based computational demography. Using simulation to improve our understanding of demographic behaviour. Physica Verlag (Springer), Heidelberg.

Billari, F.C., T. Fent, A. Prskawetz and J. Scheffran eds. (2006) Agent-based Computational Modelling : Applications in Demography, Social, Economic and Environmental Sciences. Physica Verlag (Springer), Heidelberg.

Biswas, S. (1988) Stochastic processes in demography and applications. Wiley, New Delhi.

Biswas, S. (1995) Applied stochastic processes: a biostatistical and population oriented approach. Wiley, New Delhi.

Blossfeld, H.P. (1998) A dynamic integration of micro- and macro-perspectives using longitudinal data and event history models. In: H.P. Blossfeld and G. Prein eds. Rational choice theory and large-scale data analysis. Westview Press, Boulder, Colorado, pp. 233-246.

Blossfeld, H.P. and G. Rohwer (2002) Techniques of event history modeling. New approaches to causal analysis. Lawrence Erlbaum, Mawah, New Jersey. Second Edition.

Bogue D.J., E.E. Arriaga and D.L. Anderton eds. (1993) Readings in population research methodology. Social Development Center, Chicago, and UNFPA, New York.

Booth, H. (2006) Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, 22:547-581. Also available as Working Paper 100, Demography and Sociology Programme, Australian National University, Canberra: http://demography.anu.edu.au/Publications/WorkingPapers/100.pdf

Breslow, N.E. and N.E. Day (1987) Statistical methods in cancer research. International Agency for Research on Cancer, Scientific Publication 82, Lyon.

Calot, G. and A. Franco (2002) The construction of life tables. In: G. Wunsch, M. Mouchart and J. Duchêne eds. The life table: Modeling survival and death, Kluwer Academic Publishers, Dordrecht, pp. 33-78.

Carone, G. (2005) Long-term labour force projections for the 25 EU Member States: A set of data for assessing the economic impact of ageing. Economic Papers no. 235, Directorate General for Economic and Financial Affairs, European Commission, Brussels. Available at: http://129.3.20.41/eps/lab/papers/0512/0512006.pdf

Carstensen, B. and N. Keiding (2004) Age-period-cohort models: Statistical inference in the lexis diagram. Lecture notes, Department of Biostatistics, University of Copenhagen, http://www.biostat.ku.dk/~bxc/APC/notes.pdf

Chiang, C.L. (1984) The life table and its applications. R.E. Krieger Publishing, Malabar, Fl.

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

Coale, A.J. and D. McNeil (1972) The distribution by age of the frequency of first marriage in a female cohort. *Journal of the American Statistical Association* 67:743-749.

Coleman, J.S. (1964) Introduction to mathematical sociology. The Free Press of Gencoe and Collier-Macmillan, London.

Coleman, J.S. (1990) Foundations of social theory. The Belknap Press of Harvard University Press, Cambridge, Mass.

Commenges D. 1999. "Multi-state models in epidemiology." *Lifetime Data Analysis*, 5:315-327

Conroy, R.M., K. Pyorala, A.P. Fitzgerald et al. (2003) Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *European Heart Journal*, 24:987-1003.

Cox, D.R. (1972) Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society (B)*. 34:187-220.

Crimmins,E.M., M.D. Hayward, and Y. Saito (1994) Changing mortality and morbidity rates and the health status and life expectancy of the older population. *Demography*, 31(1), pp. 159-75.

David, H.A. and M.L. Moeschberger (1978) The theory of competing risks. Charles Griffin & Co, London.

De Backer, G. and D. de Bacquer (1999) Lifetime-risk prediction: a complicated business. *The Lancet*, 353 (January 9): 82

Dias, J.G. and F.J. Willekens (2005) Model-based clustering of sequential data with an application to contraceptive use dynamics. *Mathematical Population Studies*, 12:135-157.

Director, S.W. and R.A. Rohrer (1972) Introduction to system theory. McGraw-Hill, New York.

Elbaz, A., J.H. Bower, D.M. Maraganore, S.K. McDonnell, B.J. Peterson, J.E. Ahlskog, D.J. Schaid and W.A. Rocca (2002) Risk tables for parkinsonism and Parkinson's disease. *Journal of Clinical Epidemiology*, 55:25-31.

Elder, G.H. Jr. (1985) Life course dynamics: trajectories and transitions, 1968-1980. Cornell University Press, Ithaca, New York. Extracts from Chapter 1 reprinted as 'Perspectives on the life course' in D.J. Bogue, E.E. Arriaga and D.L. Anderton eds. Readings in population research methodology. Social Development Center, Chicago, and UNFPA, New York, pp. 15.12-15.14.

Elder, G.H. Jr. (1999) The life course and aging; some reflections. Distinguished Scholar Lecture, American Sociological Association, August 1999.

Ellner, S.P. and M. Rees (2006) Integral projection models for species with complex demography. *The American Naturalist*, 167:410-428.

Fisher, R.A. (1930) The genetical theory of natural selection. Oxford University Press, Oxford. Republished in revised and enlarged version (2nd edition) by Dover Publishers, 1958, New York.

Flinn, C. and J. Heckman (1982a) New methods for analyzing structural models of labor force dynamics. *Journal of Econometrics*, 18:115-168.

Flinn, C. and J. Heckman (1982b) Models for the analysis of labor market dynamics. In: R. Bassmann and G. Rhodes eds. Advances in econometrics. Vol. 3, JAI Press, Greenwich, Conn.

Francis, B. and J. Pritchard (2000) Bertin, Lexis and the graphical representation of event histories. *Cybergeo. European Journal of Geography*, no. 152 (published 17/11/2000). Available at http://193.55.107.45/semiogra/brian/franjoh2.htm

Frank, W.A. (1998) Foundations of social evolution. Princeton University Press, Princeton, New Jersey.

Galler, H.P. (1997) Discrete-time and continuous-time approaches to dynamic microsimulaiton reconsidered. Technical Paper 13. National Centre for Social and Economic Modeling (NATSEM), University of Canberra [http://www.natsem.canberra.edu.au/pubs/tps/tp13/tp13.pdf]

Gardiner, C.W. (2004) Handbook of stochastic methods for physics, chemistry and the natural sciences. Springer Verlag, Berlin. ?? Edition (first edition 1983).

Giele, J.Z. and G.H. Elder Jr. (1998) Life course research. Development of a field. In: J.Z. Giele and G.H. Elder Jr. eds. Methods of life course research. Qualitative and quantitative approaches. Sage Publications, Thousand Oaks, Ca., pp. 5-27.

Gill, R.D. (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. (1992) Multistate life tables and regression models. *Mathematical Population Studies*, 3(4):259-276

Gill, R.D and N. Keilman (1990) On the estimation of multidimensional demographic models with population registration data. *Mathematical Population Studies*, 2(2):119-143.

Girosi, F. and G. King (2006) Demographic forecasting. Prepublication available at http://gking.harvard.edu/files/smooth.pdf

Godfrey, K. (1983) Compartmental models and their applications. Academic Press, London.

Goldman, D.P., P.G. Shekelle, J. Bhattacharya and others (2004) Health status and medical treatment of the future elderly. Final Report. Prepared for the Centers for Medicare and Medicaid Services. TR-169-CMS, Rand Corporation, Santa Monica. Available at: http://www.rand.org/labor/aging/pdfs/goldman_health.pdf

Goodman, L.A. (1967) On the reconciliation of mathematical theories of population growth. *Journal of the Royal Statistical Society A*, 130:541-553.

Goodman, L.A. (1968) An elementary approach to the population projection matrix, to the population reproductive value, and related topics in the mathematical theory of population growth. *Demography*, 5:382-409.

Greene, W.H. (1997) Econometric analysis. Third edition. Prentice-Hall, Upple Saddle River, New Jersey.

Gribble, S. (1997) LifePaths: A longitudinal microsimulation model using a synthetic approach. Paper presented in the conference "Microsimulation in Government Policy and Forecasting: International Conference on Combinatorics, Information Theory and Statistics" Maine Portland, Maine, USA, July 18-20, 1997

Haberman, S. and E. Pitacco (1999) Actuarial models for disability insurance. Chapman and Hall, London, and CRC, Boca Raton, Florida.

Hachen, D.S. (1988) The competing risk model. *Sociological Methods and Research*, 17(1):21-54. Reprinted in in D.J. Bogue, E.E. Arriaga and D.L. Anderton eds. Readings in population research methodology. Social Development Center, Chicago, and UNFPA, New York, pp. 21.85-21.101.

Halfon, N. and M. Hochstein (2002) Life course health development: an integrated freamework for developing health, policy, and research. *The Milbank Quarterly*, 80(3). Available at http://www/milbank.org/quarterly/8003feat.html

Hammel, E.A. (1990) *SOCSIM II*, Working Paper no. 29, Graduate Group in

Demography, University of California, Berkeley.

