Multistate Models in Health Insurance

Marcus Christiansen

Preprint Series: 2010-15

Fakultät für Mathematik und Wirtschaftswissenschaften
UNIVERSITÄT ULM

Multistate models in health insurance

Marcus C. Christiansen
Institut für Versicherungswissenschaften
Universität Ulm
D-89081 Ulm, Germany

Abstract

We illustrate how multistate Markov and semi-Markov models can be used for the actuarial modeling of health insurance policies, focussing on health insurances that are pursued on a similar technical basis to that of life insurance. On the basis of the general modeling framework of Helwich (2008), we study examples of permanent health insurance, critical illness insurance, long term care insurance, and German private health insurance and discuss the calculation of premiums and reserves on the safe side. In view of the rising popularity of stochastic mortality rate modeling, we present a theoretical foundation for stochastic transition rates and explain why there is a need for future statistical research.

Key Words: health insurance; Markov jump process; semi-Markov process; prospective reserve; actuarial calculation on the safe side; systematic biometric risk; worst case scenario

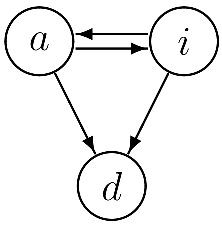
1 Introduction and motivation

Health insurances provide financial protection in case of sickness or injury by covering medical expenses or loss of earnings. Throughout the world there are various types of health insurance products, and there are many different traditions when it comes to their actuarial calculation. All of the calculation techniques are

- either pursued on a similar technical basis to that of *life insurance*
- or pursued on a similar technical basis to that of *non-life insurance*.

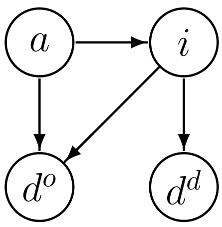
The multistate modeling approach that we are presenting here belongs to the first group. That means that we assume that (despite some financial risks) the only source of randomness is a random pattern of (finitely many) states that the policyholder goes through during the contract period. Claim sizes and occurrence times of all health insurance benefits just depend on that random pattern of states. At first let us consider four small examples that illustrate the fundamental ideas of multistate modeling in health insurance.

Example 1.1 (disability insurance)**.** A *disability insurance* or *permanent health insurance* (PHI) provides an insured with an income if the insured is prevented from working by disability due to sickness or injury. It is usually modeled by a multiple state model with state space $\mathcal{S} := \{a = \text{active/healthy}, i = \text{invalid/disabled}, d = \text{dead}\}$.

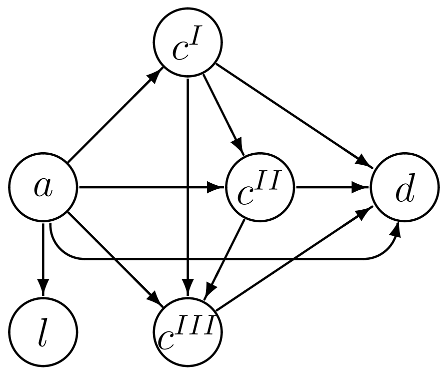


Note that in some countries, e.g. Germany, disability insurance is rather categorized as life or pension insurance than health insurance.

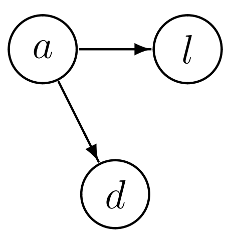
Example 1.2 (critical illness insurance)**.** A *critical illness insurance* or *dread disease insurance* (DD) provides the policyholder with a lump sum if the insured contracts an illness included in a set of diseases specified by the policy conditions. The most commonly diseases are heart attack, coronary artery disease requiring surgery, cancer and stroke. For example, it can be modeled by a multi state structure with state space $\mathcal{S} := \{a = \text{active/healthy}, i = \text{ill}, d^d = \text{dead due to dread disease }, d^o = \text{dead due to other causes}\}$.



