ACTUARIAL RESEARCH CLEARING HOUSE
1993 VOL. 1

Modelling Multi-State Processes using a Markov Assumption

Bruce L. Jones
Department of Statistics and Actuarial Science
The University of Iowa
Iowa City, IA 52242

January 1993

Abstract

Many areas of actuarial work involve situations which are conveniently viewed in terms of multi-state processes. Often an individual's presence in a particular state, or movement from one state to another, has some financial impact. The task of the actuary is then to quantify this impact, allowing for the stochastic nature of the process. The use of a Markov assumption in modelling these processes has the advantages of parsimony, mathematical tractability and ease of parameter estimation.

In this paper, we show how probabilities and actuarial values may be calculated using a time-homogeneous Markov model. Piecewise constant transition intensities are suggested as a way to extend the approach to the inhomogeneous case. In the event that the Markov assumption is found to be inappropriate, the state space can be modified as an alternative to assuming a more general stochastic process.

1 Introduction

It is very appealing intuitively to use multi-state processes to aid in the understanding of actuarial problems. We assume that the process occupies one of a finite number of states at any given point in time. Quantities of interest then depend on the state of the process over time. The state may indicate the health status of an individual, the presence or absence of certain risk factors, the cause of death of an individual, or the members of a group of lives that are still alive.

The simplest situation involves only two states and one possible transition, as shown in Figure 1. This could be used when considering life insurance or annuity products. In the case of a simple whole life insurance policy, the policy is issued to an individual in state 1. Premiums are payable while the insured is in state 1, and the death benefit is payable upon transition to state 2. For a life annuity, a single premium is paid when the contract is issued to an individual in state 1. Benefits are then payable until transition to state 2. The probabilistic behavior of this process depends only on the distribution of the time until death. Of course, when dealing with life insurance, we usually wish to consider withdrawals as well as deaths, and may be interested in the cause of death.

Figure 1: **Example of Two-State Model**

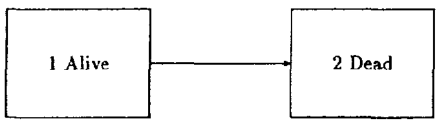
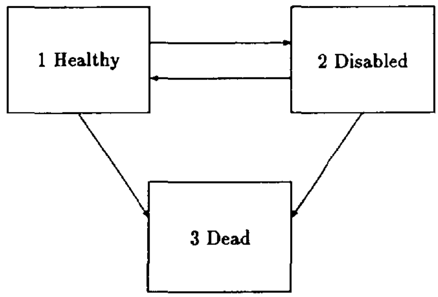


Figure 2 illustrates a three-state process which might be used in analyzing disability income insurance. Here, a policy is issued to an individual in state 1. Premiums are payable while in state 1, and benefits are payable while in state 2 (usually after a waiting period). This situation is more complicated than the two-state case since we allow individuals to cycle between states 1 and 2.

There are many other situations in which multi-state processes can be helpful. For instance, in modelling long-term care, we may have several levels of care which we wish to consider. We could also use a multi-state process to describe movement amongst risk categories such as smoking status (see Tolley and Manton 1991). In analyzing automobile insurance, we could use a multi-state process with states which represent the the insured's driving record.

Many authors have considered actuarial applications of multi-state processes. In much of this work it has been assumed that the process satisfies the Markov property. Under this assumption, Hoem (1969 and 1988) generalized a number of standard results from life contingencies. The stochastic properties of the profit earned on an insurance policy were

Figure 2: **Example of Three-State Model**



examined by Ramlau-Hansen (1988a and b). Tolley and Manton (1991) proposed models for morbidity and mortality which include various risk factors in the model state space. In modelling the mortality of individuals infected with the HIV virus, Panjer (1988) and Ramsay (1989) used a Markov process with states which represent the stage of infection. The development of formulas for probabilities and the estimation of parameters were discussed by Waters (1984). More general multi-state processes have been considered by Hoem (1972), Hoem and Aalen (1978), Seal (1970) and Waters (1989 and 1990).