Hammel, E.A., D. Hutchinson, K. Wachter, R. Lundy, and R. Deuel (1976) The SOCSIM Demographic-Sociological Microsimulation Program Operating Manual. Institute of International Studies Research Monograph No. 27, University of California, Berkeley, California.

Heckman , J.J. and B. Singer (1982) Population heterogeneity in demographic models. In: K.C. Land and A. Rogers eds. Multidimensional mathematical demography. Academic Press, New York, pp. 567-599.

Heijdra, B.J. and W.E. Romp (2005) A life-cylce overlapping generations model for the small open economy. SOM Research Report No. 05C04, University of Groningen. Downloadable from www.rug.nl/economics/rompwe

Heligman, L. and J.H. Pollard (1980) The age pattern of mortality. *Journal of the Institute of Actuaries*, 107:49-80.

Hense, H.W. (2004) Observations, predictions and decisions. Assessing cardiovascular risk assessment. *International Journal of Epidemiology*, 33:235-239.

Hernes, G. (1972) The process of entry into first marriage. *American Sociological Review* 3337:11733-182.

Hizanidis, J. (2002) The master equation. Available at http://wwwnlds.physik.tu-berlin.de/~hizanidis/talks/mastermanu.pdf

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. New York: Academic Press, pp. 155-264.

Holmer M., A. Janney and B. Cohen (2006). PENSIM overview. Policy Simulation Group, U.S. Department of Labor, Washington, D.C.

Hougaard, P. (1999) Multi-state models: a review. *Lifetime Data Analysis*, 5(3):239-264.

Hougaard, P. (2000) Analysis of multivariate survival data. Springer Verlag, New York

Kalbfleisch, J.D. and J.F Lawless (1985) The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association*, 80(392):863-871

Kaneko, R. (2003) Elaboration of the Coale-McNeil Nuptiality Model as The Generalized Log Gamma Distribution: A New Identity and Empirical Enhancements. *Demographic Research*, Vol. 9-10:223-262. http://www.demographic-research.org/Volumes/Vol9/10/9-10.pdf

Keiding, N. (1999) Event history analysis and inference from observational epidemiology. *Statistics in Medicine*, 18:2353-2363

Keyfitz, N. (1968) Introduction to the mathematics of population. Addison-Wesley, Reading, Massachusetts.

Keyfitz, N. (1982) Kan knowledge improve forecasts? *Population and Development Review*, 8(4):729-751.

Keyfitz, N. (1985) Applied mathematical demography. Second Edition. Springer Verlag, New York.

Keyfitz, H. and H. Caswell (2005) Applied mathematical demography. Third edition. Springer, New York.

Keyfitz, B.L. and N. Keyfitz (1997) The McKendrick Partial Differential Equation and its Uses in Epidemiology and Population Study, *Mathematical and Computer Modelling*, 26(6): 1-9.

Klevmarken N. A. et P. Olovsson (1996). Direct and behavioral effects of income tax changes - simulations with the Swedish model MICROHUS, in: A. Harding ed. Microsimulation and Public Policy, North-Holland-Elsevier, Amsterdam.

Kono, S. (1987) The headship rate method for projecting households. In J. Bongaarts, T. K. Burch, and K. W. Wachter eds. Family demography: methods and their application. Oxford University Press, New York/Oxford, England, pp. 287-308

Kono, S (1993) Functional population projections. Editor's introduction. In: D.J. Bogue, E.E. Arriaga and D.L. Anderton eds. (1993) Readings in population research methodology. Social Development Center, Chicago, and UNFPA, New York, pp. 18.1 – 18.2.

Kuh, D.L. and Y. Ben-Shlomo (1997) A life course approach to chronic disease epidemiology. Tracing origins of ill-health from early to adult life. Oxford University Press, Oxford.

Kuh. D.L. and R. Hardy (2002) A life course approach to women's health. Oxford University Press, Oxford.

Laird, N. and D. Olivier (1981) Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association*, 76(374): 231-240.

Lancaster, T. (1990) The econometric analysis of transition data. Cambridge University Press, Cambridge.

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

Lee, R.D. (1998) Probabilistic approaches to population forecasting. *Population and Development Review. Supplement to Volume 24*, pp. 165-190.

Légaré, J. and N. Marcil-Gratton (1990) Individual programming of life events: a challenge for demographers in the twenty-first century. In: C. Maltoni and I.J. Selikoff eds. Scientific issues of the next century: convocation of world academies. *Annals of the New York Academy of Sciences*, 610: 99-105.

Liang, Z. (2000) The Coale-McNeil model. Theory, generalization and application. Thela Thesis, Amsterdam.

Liaw, K.-L. and J. Ledent (1980) Discrete approximation of a continuous model of multistate demography. Professional Paper PP-80-14, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Linke W. (1988) The Headship Rate Approach in Modelling Households: the Case of the Federal Republic of Germany. in Keilman N., Kuijsten A., Vossen A. eds. Modelling Household Formation and Dissolution, Clarendon Press, Oxford, pp. 108-122

Leslie P. H. (1945) On the use of matrices in certain population mathematics. *Biometrika*, 33, pp.183-212.

Lloyd-Jones, D.M., M.G. Larson, A. Beiser and D. Levy (1999) Lifetime risk of developing coronary heart disease. *The Lancet*, 353(January 9):89-92 (see also discussion in *The Lancet*, 353 (March 13):924-925.

Lotka, A.J. (1925) Elements of physical biology. William and Wilkins, Baltimore. Reproduced as *Elements of Mathematical Biology,* Dover, New York, 1956

Lutz, W., W. C. Sanderson, and S. Scherbov (1997) Doubling of world population unlikely. *Nature* 387: 803-805, 19 June 1997.

Maltz, M.D. and J.M. Mullany (2000) Visualizing lives: new pathways for analysing life course trajectories. *Journal of Quantitative Criminology*, 16(2):255-281.

Mamun, AA. (2003) Life history of cardiovascular disease and its risk factors. Rozenberg Publishers, Amsterdam.

Manotonino, A. and D. Young (2004) Investment in education: the implications for economic growth and public finances. Economic Papers no. 217. Directorate General for Economic and Financial Affairs, Economic Commission, Brussels. Available at: http://ec.europa.eu/economy_finance/publications/economic_papers/2004/ecp217en.pdf

Manton, K.G. and E. Stallard (1988) Chronic disease modeling: measurement and evaluation of the risks of chronic disease processes. Oxford University Press, New York.

Manton, K.G., B.H. Singer and R.M. Suzman (1993) The scientific and policy needs for improved health forecasting models for elderly populations. In: K.G. Manton, B.H. Singer and R.M. Suzman eds. Forecasting the health of the elderly population. Springer Verlag, New York, pp. 3-35.

Mathers, C.D. and J.M. Robine (1997) How good is Sullivan's method for monitoring changes in population health expectancies? *Journal of Epidemiology and Community Health*, 51(1): 80-86.

McKendrick, A.G. (1926) Application of mathematics to medical problems. *Proceedings of th Edinburgh Mathematical Society* (Edinburgh, Scotland), 44:98-130.
McLachlan, G.J. and Peel, D. (2000). Finite Mixture Models. Wiley, New York.

Metz, J.A.J. and O. Diekmann (1986) The dynamics of physiologically structured populations. Springer Verlag, Berlin.

Namboodiri, K. (1991) Demographic analysis. A stochastic approach. Academic Press, San Diego.

Namboodiri, K. and C.M. Suchindran (1987) Life table techniques and their applications. Academic Press, Orlando

Newman, S.C. (1988) A Markov process interpretation of Sullivan's index of morbidity and mortality. *Statistics in Medicine*, 7:787 - 794

O'Donoghue, C, (n.d.) Dynamic microsimulation. A methodological survey. http://www.beje.decon.ufpe.br/v4n2/cathal.pdf

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

Peeters, A., A.A. Mamun, F.J. Willekens and L. Bonneux (2002) A cardiovascular life course. A life course analysis of the original Framingham Heart Study cohort. *European Heart Journal*, 23 (2002), pp. 458- 466.

Pencina, M.J., M.G. Larson and R.B. D'Agostino (2006) Choice of time scale and its effect on significance of predictors in longitudinal studies. *Statistics in Medicine*, in press.

Pressat, R. (1995) Eléments de démographie mathématique. Association Internationale des Démographes de Langue Francaise, Paris.

Preston, S., P. Heuveline and M. Guillot (2001) Demography. Measuring and modelling population processes. Blackwell, Oxford.

Rajulton, F. (1999) LIFEHIST: Analysis of life histories: a state-space approach. Paper presented at the Workshop on Longitudinal Research in Social Science: A Canadian Focus, Windermere Manor, London, Ontario, Canada, October 25-27, 1999.