Example 1.3 (long term care insurance)**.** A *long term care insurance* (LTC) provides financial support for insured who are in need of nursing or medical care. The need for care due to the frailty of an insured is classified according to the individuals ability to take care of himself by performing activities of daily living such as eating, bathing, moving around, going to the toilet, or dressing. LTC policies are commonly modeled by multistate models, and the state space usually consists of the states active, dead, and the corresponding levels of frailty. For example, in Germany three different levels of frailty are used and, moreover, lapse plays an important role. Thus, we have a state space of $\mathcal{S} = \{a = \text{active/healthy}, c^I = \text{need for basic care}, c^{II} = \text{need for medium care}, c^{III} = \text{need for comprehensive care}, l = \text{lapsed/canceled}, d = \text{dead}\}$.



Example 1.4 (German private health insurance)**.** A *German private health insurance* primarily covers actual medical expenses of the policyholder. In fact, the individual future medical expenses are unknown and, thus, stochastic. However, German private health insurers use deterministic forecasts for individual medical expenses that only depend on the age of the policyholder, so that the state space consists only of $\mathcal{S} := \{a = \text{alive}, l = \text{lapsed/canceled}, d = \text{dead}\}$. There is no differentiation made between different types and levels of morbidity, and so the only source of randomness is the time of death or lapsation, whichever occurs first.



For health insurances pursued on a similar technical basis to that of life insurance (the type of products we are studying here), the contractual guaranteed payments between insurer and policyholder are defined as deterministic functions of time and of the pattern of states of the policyholder. From a mathematical point of view, an insurance contract is just the set of that deterministic payment functions. Beside the specification of the contract terms (the payment functions), one of the main tasks of an actuary is to formulate a stochastic model for the random pattern of states of the policyholder. In many cases it is reasonable to assume that the random pattern of states is a Markovian process. The Markovian property reduces complexity and leads to an easily manageable model. However, in some cases we have significant durational effects, that is, the duration of stay in certain states has an effect on the likeliness of future transitions between states. For example, for disabled insured both the probability of recovering and the probability of dying usually significantly decrease with increasing duration of disability (see, for example, Segerer (1993)). In such cases the Markovian assumption clearly oversimplifies matters. The literature shows that then the best compromise between accuracy and feasibility is a semi-Markovian model. That is a model where (1) the present state of the policyholder and (2) the actual duration of stay in the present state together form a two-dimensional Markovian process. Nearly all multistate health insurance models presented in the literature fit into that semi-Markovian framework. (An interesting exception is the concept of Davis (1984).) By inserting the pattern of states process into the payment functions, we obtain the random future cash flow between insurer and policyholder. The job of the actuary is then to analyze this cash flow with the objective of determining premiums, solvency reserves, portfolio values, profits and losses, and so on.

For the interested reader we recommend the monograph of Haberman and Pitacco (1999), which gives a comprehensive survey of actuarial modeling of disability insurance, critical illness cover, and long-term care insurance. A detailed overview of actuarial modeling of private health insurance in Germany is given in Milbrodt (2005). In the insurance literature, the Markovian multistate model first appeared in Amsler (1968) and Hoem (1969). Since then the literature offers a range of papers that study this model, most of them under the topic 'life insurance'. Comprehensive presentations of the Markovian model and lots of further references can be found in the monographs of Wolthuis (1994), Milbrodt and Helbig (1999), and Denuit and Robert (2007). The semi-Markovian approach is much less studied in the literature, and we are not aware of a comprehensive monograph here. For the reader who is interested in the mathematical details of the semi-Markovian multistate modeling, we recommend Nollau (1980), Janssen and De Dominicis (1984), and Helwich (2008). In the insurance literature the semi-Markovian approach first appeared in Janssen (1966) and Hoem (1972). Further references are, for example, Waters (1989), Møller (1993), Segerer (1993), Gatenby and Ward (1994), Möller and Zwiesler (1996), Rickayzen and Walsh (2000), and Wetzel and Zwiesler (2003).