This paper investigates how we can obtain results of interest for multi-state processes by taking advantage of the tractability of the time-homogeneous Markov model. In Section 2, we introduce the Markov assumption and examine some of the properties of the Markov process. Section 3 considers the calculation of actuarial values. In Section 4, we discover the advantage of the time-homogeneity or constant intensity assumption. We relax this assumption, but retain its benefits in Section 5 by allowing the transition intensity functions to be piecewise constant. In Section 6, we look at how we can alter the state space of a semi-Markov model in order to create a Markov model which adequately describes the process of interest. The results of the previous sections may then be used.

2 The Markov Assumption

Consider a process $\{X(t), t \geq 0\}$ with state space $\{1, 2, \ldots, k\}$. Thus, $X(t)$ represents the state of the process at time t. $\{X(t), t \geq 0\}$ is a Markov process (or continuous-time Markov chain) if, for all $s, t \geq 0$ and $i, j, x(u) \in \{1, 2, \ldots, k\}$,

$$\begin{aligned} & \Pr\{X(s+t) = j | X(s) = i, X(u) = x(u), 0 \leq u < s\} \\ = & \Pr\{X(s+t) = j | X(s) = i\}. \end{aligned}$$

Thus, the future of the process (after time s) depends only on the state at time s and not on the history of the process up to time s. We define

$$p_{ij}(s, s+t) \equiv \Pr\{X(s+t) = j | X(s) = i\},$$

and assume that

$$\sum_{j=1}^{k} p_{ij}(s, s+t) = 1 \text{ for all } t \geq 0.$$

We also assume the existence of the transition intensity functions

$$\mu_{ij}(t) = \lim_{h \to 0^+} \frac{p_{ij}(t, t+h) - \delta_{ij}}{h},$$

where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

The functions $\mu_{ij}(t)$ correspond to the forces of decrement considered in multiple decrement theory.

It is easily seen that

$$p_{ij}(s, s+t+u) = \sum_{l=1}^{k} p_{il}(s, s+t) p_{lj}(s+t, s+t+u), \tag{1}$$

for $i, j \in \{1, 2, \ldots, k\}$ and $s, t, u \geq 0$. These are known as the Chapman-Kolmogorov equations.

The transition intensity functions and the transition probability functions are related by the Kolmogorov forward and backward equations. These are

$$\frac{\partial}{\partial t} p_{ij}(s, s+t) = \sum_{l=1}^{k} p_{il}(s, s+t) \mu_{lj}(s+t), \tag{2}$$

and

$$\frac{\partial}{\partial s} p_{ij}(s, s+t) = -\sum_{l=1}^{k} \mu_{il}(s) p_{lj}(s, s+t), \tag{3}$$

respectively, with boundary conditions $p_{ij}(s, s) = \delta_{ij}$. Further properties and examples of continuous-time Markov chains are discussed by Ross (1983, ch. 5).

The main advantage of the Markov assumption is its simplicity. In stochastic modelling, we seek the simplest model which reasonably describes the process under consideration. Such a model is more convenient mathematically and easier to fit to data than more complicated models.

The reasonableness of the Markov assumption depends somewhat on the level of detail in the state description. For example, consider the three-state process shown in Figure 2.

In this case, the Markov assumption may be inappropriate. The future health of a recently disabled individual is likely to differ from that of someone who has been disabled for a long period of time. In Section 6, we suggest a way to get around this problem.

Often the data available does not allow us to assume a more general process. If transition times are unknown, we may have to assume a Markov process. In the insurance context, data is generally fairly good. Here there is a mechanism in place for keeping track of certain transitions (i.e. claims administration).

3 Actuarial Values

Consider a general policy with a continuous annual cash flow of $b_j(t)$ for an individual in state j at time t and lump sum cash flows of $b_{ij}(t)$ payable upon transition from i to j at time t. Amounts may be either positive or negative. We are interested in the random variable $\Gamma_i(s)$, representing the present value of future cash flows for an individual in state i at time s. We find that the expected value of this quantity is

$$\begin{aligned} E[\Gamma_i(s)] &\equiv V_i(s) \\ &= \sum_{j=1}^{k} \int_s^\infty v^{t-s} p_{ij}(s,t) b_j(t) dt \qquad (4) \\ &\quad + \sum_{l \neq j} \sum_{j=1}^{k} \int_s^\infty v^{t-s} p_{ij}(s,t) \mu_{jl}(t) b_{jl}(t) dt. \end{aligned}$$