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

Rogers, A. (1986) Parameterized multistate population dynamics and projections. *Journal of the American Statistical Association*, 81(?):48-61. Reprinted in D.J. Bogue, E.E. Arriaga and D.L. Anderton eds. Readings in population research methodology. Social Development Center, Chicago, and UNFPA, New York, pp. 22.83-22.9

Rogers, A. (1995) Multiregional demography. Principles, methods and extensions. Wiley, Chichester

Rogers A. and L.J. Castro (1981) Model migration schedules. Research Report *RR-81-30*, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Rogers, A. and L. Castro (1986) Migration. In: A. Rogers and F. Willekens eds. Migration and settlement. A multiregional comparative study. D. Reidel Publishing Company, Dordrecht, The Netherlands, pp. 157-208.

Rogers, A. and F.J. Willekens (1978) The spatial reproductive value and the spatial momentum of zero population growth. *Environment and Planning A*, 10(5):503-518.

Rogers, A. and F.J. Willekens (1986) A short course in multiregional mathematical demography. In: A. Rogers and F.J. Willekens eds. Migration and settlement. A multiregional comparative study. Reidel Publishing Company, Dordrecht, pp. 355-384.

Rohwer, G. and U. Pötter (1999) TDA User's manual. Ruhr-Universität Bochum. Fakultät für Sozialwissenschaften, Bochum, Germany.

Romo, V.C. (2003) Decomposition methods in demography. Rozenberg Publishers, Amsterdam.

Ross, S.M. (2006) Simulation. Fourth edition. Elsevier/Academic Press, Amsterdam.

Ryder, N.B. (1965) The cohort as a concept in the study of social change. *American Sociological Review*, 30:843-861. Reprinted in W.M. Mason and S.E. Fienberg eds. (1985) Cohort analysis in social research. Beyond the identification problem. Springer Verlag, New York, pp. 9-44.

Sanderson, W. (1998) Knowledge can improve forecasts: a review of selected socioeconomic population projection models. *Population and Development Review*, Vol. 24, Supplement: Frontiers of Population Forecasting (1998) , pp. 88-117

Scherbov, S., A. Yashin, and V. Grechucha (1986) Dialog System for Modeling Multidimensional Demographic Processes, WP-86-29, Laxenburg, Austria: International Institute for Applied Systems Analysis

Schoen, R. (1988) Modeling multigroup populations. Plenum Press, New York.

Schoen, R. and K. 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.

Schouten, L.J., H. Straatman, L.A.L.M. Kiemeney and A.L.M. Verbeek (1994) Cancer incidence: life table risk *versus* cumulative risk. *Journal of Epidemiology and Community Health*, 48:596-600

Sen, A. and T. Smith (1995) Gravity models of spatial interaction behavior. Springer Verlag, Berlin.

Shieh, L.S., R. Yates and J. Navarro (1978) Representation of continuous time state equations by discrete-time equations. *IEEE*, 8(6):485-492.

Singer, B. and S. Spilerman (1979) Mathematical representations of development theories. In: J.R. Nesselroade and P.B. Baltes eds. Longitudinal research in the study of behavior and development. Academic Press, New York, pp. 155-177.

Smith, S. (1982) Tables of working life: the increment-decrement model. U.S. Government Printing Service for U.S. Department of Labour, Bureau of Labor Statistics, Washington.

Statistics Canada (2001). The LifePaths Microsimulation Model: An Overview. http://statcan.ca/english/spsd/LifePathsOverview_E.pdf

Strang, G. (1980) Linear algebra and its applications. Second edition. Academic Press, New York.

Taylor, H.M. and S. Karlin (1994) An introduction to stochastic modelling. Revised edition. Academic Press, San Diego.

Tuma, N. and M.T. Hannan (1984) Social dynamics: models and methods. Academic Press, New York.

Van de Vegte, J. (1994) Feedback control systems. Third edition. Prentice-Hall, Englewood Cliffs, New Jersey.

Van der Gaag, N., J. de Beer and F.J. Willekens (2005) MicMac. Combing micro and macro approaches in demographic forecasting. Paper presented at the Joint Eurostat-ECE Work Session on Demographic Projections, September 21-23, 2005, Vienna.

Van der Gaag, N., J. de Beer, P. Ekamper and F. Willekens (2006) Using MicMac to project living arrangements: an illustration of biographic projections, Paper presented at the European Population Conference 2006 'Population Challenges in Ageing Societies', 21-24 June 2006, Liverpool, the United Kingdom.

Vandeschrick, C. (2001) The Lexis diagram, a misnomer. *Demographic Research*, 4(3): 94-118.

Van Imhoff, E. (1990) The exponential multidimensional demographic projection model. *Mathematical Population Studies*, 2(3):171-182

Van Imhoff, E. and W. Post (1998), Microsimulation methods for population projection. *Population: An English Selection*, special issue *New Methodological Approaches in the Social Sciences*, 97-138.]

Van Imhoff, E. and N. Keilman (1991) *LIPRO 2.0: An application of a dynamic demographic projection model to household structure in the Netherlands*. Amsterdam: Sweets&Zeitlinger.

Vaupel, J.W. and V.C. Romo (2002) Decomposing demographic change into direct vs. compositional components. *Demographic Research*, 7(1): 1 - 14

Vermunt, J. (2002) A general latent class approach to unobserved heterogeneity in the analysis of event history data. http://spitswww.uvt.nl/~vermunt/hagenaars2002a.pdf

Von Förster, H. (1959) Some remarks on changing populations. In: F. Stohlman, Jr. ed. The kinetics of cellular proliferation. Greene and Stratton, New York.

Weidlich, W. and G. Haag (1983) Concepts and models of quantitative sociology: the dynamics of interacting populations. Springer Verlag, Berlin.

Weidlich, W. and G. Haag eds (1988) Interregional migration. Dynamic theory and comparative analysis. Springer Verlag, Berlin.

Willekens, F.J. (1977) The spatial reproductive value. Theory and applications. Rsearch Memorandum RM-77-9, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria.

Willekens, F.J. (1987) The marital status life-table. In: J. Bongaarts, T. Burch and K.W. Wachter eds. Family demography: models and applications. Oxford: Clarendon Press (for IUSSP), pp. 125-149.

Willekens F.J. (1990) Demographic forecasting: state-of-the-art and research needs. In C.A. Hazeu and C.A.B. Frinking eds., Emerging issues in demographic research, Elsevier Science, Amsterdam, pp. 9–66.

Willekens, F.J. (1995) MUDEA. Version 2.0. Manual and tutorial. Report. Faculty of Spatial Sciences, University of Groningen.

Willekens, F.J. (1998) Demographic projection models. A technical introduction. Manuscript.
Willekens, F.J. (2001) Gompertz in context: the Gompertz and related distributions. In: E. Tabeau, A, van den Berg Jeths and C. Heathcote eds. Forecasting mortality in developed countries. Insights from a statistical, demographic and epidemiological perspective. Kluwer Academic Publishers, Dordrecht, 2001, pp. 105 –126.

Willekens, F.J. (2002) Forecasting the life course. Paper presented at the Annual Meeting of the Population Association of America, Atlanta.

Willekens, F.J. (2004) Biographies. Real and synthetic. Manuscript.

Willekens, F.J. (2006) Continuous-time microsimulation. Manuscript.

Willekens, F.J. (2007) Financial demography. Mastering the financial consequences of life contingencies and demographic change. Paper presented at the Annual Meeting of the Population Association of America, New York, March 2007.

Willekens, F.J. and R. Hakkert (1992) Directory of demographic software. Outcome of the IUSSP Working Group on Demographic Software and Micro-computing), published in 1998 on the internet: http://www.unfpacst.cl (site United Nations Population Fund, CELADE, Chili).

Willekens, F.J. and A. Rogers (1978) Spatial population analysis. Methods and computer programs. IIASA, Research Report RR-78-18.

Willekens, F.J. and P. Drewe (1984) A Multiregional model for regional demographic projection. In: H. ter Heide and F. Willekens eds. Demographic research and spatial policy. London: Academic Press, 1984, pp. 309-334. Reprinted in: D.J. Bogue, E.E. Arriaga and D.L. Anderton eds. Readings in population research methodology. Social Development Center, Chicago, and UNFPA, New York, 1993.) pp. 17.49 - 17.57.

Willekens, F. and D. Philipov (1981) Dynamics of multiregional population systems: a mathematical analysis of the growth path. Working Paper no. 22, Voorburg: NIDI.

Willekens, F.J., J. de Beer and N. van der Gaag (2005) MicMac. From demographic to biographic forecasting. Paper presented at the Joint Eurostat-ECE Work Session on Demographic Projections, September 21-23, 2005, Vienna.

WHO (World Health Organization) (2002) Life course perspectives on coronary heart disease, stroke and diabetes. The evidence and implications for policy and research.