2 Random pattern of states

For health insurances pursued on a similar technical basis to that of life insurance, contractual guaranteed payments between insurer and policyholder are defined as deterministic functions of time and of the pattern of states of the policyholder. Therefore, at first we need a model for that pattern of states.

Definition 2.1. The *random pattern of states* is a pure jump process $(\Omega, \mathfrak{F}, P, (X_t)_{t\geq 0})$ with finite state space $\mathcal{S}$ and right continuous paths with left-hand limits, where X_t represents the state of the policyholder at time $t \geq 0$.

We further define the transitions space $J := \{(i,j) \in \mathcal{S} \times \mathcal{S} | i \neq j\}$, the counting processes

$$N_{jk}(t) := \#\big\{\tau \in (0,t] \,\big|\, X_\tau = k, X_{\tau-} = j\big\}\,, \quad (j,k) \in J\,,$$

the time of the next jump after t

$$T(t) := \min\big\{\tau > t \,\big|\, X_\tau \neq X_t\big\}\,,$$

the series of the jump times

$$S_0 := 0\,, \quad S_n := T(S_{n-1})\,, \quad n \in \mathbb{N}\,,$$

and a process that gives for each time t the time elapsed since entering the current state,

$$U_t := \max\big\{\tau \in [0,t] \,\big|\, X_u = X_t \text{ for all } u \in [t-\tau, t]\big\}\,.$$

Instead of using a jump process $(X_t)_{t\geq 0}$, some authors describe the random pattern of states by a chain of jumps. The two concepts are equivalent.

2.1 The semi-Markovian approach

The random pattern of states $(X_t)_{t\geq 0}$ is called *semi-Markovian*, if the bivariate process $(X_t, U_t)_{t\geq 0}$ is a Markovian process, which means that for all $i \in \mathcal{S}$, $u \geq 0$, and $t \geq t_n \geq ... \geq t_1 \geq 0$ we have

$$P\big((X_t, U_t) = (i,u) \big| X_{t_n}, U_{t_n}, ..., X_{t_1}, U_{t_1}\big) = P\big((X_t, U_t) = (i,u) \big| X_{t_n}, U_{t_n}\big)$$

almost surely. In the following we always assume that the initial state (X_0, U_0) is deterministic. (Note that $U_0 = 0$ by definition.) In practice that means that we know the state of the policyholder when signing the contract. With this assumption and the Markov property of $(X_t, U_t)_{t\geq 0}$ we have that the probability distribution of $(X_t, U_t)_{t\geq 0}$ is already uniquely defined by the *transition probability matrix*

$$p(s,t,u,v) := \Big(P(X_t = k, U_t \leq v \,|\, X_s = j, U_s = u)\Big)_{(j,k)\in\mathcal{S}^2}\,, \qquad 0 \leq u \leq s \leq t < \infty\,, \, v \geq 0\,.$$

Practitioners usually prefer the notation

$$_{v,t-s}p^{jk}_{x+s,u} := p_{jk}(s,t,u,v)\,,$$

where x is the age of the policyholder at contract time zero. Alternatively, we can also uniquely define the probability distribution of $(X_t, U_t)_{t\geq 0}$ by specifying the probabilities

$$\begin{aligned}
&\overline{p}(s,t,u) = \Big(\overline{p}_{jk}(s,t,u)\Big)_{(j,k)\in\mathcal{S}^2}\,,\\
&\overline{p}_{jk}(s,t,u) := P(T(s) \leq t, X_{T(s)} = k \,|\, X_s = j, U_s = u)\,, \quad j \neq k\,,\\
&\overline{p}_{jj}(s,t,u) := P(T(s) \leq t \,|\, X_s = j, U_s = u)
\end{aligned}$$

for $0 \le u \le s \le t < \infty$. For $j = k$ the corresponding actuarial notation is

$$ {}_{t-s}p^{jj}_{x+s,u} := 1 - \overline{p}_{jj}(s,t,u) = P(T(s) > t \,|\, X_s = j, U_s = u)\,. $$