From Ramlau-Hansen (1988a), the variance is

$$Var[\Gamma_i(s)] = \sum_{l \neq j} \sum_{j=1}^{k} \int_s^\infty v^{2(t-s)} p_{ij}(s,t) \mu_{jl}(t) [V_l(t) + b_{jl}(t) - V_j(t)]^2 dt. \qquad (5)$$

This general policy does not allow for the possibility that an amount becomes payable after the policyholder has been in a particular state for a given period of time. This is usually the case with disability income insurance. In such a situation, expressions for the expected value and variance of the present value of future cash flows are somewhat more complicated.

In order to apply (4) and (5) we must obtain the transition probability functions, $p_{ij}(s,t)$. As explained by Waters (1984), this involves the numerical solution of a system of simultaneous differential equations. In the very special case of transition intensity functions which are constant with respect to time, explicit expressions for the transtion probability functions can be obtained. These expressions become extremely complicated as the number of states increases. However, in the next section we describe an approach in which concise representation of the transition probability functions is possible even when the number of states is large.

4 Time-Homogeneous Markov Processes

In this section, we assume that $\mu_{ij}(t) = \mu_{ij}$ for all t. Such a Markov process is referred to as time-homogeneous or stationary. The assumption of constant intensities implies that the time spent in each state is exponentially distributed . Also, the functions $p_{ij}(s, s+t)$ are the same for all s. Thus,

$$p_{ij}(s, s+t) \equiv p_{ij}(t).$$

It is convenient to express the transition intensity and transition probability functions in matrix form. Let Q be the $k \times k$ matrix with (i, j) entry μ_{ij} and $P(t)$ be the $k \times k$ matrix with (i, j) entry $p_{ij}(t)$. Corresponding to (1), the Chapman-Kolmogorov equations are given by

$$P(t+u) = P(t)P(u). \tag{6}$$

Also, corresponding to (2) and (3), the Kolmogorov differential equations may be written

$$P'(t) = P(t)Q \tag{7}$$

and

$$P'(t) = QP(t). \tag{8}$$

Equations (7) and (8) have the solution

$$\begin{aligned} P(t) &= e^{Qt} \\ &= I + Qt + \frac{Q^2t^2}{2!} + \cdots \end{aligned}$$

This is of limited use since the series may converge rather slowly. However, as noted by Cox and Miller (1965), if Q has distinct eigenvalues, $d_1, \ldots, d_k$, then $Q = ADC$ where $C = A^{-1}$, $D = \mathrm{diag}(d_1, \ldots, d_k)$ and the ith column of A is the right eigenvector associated with d_i. Furthermore,

$$P(t) = A\,\mathrm{diag}(e^{d_1t}, \ldots, e^{d_kt})\,C. \tag{9}$$

Therefore, the problem of finding the transition probability functions is reduced to a problem of determining the eigenvalues and eigenvectors of the transition intensity matrix, Q. Software to perform this task is readily available.

The requirement that Q have distinct eigenvalues imposes no practical restriction. In the situations we consider, this will be the case for almost all parameter values (exceptions are a set of measure zero).

From (9), we may now write

$$p_{ij}(t) = \sum_{n=1}^{k} a_{in}c_{nj}e^{d_nt}, \tag{10}$$

where a_{ij} and c_{ij} are the (i, j) entries of A and C, respectively. The calculation of actuarial values such as (4) and (5) is easily accomplished using transition probability functions obtained from (10).

5 Piecewise Constant Intensities

In the previous section, it was assumed that transition intensities are constant with respect to time. This permits convenient representation of the transition probability functions. Unfortunately, in many actuarial applications, this is impractical. We require intensities which vary with age. We can accomplish this, while preserving the tractability of constant intensities, by using intensity functions which are piecewise constant.