WHO/NMH/NPH/02.1, Department of Noncommunicable Disease Prevention and Health Promotion, World Health Organization, Geneva.

Wilson, W.G. (1996) Lotka's game in predator-prey theory: linking populations to individuals. *Theoretical Population Biology*, 50:368-393.

Wilson, T. and Ph. Rees (2005) Recent developments in population projection methodology: a review. *Population, Space and Place*, 11(5):337-360.

Wolf, D. (2000) The role of microsimulation in longitudinal data analysis. http://www.ssc.uwo.ca/sociology/longitudinal/wolf.pdf

Wun , L-M., R.M. Merrill and E.J. Feuer (1998) Estimating lifetime and age-conditional probabilities of developing cancer. *Lifetime Data Analysis*, 4(2):169-186.

Wunsch, G.J. and M.G. Termote (1978) Introduction to demographic analysis. Principles and methods. Plenum Press, New York.

Wunsch, G., M. Mouchart and J. Duchêne eds. (2002) The life table. Modelling survival and death. Kluwer Academic Publishers, Dordrecht.

Yaari, M. E. (1965) Uncertain lifetime, life insurance, and the theory of the consumer. *Review of Economic Studies*, 32:137–150.

Yamaguchi, K. (1991) Event history analysis. Sage Publications, Newbury Park, USA.

Zaidi, A. and K. Rake (2001) Dynamic microsimulation models: a review and some lessons for SAGE. SAGE Discussion Paper no. 2 [SAGEDP/02]. ESRC SAGE Research Group, The London School of Economics, London (available at www.lse.ac.uk/depts/sage)

Zeng Yi (1991) Family dynamics in China. A multistate life table analysis. Wisconsin University Press, Madison, WI.

Zeng Yi, J.W. Vaupel and W. Zhenglian (1997) A multidimensional model for projecting family households. With an illustrative numerical application. *Mathematical Population Studies*, 6(3):187-216

Zeuner G. (1869) Abhandlungen aus der Mathematischen Statistik. Verlag von Arthur Felix, Leipzig.

Annex A
Calculation of an exponential of a matrix

Note: If M is small (dt), P os approximately [I+dt M]

A number of methods exists to determine the value of the transition matrix or matrix of transition probabilities **P**(h) = exp[-h**M**], where **M** is the matrix of transition rates (see e.g. Director and Rohrer, 1972, pp. 431ff; Aoki, 1976, p. 387ff; Strang, 1980, p. 206). We describe four methods and provide a numerical illustration for a state space with three states. The three states may be North, Central and South in which case the transitions refer to migration and the transition rates are migration rates. Let

$$\mathbf{M} = \begin{bmatrix} m_{11} & -m_{21} & -m_{31} \\ -m_{12} & m_{22} & -m_{32} \\ -m_{13} & -m_{23} & m_{33} \end{bmatrix} = \begin{bmatrix} 0.14 & -0.02 & -0.05 \\ -0.10 & 0.10 & -.020 \\ -0.04 & -0.08 & 0.25 \end{bmatrix}$$

A.1 Taylor series expansion

The power series of exp[h] is

$$e = 1 + h + \frac{h^2}{2!} + \frac{h^3}{3!} + \frac{h^4}{4!} + \ldots = \sum_{j=0}^{\infty} \frac{h^j}{j!}$$

For h = 1, e = 2.71828

The function P**(h) = exp[-h**M**]** has a Taylor series expansion around 0:

$$\mathbf{P}(h) = \exp[-h\mathbf{M}] = \mathbf{I} - h\mathbf{M} + \tfrac{1}{2!}(h\mathbf{M})^2 - \tfrac{1}{3!}(h\mathbf{M})^3 + \ldots = \sum_{k=0}^{\infty} \frac{-(h\mathbf{M})^k}{k!}$$

where 0! = 1 and $\mathbf{M^0 = I}$.

The infinite series can be reduced to a polynomial of a degree equal to the rank of the matrix minus 1:

$$\exp[-h\mathbf{M}] = \alpha_0\mathbf{I} - \alpha_1 h\mathbf{M} + \alpha_2[h\mathbf{M}]^2 - \alpha_3[h\mathbf{M}]^3 + \ldots\ \alpha_{I-1}[h\mathbf{M}]^{I-1}$$

where I is the rank of the matrix.

P(h) = exp[-h**M**] is obtained using **M** given above and the Taylor series expansion, with h = 1:

$$\mathbf{P}(h) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - h \begin{bmatrix} 0.14 & -0.02 & -0.05 \\ -0.10 & 0.10 & -0.20 \\ -0.04 & -0.08 & 0.25 \end{bmatrix} + \frac{h^2}{2} \begin{bmatrix} 0.14 & -0.02 & -0.05 \\ -0.10 & 0.10 & -0.20 \\ -0.04 & -0.08 & 0.25 \end{bmatrix}^2$$

$$- \frac{h^3}{6} \begin{bmatrix} 0.14 & -0.02 & -0.05 \\ -0.10 & 0.10 & -0.20 \\ -0.04 & -0.08 & 0.25 \end{bmatrix}^3 + \ldots$$

The series converges at the 5-th power of **M**. At the 6-th power, the estimate of the matrix of transition probabilities is

$$\mathbf{P}(h) = \exp[-h\mathbf{M}] = \sum_{k=0}^{6} \frac{-(h\mathbf{M})^k}{k!} = \begin{bmatrix} 0.8712 & 0.0195 & 0.0430 \\ 0.0924 & 0.9127 & 0.1707 \\ 0.0364 & 0.0678 & 0.7863 \end{bmatrix}$$

Note that the columns indicate the states of origin and the rows the states of destination.

A.2 Linear approximation

The exponential **P**(h) = exp[-h**M**], which assumes that the transition rates are constant during the interval of length h, can be approximated by a linear function which implies that the events are uniformly distributed during the observation interval. In multistate demography, the linear approximation is the often used. It was originally developed by The linear model is

$$\mathbf{P}(h) = \exp[-h\mathbf{M}] \cong \left[\mathbf{I} + \frac{h}{2}\mathbf{M}\right]^{-1} \left[\mathbf{I} - \frac{h}{2}\mathbf{M}\right]$$

Using the matrix of transition rates given above, the linear approximation leads to

$$\mathbf{P}(h) = \begin{bmatrix} 1+\frac{1}{2}0.14 & -\frac{1}{2}0.02 & -\frac{1}{2}0.05 \\ -\frac{1}{2}0.10 & 1+\frac{1}{2}0.10 & -\frac{1}{2}0.20 \\ -\frac{1}{2}0.04 & -\frac{1}{2}0.08 & 1+\frac{1}{2}0.25 \end{bmatrix}^{-1} \begin{bmatrix} 1-\frac{1}{2}0.14 & \frac{1}{2}0.02 & \frac{1}{2}0.05 \\ \frac{1}{2}0.10 & 1-\frac{1}{2}0.10 & \frac{1}{2}0.20 \\ \frac{1}{2}0.04 & \frac{1}{2}0.08 & 1-\frac{1}{2}0.25 \end{bmatrix}$$

$$= \begin{bmatrix} 0.9354 & 0.0097 & 0.0217 \\ 0.0463 & 0.9561 & 0.0860 \\ 0.0183 & 0.0342 & 0.8923 \end{bmatrix} \begin{bmatrix} 0.9300 & 0.0100 & 0.0250 \\ 0.0500 & 0.9500 & 0.1000 \\ 0.0200 & 0.0400 & 0.8750 \end{bmatrix}$$

$$= \begin{bmatrix} 0.8709 & 0.0195 & 0.0433 \\ 0.0926 & 0.9122 & 0.1720 \\ 0.0366 & 0.0683 & 0.7847 \end{bmatrix}$$

A.3 Diagonalization of the matrix M

Assume **M** has distinct eigenvalues λ_i with i = 1, 2, . . . , I. **M** may be transformed into a diagonal matrix by the following expression:

$$\mathbf{M} = \mathbf{D}\boldsymbol{\Lambda}\mathbf{D}^{-1}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix of the distinct eigenvalues or characteristics roots of **M**. It is known as the *spectral matrix*. **D** is a square matrix with the right eigenvectors of **M** associated with each of the eigenvalues as its columns: $\mathbf{D} = [\boldsymbol{\xi}_1 \ \boldsymbol{\xi}_2 \ \boldsymbol{\xi}_3]$, where $\boldsymbol{\xi}_i$ is the right eigenvector associated with the i-th eigenvalue λ_i. The matrix **D** of eigenvectors is known as the *modal matrix*. It is a transformation matrix transforming **M** into a diagonal matrix. Since the eigenvalues are distinct, the eigenvectors, and hence the columns of **D**, are linearly independent. As a consequence, **D** is non-singular and the inverse exists. The inverse $\mathbf{D}^{-1}$ is $\mathbf{D}^{-1} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3]' = \mathbf{N}'$ is a matrix with the left eigenvectors (row vectors) as its rows. The left eigenvectors and the right eigenvectors are such their inner product is $\mathbf{v}_i' \boldsymbol{\xi}_i = 1$ and $\mathbf{v}_i' \boldsymbol{\xi}_j = 0$ for $i \neq j$. The vectors $\mathbf{v}_i'$ and $\boldsymbol{\xi}_i$ are normalized and $\mathbf{v}_i'$ and $\boldsymbol{\xi}_j$ $(j \neq i)$ are orthogonal.