A third way to uniquely define the probability distribution of $(X_t, U_t)_{t\ge 0}$ is to specify the *cumulative transition intensity matrix*

$$ \begin{aligned} q(s,t) &= \Big(q_{jk}(s,t)\Big)_{(j,k)\in\mathcal{S}^2}\,, \\ q_{jk}(s,t) &:= \int_{(s,t]} \frac{\overline{p}_{jk}(s,\mathrm{d}\tau,0)}{1-\overline{p}_{jj}(s,\tau-,0)}\,, \qquad 0 \le s \le t < \infty\,. \end{aligned} \tag{2.1} $$

If $q(s,t)$ is differentiable with respect to t, we can also define the so-called *transition intensity matrix*

$$ \mu(t,t-s) := \frac{\mathrm{d}}{\mathrm{d}t} q(s,t) = \bigg(\frac{\frac{\mathrm{d}}{\mathrm{d}t}\overline{p}_{jk}(s,t,0)}{1-\overline{p}_{jj}(s,t,0)}\bigg)_{(j,k)\in\mathcal{S}\times\mathcal{S}}\,, $$

which is some form of multi-state hazard rate. The quantity $\mu_{jk}(t, t-s)$ gives the rate of transitions from state j to state k at time t given that the current duration of stay in j is $t-s$. The corresponding actuarial notation is

$$ \mu^{jk}_{x+t,u} := \mu_{jk}(t,u)\,. $$

If the transition intensity matrix μ exists, it also already uniquely defines the probability distribution of $(X_t, U_t)_{t\ge 0}$. In order to obtain the transition probabilities from the cumulative transition intensities, we can use the *Kolmogorov backward integral equation system*

$$ \begin{aligned} p_{ik}(s,t,u,v) - \mathbf{1}_{\{i=k\}}\,\mathbf{1}_{\{v\ge u+t-s\}} &= \sum_{j:j\neq i} \int_{(s,t]} p_{jk}(\tau,t,0,v)\, q_{ij}(s-u,\mathrm{d}\tau) \\ &\quad + \int_{(s,t]} p_{ik}(\tau,t,u+\tau-s,v)\, q_{ii}(s-u,\mathrm{d}\tau)\,. \end{aligned} \tag{2.2} $$

We will base all our actuarial calculations on the cumulative transition intensity matrix $q(s,t)$ and the probabilities $\overline{p}_{jj}(s,t,u)$, where the latter can be derived from the former by the so-called exponential formula

$$ 1 - \overline{p}_{ii}(s,t,u) = e^{q^c_{ii}(s-u,t) - q^c_{ii}(s-u,s)} \prod_{s<\tau\le t} \Big(1 + q_{ii}(s-u,\tau) - q_{ii}(s-u,\tau-)\Big)\,. \tag{2.3} $$

Here, $q^c_{ii}(s-u,t)$ is the continuous part of $t \mapsto q_{ii}(s-u,t)$. Thus, for our actuarial calculations following later on, we have to specify $q(s,t)$, and only $q(s,t)$. The following concepts are frequently used in practice.

(a) We estimate the $\mu^{jk}_{x+t,u} = \mu_{jk}(t,u)$ from statistical data, often by using a parametric method (cf. Haberman and Pitacco (1999), section 4). The cumulative transition intensities $q(s,t)$ are obtained by integrating the transition intensities. In case we are also interested in the transition probability matrix $p(s,t,u,v)$, we can obtain it as the unique solution of the *Kolmogorov backward differential equation system*

$$\frac{\partial}{\partial s} p(s,t,u,v) = -\mu(s,u)\, p(s,t,0,v) - \frac{\partial}{\partial u} p(s,t,u,v)$$

with boundary condition $p(t,t,u,v) = \mathbf{1}_{\{i=k\}}\, \mathbf{1}_{\{v \geq u+t-s\}}$.