Let $\mu_{ij}(t) = \mu_{ij}^{(m)}$ if $t \in (t_{m-1}, t_m]$, for $m = 1, 2, \ldots$, where $t_0 = 0$. Also, let $p_{ij}^{(m)}(t)$ be the transition probability functions associated with time intervals contained in $(t_{m-1}, t_m]$. In matrix form, we have $Q^{(m)}$ and $P^{(m)}(t)$. Now define m_t to be the time interval which contains time t. Then from (6), we have

$$P(s,t) = P^{(m_s)}(t_{m_s} - s)P^{(m_s+1)}(t_{m_s+1} - t_{m_s}) \cdots P^{(m_t)}(t - t_{m_t-1}).$$

Thus, given s and t, the transition probability matrix can be computed. We first determine $A^{(m)}$, $D^{(m)}$ and $C^{(m)} = (A^{(m)})^{-1}$ as described in Section 4. Furthermore, if we let $H^{(m)} = P(s, t_{m-1})A^{(m)}$, then for $t \in (t_{m-1}, t_m]$,

$$p_{ij}(s,t) = \sum_{n=1}^{k} h_{in}^{(m)} c_{nj}^{(m)} e^{d_n^{(m)}(t - t_{m-1})}.$$

6 Modification of the State Space

In some situations, the Markov assumption is clearly inappropriate for modelling a process. An example is select mortality. Here the mortality rate depends on the time that an individual has been in the insured state, as well as the age of the individual. Another example arises in disability insurance. The rate of disability termination is typically larger for those more recently disabled. In these cases, a semi-Markov assumption can be used. The future of a semi-Markov process depends on the time since entry to the current state. Hoem (1972) discussed applications of semi-Markov processes.

An alternative way to reflect this duration dependence is to treat each state as a collection of one or more substates. We then assume that the Markov property holds for the process indicating the substate occupied by an individual at each point in time, but not for the original process. For example, we might assume that the "insured" state in a life insurance situation consists of two substates: "insured, insurable" and "insured, uninsurable." This was suggested by Norberg (1988) as a way of explaining select mortality. Norberg showed that, if

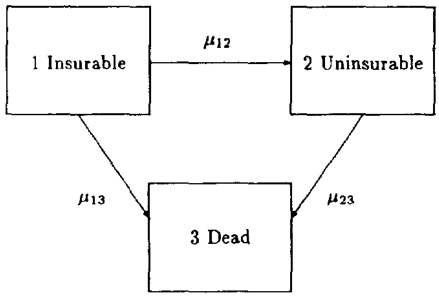
1. only insurable lives may enter the insured state
2. insurable lives may become uninsurable
3. uninsurable lives may not return to the insurable state

4. *the force of mortality for insurable lives is less than that for uninsurable lives*

then, for a fixed age, the force of mortality increases with duration since becoming insured. Møller (1990) showed that the result also holds if assumption 3 is relaxed, and explored the selection effect using more than one uninsurable state.

To examine this model more closely, consider the setup shown in Figure 3. Suppose that

Figure 3: **Model for Select Mortality**



the μ_{ij} represent transition intensities for some age group. As a function of the duration since issue of the insurance policy, the force of mortality may be written

$$\begin{aligned} \mu(t) &= \frac{e^{-(\mu_{12}+\mu_{13})t}\mu_{13} + \int_0^t e^{-(\mu_{12}+\mu_{13})x}\mu_{12}e^{-\mu_{23}(t-x)}dx\mu_{23}}{e^{-(\mu_{12}+\mu_{13})t} + \int_0^t e^{-(\mu_{12}+\mu_{13})x}\mu_{12}e^{-\mu_{23}(t-x)}dx} \\ &= \frac{(\mu_{23}-\mu_{13})(\mu_{12}+\mu_{13})e^{-(\mu_{12}+\mu_{13})t} - \mu_{12}\mu_{23}e^{-\mu_{23}t}}{(\mu_{23}-\mu_{13})e^{-(\mu_{12}+\mu_{13})t} - \mu_{12}e^{-\mu_{23}t}} \end{aligned} \quad (11)$$

Thus, μ_{12}, μ_{13} and μ_{23} should be chosen so that (11) best represents the selection effect for this age group.