The matrix of transition probabilities **P**(h) and the matrix of transition rates **M** have the same modal matrix. That can be shown using the Taylor series expansion. Note that $\left[\mathbf{D}\boldsymbol{\Lambda}\mathbf{D}^{-1}\right]^T = \mathbf{D}\boldsymbol{\Lambda}\mathbf{D}^{-1}\mathbf{D}\boldsymbol{\Lambda}\mathbf{D}^{-1}\mathbf{D}\boldsymbol{\Lambda}\mathbf{D}^{-1}\ldots = \mathbf{D}\boldsymbol{\Lambda}^T\mathbf{D}^{-1}$ (see also e.g. Strang, 1980, p. 206). Hence we may write

$$\mathbf{P}(h) = \exp[-h\mathbf{M}] = \exp\left[-h\mathbf{D}\boldsymbol{\Lambda}\mathbf{D}^{-1}\right] = \mathbf{D}\exp[-h\boldsymbol{\Lambda}]\mathbf{D}^{-1}$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues of **M**. The exponent of a diagonal matrix is simply the exponent of each element. Hence the i-th element of $\exp[-h\boldsymbol{\Lambda}]$ is $\exp[-h\,\lambda_i]$.

Consider the matrix **M** shown above. **M** has three distinct real eigenvalues (imaginary part of the complex number is zero). They are: $\lambda_1 = 0.0$, $\lambda_2 = 0.3226$, and $\lambda_3 = 0.1674$. The right eigenvectors associated with the eigenvalues are shown in Table B.1

Right eigenvectors $\boldsymbol{\xi}$(j) of **M** associated with the eigenvalues λ_j			
	Eigenvalue		
	λ_1	λ_2	λ_3
$\xi(1)$	0.2483	0.1415	0.7835
$\xi(2)$	0.9104	0.6256	-0.5908
$\xi(3)$	0.3310	-0.7672	-0.1927

The modal matrix **D** is obtained by entering the eigenvectors in the columns. Hence (h = 1):

$$\mathbf{P}(h) = \exp[-h\mathbf{M}] = \mathbf{D}\exp[-h\ \boldsymbol{\Lambda}\ ]\mathbf{D}^{-1}$$

$$= \begin{bmatrix} 0.2483 & 0.1415 & 0.7835 \\ 0.9104 & 0.6256 & -0.5908 \\ 0.3310 & -0.7672 & -0.1927 \end{bmatrix} \exp\left[-h \begin{bmatrix} 0.0000 & 0 & 0 \\ 0 & 0.3226 & 0 \\ 0 & 0 & 0.1674 \end{bmatrix} \right] \begin{bmatrix} 0.2483 & 0.1415 & 0.7835 \\ 0.9104 & 0.6256 & -0.5908 \\ 0.3310 & -0.7672 & -0.1927 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 0.2483 & 0.1415 & 0.7835 \\ 0.9104 & 0.6256 & -0.5908 \\ 0.3310 & -0.7672 & -0.1927 \end{bmatrix} \begin{bmatrix} 1.0000 & 0 & 0 \\ 0 & 0.7243 & 0 \\ 0 & 0 & 0.8459 \end{bmatrix} \begin{bmatrix} 0.6713 & 0.6713 & 0.6713 \\ 0.0235 & 0.3594 & -1.0060 \\ 1.0594 & -0.2776 & -0.0310 \end{bmatrix}$$

$$= \begin{bmatrix} 0.8712 & 0.0195 & 0.0430 \\ 0.0924 & 0.9127 & 0.1707 \\ 0.0364 & 0.0678 & 0.7863 \end{bmatrix}$$

The diagonalization of the matrix of transition rates gives the same transition probabilities as the Taylor series expansion. The linear model results in an approximation of the transition probabilities.

The results of the diagonalization may be used to derive a simple expression of the projection model. The method is known as the *spectral decomposition* of the trajectory of change (or growth) (see e.g. Strang, 1980, p. 207; Van de Vegte, 1994, pp. 337ff). Without loss of argument, we assume that the transition rates are constant. The population at time t is related to the population in the base period by the following equation:

$$\mathbf{k}(t) = \mathbf{P}(0,t) = \exp[-t\,\mathbf{M}]\mathbf{k}(0) = \mathbf{D}\exp[-t\ \mathbf{\Lambda}\ ]\mathbf{D}^{-1}\mathbf{k}(0)$$

where an element $k_i(t)$ of $\mathbf{k}(t)$ is the number of people in state i at time t. $\mathbf{P}(0,t)$ is the matrix of discrete-time transition probabilities between 0 an t. An element $p_{ij}(0,t)$ give the probability that an individual who occupies state i at 0 occupies state j at t. The projection model can be written as a linear combination of weighed sum of pure exponentials:

$$\mathbf{k}(t) = c_1 \exp[-\lambda_1 t]\xi_1 + c_2 \exp[-\lambda_2 t]\xi_2 + c_3 \exp[-\lambda_3 t]\xi_3$$

where λ_i is the i-th eigenvalue of $\mathbf{M}$, ξ_i is the right eigenvector associated with the eigenvalue λ_i and the coefficient c_i is an element of the vector $\mathbf{c} = \mathbf{D}^{-1}\ \mathbf{k}(0)$ that matches the initial condition. Hence the weights in the linear combination of pure exponentials depend on the initial condition. An element c_i is the weighted sum of the initial population, with weights equal to the elements of $\mathbf{v}_i$.

For an illustration, assume that everyone starts out in state 1, hence $k_1(0) = 1$ and $k_i(0) = 0$ for $i > 1$. The vector $\mathbf{c}$ of coefficients is $\mathbf{c} = \mathbf{D}^{-1}\ \mathbf{k}(0)$:

$$\mathbf{c} = \begin{bmatrix} 0.6713 & 0.6713 & 0.6713 \\ 0.0235 & 0.3594 & -1.0060 \\ 1.0594 & -0.2776 & -0.0310 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.6713 \\ 0.0235 \\ 1.0594 \end{bmatrix}$$

Because of the interpretation of $\mathbf{D}^{-1}$, c_1 is the product of the first left eigenvector of M and the initial population.

The projection model is

$$\mathbf{k}(t) = 0.6713\exp[-1.0000t]\begin{bmatrix}0.2483\\0.9104\\0.3310\end{bmatrix} + 0.0235\exp[-0.7243t]\begin{bmatrix}0.1415\\0.6256\\-0.7672\end{bmatrix} + 1.0594\exp[-0.8459t]\begin{bmatrix}0.7835\\-0.5908\\-0.1927\end{bmatrix}$$

For t = 1, $\mathbf{k}(1) = \begin{bmatrix}0.1667\\0.6112\\0.2222\end{bmatrix} + \begin{bmatrix}0.0024\\0.0106\\-0.0131\end{bmatrix} + \begin{bmatrix}0.7021\\-0.5294\\-0.1727\end{bmatrix} = \begin{bmatrix}0.8712\\0.0924\\0.0365\end{bmatrix}$

Note that for t = 0, an observed population structure, i.e. the observed distribution of the population between the states of the state space, is expressed as a sum of

In this simple illustration, the eigenvectors are real numbers. As a consequence, the population trajectory is monotonous. In many practical applications, the eigenvalues are complex numbers and the spectral decomposition identifies the waves that are characteristic for many demographic changes. Spectral decomposition has been used in the studies of regimes of demographic change and to characterize change as composed of a stable pattern and deviations from the stable pattern (Keyfitz, 1968, pp. 59ff; Keyfitz and Caswell, 2005, pp. 150ff; Willekens and Philipov, 1981). Bartholomew (1982, pp. 89 ff) applies the theory to study social change processes. The stable pattern is given by the right eigenvector associated with the first eigenvalue. If the vector is normed to a total of unity, it gives the relative distribution (in proportions of total) of the population in dynamic or steady state equilibrium. In mathematical demography hat population is known as the *stable population*. Departures from stability are covered by the contributions of the second and third eigenvalues. The three-state example shows that, although the entire population starts out in state 1 (North), in the long-run about two third will occupy state 2 (Central) and only 17 percent will reside in the North. That is entirely the result of migration. The contribution of the North to the population distribution in large initially, but declines over time when the Central gains in size. The large-term trend of Central is significant although its initial contribution is very small.