(b) We estimate ${}_{\infty,t-s}p^{jk}_{x+s,u} := p_{jk}(s,t,u,\infty)$ from statistical data, but only at integer times s,t,u. We generally assume that transitions can only occur at the turn of the years. Then from (2.1) we obtain that $q(m,\mathrm{d}t) = 0$ for non-integer times t, and from (2.2) we get

$$q_{jk}(m,n) - q_{jk}(m,n-) := \begin{cases} p_{jk}(n-1,n,n-1-m,\infty) & : j \neq k \\ p_{jj}(n-1,n,n-1-m,\infty) - 1 & : j = k \end{cases}$$

for integer times $m < n$.

In practice, the probabilities $p_{jk}(n-1,n,n-1-m,\infty)$ are often provided by so-called select-and-ultimate tables (cf. Bowers et al. (1997), section 3.8). Such tables generally contain annual rates that depend on the age at selection (e.g. onset of disability) and on the duration since selection. The common actuarial notation is

$$p_{jk}(n-1,n,n-1-m,\infty) =: p^{jk}_{[x+m]+n-1-m}\,.$$

(The value in the square brackets is the age of the policyholder at the last transition, and the addend after the brackets gives the duration of stay in the actual state.) Usually, beyond a certain period such as 5 years, the dependence on the time since selection is neglected and the corresponding transition probabilities are combined, resulting in the so-called ultimate table $p^{jk}_{x+n-1} := p^{jk}_{[x+m]+n-1-m}$ for $n-1-m \geq 5$.

(c) Similarly to (b), we estimate the ${}_{\infty,t-s}p^{jk}_{x+s,u} := p_{jk}(s,t,u,\infty)$ from statistical data at integer times s,t,u. But differing from (b), we distribute the mass $q_{jk}(m,n) - q_{jk}(m,n-1)$ equally on the interval $(n-1,n]$ by defining

$$\mu_{jk}(t,t-m) := \begin{cases} p_{jk}(n-1,n,n-1-m,\infty) & : j \neq k\,, t \in (n-1,n] \\ p_{jj}(n-1,n,n-1-m,\infty) - 1 & : j = k\,, t \in (n-1,n] \end{cases}$$

for all integer times m.

(d) We estimate the ${}_{\infty,t-s}p^{jk}_{x+s,u} := p_{jk}(s,t,u,\infty)$ from statistical data at integer times s,t,u. For times in between we use linear interpolation. See Helwich (2008, p. 93).

If the cumulative transition intensity matrix $q(t,s)$ is regular (the matrix elements $q_{jk}(s,\cdot)$, $j \neq k$, are monotonic non-decreasing and right-continuous functions that are zero at time s, $q_{jj} = -\sum_{k \neq j} q_{jk}$, $q_{jj}(s,t) - q_{jj}(s,t-) \geq -1$, and in case of $q_{jj}(s,t_0) - q_{jj}(s,t_0-) = -1$ the function $q_{jj}(s,\cdot)$ is constant from t_0 on), then there always exists a corresponding semi-Markovian process $((X_t,U_t))_{t \geq 0}$, which is even strong Markovian (see Helwich (2008), Remark 2.34).

In the insurance literature the semi-Markov model based on transition intensities was first described by Hoem (1972). The more general model based on cumulative intensities was introduced by Helwich (2008).