We find that, for any choice of the three parameter values, there is a second choice which produces exactly the same $\mu(t)$. That is, if $\mu_{12} = \hat{\mu}_{12}$, $\mu_{13} = \hat{\mu}_{13}$, and $\mu_{23} = \hat{\mu}_{23}$, then we can achieve the same $\mu(t)$ by letting $\mu_{12} = \tilde{\mu}_{12}$, $\mu_{13} = \tilde{\mu}_{13}$, and $\mu_{23} = \tilde{\mu}_{23}$, where $\tilde{\mu}_{12} = \hat{\mu}_{23} - \hat{\mu}_{13}$, $\tilde{\mu}_{13} = \hat{\mu}_{13}$, and $\tilde{\mu}_{23} = \hat{\mu}_{12} + \hat{\mu}_{13}$. If we restrict our attention to the subset of the parameter space for which $\mu_{23} > \mu_{12} + \mu_{13}$ or $\mu_{23} < \mu_{12} + \mu_{13}$, then, for each $\mu(t)$, the parameterization is unique. In the absence of prior information about the parameter values, an arbitrary choice of subset may be made. Our objective is simply to find the best $\mu(t)$ based on this three-state setup. Ordinarily, we have no data on the three transitions shown in Figure 3, but only on transitions from states 1 and 2 combined to state 3.

References

[1] Cox, D.R., Miller, H.D. (1965). *The Theory of Stochastic Processes.* Chapman and Hall, London.

[2] Hoem, J.M. (1969). Markov Chain Models in Life Insurance. *Blätter der Deutschen Gesellschaft für Versicherungsmathematik* **IX**, 91–107.

[3] Hoem, J.M. (1972). Inhomogeneous Semi-Markov Processes, Select Actuarial Tables, and Duration-Dependence in Demography. Pp. 251–296 in T.N.E. Greville (ed.), *Population Dynamics.* Academic Press, New York.

[4] Hoem, J.M. (1988). The Versatility of the Markov Chain as a Tool in the Mathematics of Life Insurance. *Record of Proceedings, International Congress of Actuaries.* Helsinki, Finland, 171–202.

[5] Hoem, J.M., Aalen, O.O. (1978). Actuarial Values of Payment Streams. *Scandinavian Actuarial Journal* **61**, 38–47

[6] Møller, C. M. (1990). Select Mortality and Other Durational Effects Modelled by Partially Observed Markov Chains. *Scandinavian Actuarial Journal* 1990, 177–199.

[7] Norberg, R. (1988). Select Mortality: Possible Explanations. *Record of Proceedings, International Congress of Actuaries.* Helsinki, Finland, 215–224.

[8] Panjer, H.H. (1988). AIDS: Survival Analysis of Persons Testing HIV+. *Transactions of the Society of Actuaries* **XL**, Part I, 517–530.

[9] Ramsay, C.M. (1989). AIDS and the Calculation of Life Insurance Functions. *Transactions of the Society of Actuaries* **XLI**, 393–422.

[10] Ramlau-Hansen, H. (1988a). Hattendorf's Theorem: A Markov Chain and Counting Process Approach. *Scandinavian Actuarial Journal* 1988, 143–156.

[11] Ramlau-Hansen, H. (1988b). The Emergence of Profit in Life Insurance. *Insurance: Mathematics and Economics* **7**, 225–236.

[12] Ross, S.M. (1983). Stochastic Processes. Wiley, New York.

[13] Seal, H.L. (1970). Probability Distributions of Aggregate Sickness Durations. *Skandinavisk Aktuarietidskrift* **53**, 193–204.

[14] Tolley, H.D., Manton, K.G. (1991). Intervention Effects Among a Collection of Risks. *Transactions of the Society of Actuaries* **XLIII**, 117–142.

[15] Waters, H.R. (1984). An Approach to the Study of Multiple State Models. *Journal of the Institute of Actuaries* Vol. 111 part II, 363–374.

[16] Waters, H.R. (1989). Some Aspects of the Modelling of Permanent Health Insurance. *Journal of the Institute of Actuaries* Vol. 116 part III, 611–624.

[17] Waters, H.R. (1990). The Recursive Calculation of the Moments of the Profit on a Sickness Insurance Policy. *Insurance: Mathematics and Economics* **9**, 101–113.