An element c_i is the weighted sum of the initial population, with weights equal to the elements of $\mathbf{v}_i$. Hence the elements of $\mathbf{v}_i$ measure the contribution of the population initially in state i to the population size in the long run, when the population has reached the dynamic equilibrium. That contribution is known as the *reproductive value* of the population in state i. The concept of reproductive value was introduced by Fisher (1930, 1958 edition, p. 27) in his study of natural selection and extinction of species to measure the extent to which subpopulations contribute to the ancestry of future generations. The concept was introduced in demography by Goodman (1967, 1968) and Keyfitz (1968) and was extended to multistate populations by Willekens (1977) and Rogers and

Willekens (1978). For recent reviews see Keyfitz and Caswell (2005, Chapters 8 and 9) in the field of demography and Frank (1998) in the study of social dynamics.

A.4 Laplace transform

A convenient way to evaluation **P**(h) = exp[-h**M**] is the Laplace transform. The Laplace transform is a method to solve the system of differential equations. Consider an open multistate population. In Section 8, the dynamics of a population of age x with entries (immigration) and exits (deaths and emigration) is described by the differential equation

$$\frac{d\,\mathbf{K}(x+\tau,t+\tau)}{d\,\tau} = -\boldsymbol{\mu}(x,t)\,\mathbf{K}(x,t) + \mathbf{I}_m(x,x+h,t)$$

where **K**(x,t) is a vector of state occupancies for individuals of age x at time t. $\mathbf{I}_m$(x,x+h,t) is the vector of immigrants aged x to x+h at t during the *unit* interval from t to t+1 by state of existence. The diagonal elements of the matrix of transition intensities $\boldsymbol{\mu}$(x,t) includes the instantaneous rates of transition to outside the population system:

$$\mu_{ii}(x,t) = \mu_{id}(x,t) + \sum_{j\neq i}^{I} \mu_{ij}(x,t) + \mu_{io}(x,t)$$

where μ_{id}(x,t) is the instantaneous death rate in state i for individuals aged x and born at t-x and μ_{io}(x,t) the instantaneous rate at which individuals aged x in state i leave the population system during the small interval from t to t+dτ (emigration rate).

The solution to the system of equations is

$$\mathbf{K}(x+h,t+h) = \exp\left[-h\mathbf{M}(x,x+h,t)\right]\mathbf{K}(x,t) + \int_0^h \exp\left[-(h-\tau)\mathbf{M}(x,x+h,t)\right]h\,\mathbf{I}_m(x,x+h,t)\,d\tau$$

The system of equations may be written in a more general form as

$$\frac{d\mathbf{K}(t)}{dt} = -\boldsymbol{\mu}(t)\,\mathbf{K}(t) + \mathbf{F}(t)\,\mathbf{u}(t)$$

where K(t) is the state vector, representing the population at time t and u(t) is the vector of immigrants during the unit interval. F(t) is a coefficient matrix, which in this case is the identity matrix. We assume that $\boldsymbol{\mu}$(t) and **F**(t) are constant during the interval from 0 to t (where t may be replaced by h to denote age intervals).

The solution is

$$\mathbf{K}(t) = \exp\left[-t\,\mathbf{M}\right]\mathbf{K}(0) + \int_0^t \exp\left[-(t-\tau)\mathbf{M}\right]\mathbf{F}\,\mathbf{u}(\tau)\,d\tau \qquad \text{(A.1)}$$

The Laplace transform of a function of time f(t) is defined as:

$$La\left[f(t)\right] = F(s) \cong \int_0^\infty \exp\left[-st\right]f(t)\,dt$$

where s is the complex variable σ+jω with σ the real part and jω the imaginary part and j is the square root of –1. The Laplace transform of the system of differential equations is (see e.g. van de Vegte, 1994, pp. 330ff, Godfrey, 1983, pp. 26ff)

$$La\left[\frac{d\mathbf{K}(t)}{dt}\right] = \int_0^\infty \exp[-st]\frac{d\mathbf{K}(t)}{dt}dt$$

where the function f(t) is $\left[\frac{d\mathbf{K}(t)}{dt}\right]$.

Integration by parts gives

$$La\left[\frac{d\mathbf{K}(t)}{dt}\right] = [\mathbf{K}(t)\exp(-st)]_0^\infty + \int_0^\infty s\exp[-st]\frac{d\mathbf{K}(t)}{dt}dt$$

$$= -\mathbf{K}(0) + s\,{}^*\mathbf{K}(s)$$

where ${}^*\mathbf{K}(s)$ is the Laplace transform of the function **K**(t).

The Laplace transform of -**μ K**(t) is $La[-\boldsymbol{\mu}\mathbf{K}(t)] = -\boldsymbol{\mu}\,{}^*\mathbf{K}(s)$ and $La[\mathbf{F}\mathbf{u}(t)] = \mathbf{F}\,\mathbf{U}(s)$
Hence the Laplace transform of the system of differential equations is

$$-\mathbf{K}(0) + s\,{}^*\mathbf{K}(s) = -\boldsymbol{\mu}\,{}^*\mathbf{K}(s) + \mathbf{F}\,\mathbf{U}(s)$$
$$[s\mathbf{I} + \boldsymbol{\mu}]\,{}^*\mathbf{K}(s) = \mathbf{K}(0) + \mathbf{F}\,\mathbf{U}(s)$$

which gives

$${}^*\mathbf{K}(s) = [s\mathbf{I} + \boldsymbol{\mu}]^{-1}\mathbf{K}(0) + [s\mathbf{I} + \boldsymbol{\mu}]^{-1}\mathbf{F}\,\mathbf{U}(s)$$

The first term gives the impact (response) on the population of the initial condition and the second term gives the impact of the input to the system which in this case is immigration. Having found the Laplace transform F(s) of a function f(t), the solution of the function f(t) in the time domain is obtained by inverse Laplace transformation La^{-1}. Comparing the Laplace solution with equation A.1, it may be seen that

$$\exp[-t\mathbf{M}] = La^{-1}\left[(s\mathbf{I} + \boldsymbol{\mu})^{-1}\right]$$

and

$$\int_0^t \exp[-(t-\tau)\mathbf{M}]\mathbf{F}\mathbf{u}(\tau)\,d\tau = La^{-1}\left[(s\mathbf{I} + \boldsymbol{\mu})^{-1}\mathbf{F}\,\mathbf{U}(s)\right] = La^{-1}[\mathbf{G}(s)\,\mathbf{U}(s)]$$

where **G**(s) is the transfer function matrix which relates the input vector (immigration) to the output vector (population) for zero initial condition **K**(0)=**0**.

Consider a simple 2 * 2 matrix

$$M = \begin{bmatrix} 0.10 & -0.05 \\ -0.10 & 0.05 \end{bmatrix}$$

$[s\mathbf{I}+\boldsymbol{\mu}] = \begin{bmatrix} s+0.10 & -0.05 \\ -0.10 & s+0.05 \end{bmatrix}$ and $[s\mathbf{I}+\boldsymbol{\mu}]^{-1} = \frac{adj[s\mathbf{I}+\boldsymbol{\mu}]}{|s\mathbf{I}+\boldsymbol{\mu}|}$ with $|s\mathbf{I}+\boldsymbol{\mu}|$ the determinant of the s**I**+**μ**. The determinant is (s+0.10)(s+0.05) – (-0.10)(-0.05) = s(s+0.15).

Hence

$$[s\mathbf{I}+\boldsymbol{\mu}]^{-1} = \frac{1}{s(s+0.15)}\begin{bmatrix} s+0.05 & 0.05 \\ 0.10 & s+0.10 \end{bmatrix} = \begin{bmatrix} \frac{1/3}{s}+\frac{2/3}{s+0.15} & \frac{1/3}{s}-\frac{1/3}{s+0.15} \\ \frac{2/3}{s}-\frac{2/3}{s+0.15} & \frac{2/3}{s}+\frac{1/3}{s+0.15} \end{bmatrix}$$

The inverse Laplace transform of a/(s+b) is $La^{-1}\frac{a}{s+b} = s\exp[-bt]$. Hence

$$\exp[-t\mathbf{M}] = La^{-1}\left[(s\mathbf{I}+\boldsymbol{\mu})^{-1}\right] = \begin{bmatrix} \frac{1}{3}+\frac{2}{3}\exp[-0.15t] & \frac{1}{3}-\frac{1}{3}\exp[-0.15t] \\ \frac{2}{3}-\frac{2}{3}\exp[-0.15t] & \frac{2}{3}+\frac{1}{3}\exp[-0.15t] \end{bmatrix}$$

which, for t=1, is

$$\exp[-t\mathbf{M}] = \begin{bmatrix} 0.9071 & 0.0464 \\ 0.0929 & 0.9536 \end{bmatrix}$$

When t gets large, the exponent tends to zero and the multistate population reaches the steady-state equilibrium (stable population). In the steady-state, 1/3 of the population is in state 1 and 2/3 is in state 2.