2.2 The Markovian approach

For some insurance products we can simplify the semi-Markovian model to the special case where $(X_t)_{t\geq 0}$ on its own is a Markovian process, which means that for all $i \in \mathcal{S}$ and $t \geq t_n \geq ... \geq t_1 \geq 0$ we have

$$P\big(X_t = i\big|X_{t_n}, ..., X_{t_1}\big) = P\big(X_t = i\big|X_{t_n}\big)$$

almost sure. As a consequence, the transition probabilities $p_{jk}(s,t,u,v)$ and the probabilities $\overline{p}_{jk}(s,t,u)$ do not depend on u or v anymore, the cumulative transition intensity matrix $q(s,t)$ is constant with respect to its first argument s, and the transition intensity matrix $\mu(t,u)$ is constant with respect to the second argument u. Actuaries then use the notation

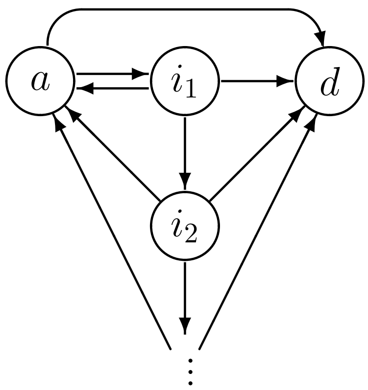
$$\begin{aligned}
{}_{t-s}p^{jk}_{x+s} &:= P(X_{x+t} = k \,|\, X_{x+s} = j) = p_{jk}(s,t,u,\infty)\,,\\
{}_{t-s}p^{\underline{jj}}_{x+s} &:= P(T(s) > t \,|\, X_s = j) = 1 - \overline{p}_{jj}(s,t,u)\,,\\
\mu^{jk}_{x+t} &:= \mu_{jk}(t) = \mu_{jk}(t,u)\,.
\end{aligned}$$

In the insurance literature the Markov model based on transition intensities $\mu_{jk}(t)$ and a differentiable discounting function was already described by Sverdrup (1965) and Hoem (1969). The more general cumulative intensity approach was introduced by Milbrodt and Stracke (1997).

Markovian models are sometimes used to approximate semi-Markovian models. The idea is to split all states i with durational effects into a set of states $i_1, i_2, ..., i_n$ with the meaning

- i_k = 'insured in state i with a duration between $k-1$ and k' for $k = 1, ..., n-1$,
- i_n = 'insured in state i with a duration greater than $n-1$'.

For example, for disability policies we have the well-known effect that for different durations since disability inception the observed recovery and mortality rates differ. Instead of using the three state model $\mathcal{S} = \{a, i, d\}$ of Example 1.1 in a semi-Markovian setup, we could use the multistate model $\mathcal{S} = \{a, i_1, ..., i_n, d\}$ in a Markovian setup.



We have to specify $q(t)$ in such a way that the transitions $(i_1, i_2), ..., (i_{n-1}, i_n)$ occur with probability one at the turn of the years. Durational effects are taken into account by using different recovery and mortality rates for the states $i_1, ..., i_n$. This differentiation of disability states has been originally proposed by Amsler (1968).

2.3 Other approaches

More realistic and sophisticated models can be built by considering for example

- the dependence on the age x at policy issue,
- the dependence on the total time spent in some states since policy issue, stressing the health history of the insured.

Both approaches complicate the actuarial models. The first one can be included by estimating ${}_{v,t-s}p^{jk}_{x+s,u}$, ${}_{t-s}p^{\overline{jj}}_{x+s,u}$, or $\mu^{jk}_{x+t,u}$ separately for each starting age x. Another interesting approach is given by Davis (1984). However, reasonable models can be defined by considering just the semi-Markovian case of above.

3 Payment functions

As already mentioned in the introduction, premiums and benefits are paid according to the pattern of states of the policyholder.

Definition 3.1 (contractual payments)**.** Payments between insurer and policyholder are of two types:

(a) Lump sums are payable upon transitions between two states and are specified by deterministic nonnegative functions $b_{jk}(t,u)$ with bounded variation on compacts. The amount $b_{jk}(t,u)$ is payable if the policy jumps from state j to state k at time t and the duration of stay in state j was u.

In order to distinguish between payments from insurer to insured and vice versa, benefit payments get a positive sign and premium payments get a negative sign.