The result of the Laplace transform is identical to the one obtained by the Cayley-Hamilton theorem that any matrix satisfies its characteristic equation and the diagonalization of the matrix **M**.

Note that the linear approximation is

$$[\mathbf{I}+0.5\mathbf{M}]^{-1}[\mathbf{I}-0.5\mathbf{M}] = \begin{bmatrix} 0.9070 & 0.0465 \\ 0.0930 & 0.9535 \end{bmatrix}$$

Annex B
The linear model

The exponential model may be approximated by the linear model. The linear model is the dominant method in demography and actuarial sciences and is sometimes referred to as the *actuarial approach*. The linear model postulates a uniform distribution of events during age and time intervals. The postulate is referred to as the *linear integration hypothesis*. In this paper, the intervals are of equal length h. The uniform distribution of events implies a particular variation in the transition intensities during the interval. Consider the differential equation that applies to individual k born at instantaneous time t-x:

$$\frac{d\,{}_k^{\tau}\mathbf{P}(x,t)}{d\tau} = -\boldsymbol{\mu}(x+\tau,t+\tau)\;{}_k^{\tau}\mathbf{P}(x,t)$$

or the equivalent expression

$$\frac{d\,{}_k\mathbf{P}(x,x+\tau,t)}{d\tau} = -\boldsymbol{\mu}(x+\tau,t+\tau)\;{}_k\mathbf{P}(x,x+\tau,t)$$

The equation describes the changes in transition probabilities for individual k when the interval of length τ increases marginally by the amount dτ. In the linear model, that change is linear. The solution to the differential equation may be written as a linear model (Alho and Spencer, 2005, p. 170) (the index k and the argument t are omitted for convenience):

$$\mathbf{P}(x+\tau) = \left[\mathbf{I} + \tau\,\mathbf{B}(x)\right]$$

where **B**(x) will be determined soon.

$$\frac{d\mathbf{P}(x+\tau)}{d\tau} = -\boldsymbol{\mu}(x+\tau)\,\mathbf{P}(x+\tau)$$

Hence

$$\frac{d\left[\mathbf{I} + \tau\,\mathbf{B}(x)\right]}{d\tau} = \mathbf{B}(x) = -\boldsymbol{\mu}(x+\tau)\left[\mathbf{I} + \tau\,\mathbf{B}(x)\right]$$

Hence $\boldsymbol{\mu}(x+\tau) = -\mathbf{B}(x)\left[\mathbf{I} + \tau\,\mathbf{B}(x)\right]^{-1}$ provided the inverse exists. If τ=0, **μ**(x) = -**B**(x), hence **B**(x) = -**μ**(x). The elements of **B**(x) are minus the densities of transition at the start of the interval. Recall that the off-diagonal elements of **μ**(x) are negative.

To illustrate the method, consider a single state. The linear model is $p(x) = 1 - bx$, the differential equation is $\frac{d\,p(x)}{dx} = -\mu(x)\,p(x)$ and $\frac{d\,p(x)}{dx} = \frac{d\,[1-bx]}{dx} = -b$. Hence the transition intensities vary with age according to the following equation: $\mu(x) = \frac{b}{1-bx}$

The transition rates at the beginning of the interval are usually not known, but the average rates ${}^{h}\mathbf{m}(x)$ are. They are obtained empirically by dividing the number of direct transitions during the (x,x+h)-interval by the total duration of exposure. If the total duration of exposure is approximated by the product of the width of the interval and the number of individuals born at t-x and in a particular state at mid-interval, then the average rate is the *central rate*. If the (direct) transitions are assumed to be uniformly distributed over the interval from x to x+h, then

$${}^{h}\mathbf{m}(x,) = \boldsymbol{\mu}(x+\tfrac{1}{2}h) = -\mathbf{B}(x)\left[\mathbf{I}+\tfrac{1}{2}h\mathbf{B}(x)\right]^{-1}$$

From that equation, we may derive **B**(x). Since ${}^{h}\mathbf{m}(x)\left[\mathbf{I}+\tfrac{1}{2}h\mathbf{B}(x)\right]=-\mathbf{B}(x)$,

$\left[\mathbf{I}+\tfrac{1}{2}h^{h}\mathbf{m}(x)\right]\mathbf{B}(x)=-{}^{h}\mathbf{m}(x)$ and $\mathbf{B}(x)=-\left[\mathbf{I}+\tfrac{1}{2}h\,{}^{h}\mathbf{m}(x)\right]^{-1}{}^{h}\mathbf{m}(x)$

Suppose the interval h is 10 years and the central rate is 0.2. Then $b=\dfrac{0.2}{1+\frac{1}{2}*10*0.2}=0.1$. Under the linear model with a central transition rate of 0.2, the transition rate varies from 0.1 at the start of the interval to 1 near the end and infinity at the end of the interval. The transition rate follows a hyperbolic function (Schoen and Land, 1979, p. 767).

The value of **P**(x+τ) may now be determined:

$$\mathbf{P}(x+\tau)=\mathbf{I}+\tau\mathbf{B}(x)=\mathbf{I}-\tau\left[\mathbf{I}+\tfrac{1}{2}h^{h}\mathbf{m}(x)\right]^{-1}{}^{h}\mathbf{m}(x)$$

$$\mathbf{P}(x+\tau)=\left[\mathbf{I}+\tfrac{1}{2}h\,{}^{h}\mathbf{m}(x)\right]^{-1}\left[\mathbf{I}+\left(\tfrac{1}{2}h-\tau\right){}^{h}\mathbf{m}(x)\right]$$

If τ = h, then

$$\mathbf{P}(x+h)=\left[\mathbf{I}+\tfrac{1}{2}h\,{}^{h}\mathbf{m}(x)\right]^{-1}\left[\mathbf{I}-\tfrac{1}{2}h\,{}^{h}\mathbf{m}(x)\right]$$

which is the linear approximation of the exponential model and is one of the fundamental equations of multistate demography.

The state probabilities at exact age x and time t are

$${}_{k}\mathbf{k}(x+h)={}^{h}_{k}\mathbf{P}(x)\;{}_{k}\mathbf{k}(x)=\left[\mathbf{I}+\tfrac{1}{2}h\,{}^{h}\mathbf{m}(x)\right]^{-1}\left[\mathbf{I}-\tfrac{1}{2}h\,{}^{h}\mathbf{m}(x)\right]{}_{k}\mathbf{k}(x)$$

The linear approximation implies the assumption that the events are uniformly distributed over the interval. That assumption is adequate when the transition rates are small or the interval is short, one year, say. Hoem and Funck-Jensen (1982, pp. 198-200) point out shortcomings of the linear approximation. The linear model is usually used in demography. For instance, the conventional estimation of the life table relies on the uniform distribution of events during an age interval of one or five years. Sometimes, such as in the first age group, the assumption of uniform distribution of event is seriously challenging the data. In those cases, other assumptions are made (see Keyfitz, 1968; Namboodiri and Suchindran, 1987) or smaller intervals are used, intervals of a month or smaller can be used. Oechsli (1975) followed that approach using spline functions.

Different assumptions imply different relations between transition rates and transition probabilities. It can be shown that the linear model is an approximation to the exponential model that retains the first three terms of the Taylor series expansion (Annex C).

Because of the assumption of uniform distribution of events, the expected sojourn times in the different states are given by the following equation:

$${}^{h}_{k}\mathbf{L}(x) = \int_0^h {}^{\tau}\mathbf{P}(x)\,d\tau = \int_0^h [\mathbf{I} + \tau\,\mathbf{B}(x)]\,d\tau = \left|\tau\right|_0^h \mathbf{I} + \mathbf{B}(x)\left|\tfrac{1}{2}\tau^2\right|_0^h = h\,\mathbf{I} + \tfrac{1}{2}\mathbf{B}(x)\,h^2$$

which is equal to $\frac{h}{2}\left[\mathbf{I} + {}^{h}_{k}\mathbf{P}(x,t)\right]$

and

$${}^{h}_{k}\mathbf{L}(x) = \tfrac{h}{2}\left[\mathbf{I} + \left[\mathbf{I} + \tfrac{h}{2}\,{}^{h}_{k}\mathbf{m}(x)\right]^{-1}\left[\mathbf{I} - \tfrac{h}{2}\,{}^{h}_{k}\mathbf{m}(x)\right]\right] = h\left[\mathbf{I} + \tfrac{h}{2}\,{}^{h}_{k}\mathbf{m}(x)\right]^{-1}$$

since $\mathbf{I} = \left[\mathbf{I} + \tfrac{h}{2}\,{}^{h}_{k}\mathbf{m}(x)\right]^{-1}\left[\mathbf{I} + \tfrac{h}{2}\,{}^{h}_{k}\mathbf{m}(x)\right]$.

Form the above equation, **B**(x) may be expressed in terms of discrete-time transition probabilities: $\mathbf{I} + h\,\mathbf{B}(x) = {}^{h}_{k}\mathbf{P}(x)$. Hence $\mathbf{B}(x) = \tfrac{1}{h}[\mathbf{P}(x) - \mathbf{I}]$

The linear model may also be derived from the integral equation (we introduce the subscripts k and t again):

${}^{h}_{k}\mathbf{P}(x,t) = \mathbf{I} - \int_0^h {}_{k}\boldsymbol{\mu}(x+\tau, t+\tau)\,{}^{\tau}_{k}\mathbf{P}(x,t)\,d\tau$. In the linear model, the variation of the transition probabilities between x and x+h is approximated by a linear function:

$${}^{h}_{k}\mathbf{L}(x,t) = \int_0^h {}^{\tau}_{k}\mathbf{P}(x,t)\,d\tau = \frac{h}{2}\left[\mathbf{I} + {}^{h}_{k}\mathbf{P}(x,t)\right].$$

Introducing this expression in the flow equation (1) gives

$${}^{h}_{k}\mathbf{P}(x,t) = \mathbf{I} - \frac{h}{2}\,{}^{h}_{k}\mathbf{m}(x,t)\left[\mathbf{I} + {}^{h}_{k}\mathbf{P}(x,t)\right]$$

$${}^{h}_{k}\mathbf{P}(x,t) = \mathbf{I} - \frac{h}{2}\,{}^{h}_{k}\mathbf{m}(x,t)\,\mathbf{I} - \frac{h}{2}\,{}^{h}_{k}\mathbf{m}(x,t)\,{}^{h}_{k}\mathbf{P}(x,t)$$

$${}^{h}_{k}\mathbf{P}(x,t) + \frac{h}{2}\,{}^{h}_{k}\mathbf{m}(x,t)\,{}^{h}_{k}\mathbf{P}(x,t) = \mathbf{I} - \frac{h}{2}\,{}^{h}_{k}\mathbf{m}(x,t)$$

$${}^{h}_{k}\mathbf{P}(x,t) = \left[\mathbf{I} + \frac{h}{2}\,{}^{h}_{k}\mathbf{m}(x,t)\right]^{-1}\left[\mathbf{I} - \frac{h}{2}\,{}^{h}_{k}\mathbf{m}(x,t)\right]$$

Annex C
The linear model as an approximation of the exponential model

Using Taylor series expansion, it can be shown that the linear model is an approximation of the exponential model. Three methods are considered

a. Method 1

The exponent exp[-h**M**] can be written as

$$\exp[-h\mathbf{M}] = \mathbf{I} - h\mathbf{M} + \tfrac{1}{2}(h\mathbf{M})^2 - \tfrac{1}{6}(h\mathbf{M})^3 + ... = \sum_{k=0}^{\infty} \frac{-(h\mathbf{M})^k}{k!}$$

Let k(t) denote the state vector at time t. Then

$$\mathbf{k}(t+h) = \exp[-h\mathbf{M}]\mathbf{k}(t)$$

The first two terms of the Taylor series expansion gives

$$\mathbf{k}(t+h) = [\mathbf{I} - h\mathbf{M}]\mathbf{k}(t)$$

The approximation can be improved by first premultiplying the exponential state equation by exp[h**M**/2] and then expanding to obtain (Keyfitz and Caswell, 2005, p. 450):

$$\left[\mathbf{I} + \frac{h}{2}\mathbf{M}\right]\mathbf{k}(t+h) = \left[\mathbf{I} - \frac{h}{2}\mathbf{M}\right]\mathbf{k}(t)$$

which gives the linear model

$$\mathbf{k}(t+h) = \left[\mathbf{I} + \frac{h}{2}\mathbf{M}\right]^{-1}\left[\mathbf{I} - \frac{h}{2}\mathbf{M}\right]\mathbf{k}(t)$$

b. Method 2

The exponent exp[-h**M**] can be written as

$$\exp[-h\mathbf{M}] = 1 - h\mathbf{M} + \tfrac{1}{2}(h\mathbf{M})^2 - \tfrac{1}{6}(h\mathbf{M})^3 + ... = \sum_{k=0}^{\infty} \frac{-(h\mathbf{M})^k}{k!}$$

The geometric progression of $[\mathbf{I} + \tfrac{1}{2}h\mathbf{M}]^{-1}$ is

$$[\mathbf{I} + \tfrac{1}{2}h\mathbf{M}]^{-1} = \mathbf{I} - \tfrac{1}{2}h\mathbf{M} + \tfrac{1}{4}(h\mathbf{M})^2 - \tfrac{1}{8}(h\mathbf{M})^3 + ... \text{ provided that } \left|\tfrac{1}{2}h\mathbf{M}\right| < 1$$

$$[\mathbf{I} + \tfrac{1}{2}h\mathbf{M}]^{-1}[\mathbf{I} - \tfrac{1}{2}h\mathbf{M}] = \mathbf{I} - h\mathbf{M} + \tfrac{1}{2}(h\mathbf{M})^2 - \tfrac{1}{4}(h\mathbf{M})^3 + \tfrac{1}{8}(h\mathbf{M})^4 - .... = \sum_{k=0}^{\infty} \frac{-(h\mathbf{M})^k}{(k-1)!}$$

with (-1)! = 1

c. Method 3

Liaw and Ledent (1980) show that a method developed in engineering for the discrete approximation of continuous-time state equations may be applied to show the relation between the exponential model and the linear model. It is the Matrix Continued Fraction (MCF) method developed by Shieh et. (1978). To make the MCF method transparent, Liaw and Ledent consider the expansion of the number 1.2345 into a continued fraction:

$$1.2345 = 1 + \frac{2345}{10000} = 1 + \frac{1}{100000/2345} = 1 + \frac{1}{4 + \frac{620}{2345}} = ...$$

After a few divisions, one gets

$$1.2345 = 1 + \cfrac{1}{4 + \cfrac{1}{3 + \cfrac{1}{1 + \cfrac{1}{3 + ...}}}} = H_1 + \left[H_2 + \left[H_3 + \left[H_4 + ... \right]^{-1}\right]^{-1}\right]^{-1}$$

The retention of the first few H_j results in a fairly good approximation of the original number. For instance, the retention of the first three H_j gives the number 1.2308.

Application of the MCF method to approximate exp[-(y-x)M(x,y)] gives

$$\exp[-(y-x)\mathbf{M}(x,y)] = \mathbf{H}_1 + \left[\mathbf{H}_2 + \left[\mathbf{H}_3 + \left[\mathbf{H}_4 + ... \right]^{-1}\right]^{-1}\right]$$

Shieh et al. (1978) show that

$$\mathbf{H}_1 = \mathbf{I}, \ \mathbf{H}_2 = [-(y-x)\mathbf{M}(x,y)]^{-1}, \ \mathbf{H}_3 = -2\mathbf{I}, \ \mathbf{H}_4 = [-3(y-x)\mathbf{M}(x,y)]^{-1}, \ \mathbf{H}_5 = 2\mathbf{I}$$

Let G_j be the estimate of exp[-(y-x)M(x,y)] by retaining only the first j H matrices. Then we get

$$\mathbf{G}_2 = \mathbf{H}_1 + [\mathbf{H}_2]^{-1} = \mathbf{I} - (y-x)\mathbf{M}(x,y)$$

$$\mathbf{G}_3 = \mathbf{H}_1 + \left[\mathbf{H}_2 + [\mathbf{H}_3]^{-1}\right]^{-1} = [\mathbf{H}_2\mathbf{H}_3 + \mathbf{I}]^{-1}[\mathbf{H}_1\mathbf{H}_2\mathbf{H}_3 + \mathbf{H}_1 + \mathbf{H}_3]$$

$$\mathbf{G}_3 = \left[\mathbf{I} + \frac{y-x}{2}\mathbf{M}(x,y)\right]^{-1}\left[\mathbf{I} - \frac{y-x}{2}\mathbf{M}(x,y)\right]$$

which is the linear approximation of the exponential model. The linear model is therefore obtained by retention of the first three H_j in the MCF method.

There is a difference between ignoring the higher H_j in the MCF method and disregarding the tail of the Taylor series expansion. Shieh et al. (1978) observe that

$$\mathbf{G}_3 = \mathbf{I} + \left[-(y-x)\mathbf{M}(x,y)\right] + \frac{1}{2!}\left[-(y-x)\mathbf{M}(x,y)\right]^2 + \sum_{j=3}^{\infty} \frac{1}{2^{j-1}}\left[-(y-x)\mathbf{M}(x,y)\right]^j$$

which differs from the Taylor series expansion.