(b) Annuity payments fall due during sojourns in a state and are defined by deterministic functions $B_j(s,t)$, $j \in \mathcal{S}$. Given that the last transition occurred at time s, $B_j(s,t)$ is the total amount paid in $[s,t]$ during a sojourn in state j. We assume that the $B_j(s,\cdot)$ are right-continuous and of bounded variation on compacts.

Example 3.2 (disability insurance)**.** Suppose we have a contract that pays a continuous disability annuity with rate $r > 0$ as long as the insured is in state disabled, and a constant (lump sum) premium $p > 0$ has to be paid yearly in advance as long as the policyholder is in state active. Following a modeling framework of Pitacco (1995),

- let (n_1, n_2) denote the insured period, meaning that benefits are payable if the disability inception time belongs to this interval,
- let f denote the deferred period (from disability inception),
- let m be the maximum number of years of annuity payment (from disability inception),
- and let T $(\geq n_2 + f)$ be the stopping time of annuity payment (from policy issue at time zero).

Then the contractual payment functions are

$$B_a(s,t) = \sum_{k=0}^{n_2-1} -p\,\mathbf{1}_{[k,\infty)}(t)\,,$$
$$B_i(s,t) = \mathbf{1}_{(n_1,n_2)}(s) \int_{\min\{t,s+f\}}^{\min\{t,s+f+m,T\}} r\,\mathrm{d}u\,,$$
$$B_d(s,t) = 0$$

for all $0 \leq s \leq t$. The transitions payments $b_{jk}(t,u)$ are all constantly zero.

Example 3.3 (critical illness insurance)**.** Let the state space be $\mathcal{S} := \{a = \text{active/healthy}, i = \text{ill}, d = \text{dead}\}$, and denote by T the policy term. Suppose that a constant (lump sum) premium $p > 0$ has to be paid yearly in advance as long as the policyholder is in state active. A lump sum is payed upon death or a dread disease diagnosis, whichever occurs first. The death benefit is l_d. The dread disease benefit is l_i, but only if the policyholder outlives a deferred period f, otherwise the sum l_d is paid. We assume that a recovery from state i to state a is impossible. The contractual payment functions are

$$B_a(s,t) = \sum_{k=0}^{T-1} -p\,\mathbf{1}_{[k,\infty)}(t)\,, \quad B_i(s,t) = l_i\,\mathbf{1}_{(0,T-f]}(s)\,\mathbf{1}_{[s+f,\infty)}(t)\,, \quad B_d(s,t) = 0\,,$$
$$b_{ad}(t,u) = l_d\,\mathbf{1}_{(0,T]}(t)\,, \quad b_{ai}(t,u) = 0\,, \quad b_{id}(t,u) = l_d\,\mathbf{1}_{(0,T]}(t)\,\mathbf{1}_{[0,f)}(u)\,,$$

for all $0 \leq s \leq t$ and $u \geq 0$. Without the deferred period ($f = 0$), we can alternatively write

$$B_a(s,t) = \sum_{k=0}^{T-1} -p\,\mathbf{1}_{[k,\infty)}(t)\,, \quad B_i(s,t) = 0\,, \quad B_d(s,t) = 0\,,$$
$$b_{ad}(t,u) = l_d\,\mathbf{1}_{(0,T]}(t)\,, \quad b_{ai}(t,u) = l_i\,\mathbf{1}_{(0,T]}(t)\,, \quad b_{id}(t,u) = 0\,.$$

Example 3.4 (long-term care insurance)**.** Suppose we have a German policy with state space $\mathcal{S} = \{a = \text{active/healthy}, c^I = \text{need for basic care}, c^{II} = \text{need for medium care}, c^{III} = \text{need for comprehensive care}, l = \text{lapsed/canceled}, d = \text{dead}\}$. Continuous annuities are paid with rates r^I, r^{II}, r^{III} as long as the insured is in state c^I, c^{II}, c^{III}, respectively. A constant (lump sum) premium $p > 0$ has to be paid yearly in advance till retirement at contract time R, lapse, or death, whichever occurs first. The contractual payment functions are

$$B_a(s,t) = \sum_{k=0}^{R-1} -p\,\mathbf{1}_{[k,\infty)}(t)\,,$$
$$B_{c^{I/II/III}}(s,t) = (t - \min\{t,s\})\,r^{I/II/III}\,,$$
$$B_l(s,t) = 0\,, \quad B_d(s,t) = 0$$

for all $0 \leq s \leq t$. The transitions payments $b_{jk}(t,u)$ are all constantly zero. Interestingly, the benefits $B_i(s,\mathrm{d}t)$ at time $t \geq s$ do not really depend on s. Hence, we can simplify the payment functions by skipping the first argument,

$$
\begin{aligned}
B_a(t) &= \sum_{k=0}^{R-1} -p\,\mathbf{1}_{[k,\infty)}(t)\,,\\
B_{c^{I/II/III}}(t) &= t\,r^{I/II/III}\,,\\
B_l(t) = 0\,,& \quad B_d(t) = 0\,.
\end{aligned}
$$

Example 3.5 (German private health insurance)**.** As already mentioned above, German health insurers calculate on the basis of deterministic benefit forecasts that only depend on the age of the policyholder. This deterministic approach is justified by the argument that, though the individual medical expenses are far from being deterministic, the portfolio average of medical expenses is nearly deterministic in case of a large portfolio of independent insured because of the law of large numbers. The law of large numbers does not work for systematic changes of medical expenses with progression of calendar time, but German health insurers have the exceptional right to adapt their deterministic forecasts of the medical expenses if calculated values and real values differ significantly. (The special right of German and Austrian health insurers to change the payments functions after(!) signing of the contract is unique in private health insurance.) Let the average yearly medical expenses (Kopfschäden) of a policyholder with age $x+k$ be a deterministic quantity K_{x+k}, where x is the age of the policyholder at inception of the contract. Recall that the state space is $\mathcal{S} := \{a = \text{alive}, l = \text{lapsed/canceled}, d = \text{dead}\}$. A constant (lump sum) premium $p > 0$ has to be paid yearly in advance till death or lapsation. The contractual payment functions are

$$
\begin{aligned}
B_a(s,t) &= \sum_{k=0}^{\infty} \big(K_{x+k} - p\big)\mathbf{1}_{[k,\infty)}(t)\,,\\
B_l(s,t) &= 0\,, \quad B_d(s,t) = 0
\end{aligned}
$$

for all $0 \le s \le t$. The transitions payments $b_{jk}(t,u)$ are all constantly zero. All payments are independent of the duration of stay. Hence, we can simplify the payment functions by skipping the first argument,

$$
\begin{aligned}
B_a(t) &= \sum_{k=0}^{\infty} \big(K_{x+k} - p\big)\mathbf{1}_{[k,\infty)}(t)\,,\\
B_l(t) &= 0\,, \quad B_d(t) = 0\,.
\end{aligned}
$$

4 Reserves

By statute the insurer must at any time maintain a reserve in order to meet future liabilities in respect of the contract. This reserve bears interest with some rate $\varphi(t)$. On the basis of this interest rate we define a discounting function,

$$
v(s,t) := e^{-\int_s^t \varphi(r)\mathrm{d}r}\,. \tag{4.1}
$$

We can interpret $v(s,t)$ as the value at time s of a unit payable at time $t \geq s$. At next, we study the present value of future payments between insurer and policyholder, that is, the discounted sum of all future benefit and premium payments,

$$\begin{aligned} A(t) := \sum_{j \in \mathcal{S}} \sum_{n=0}^{\infty} \int_{(t,\infty)} & v(t,\tau)\, \mathbf{1}_{\{X_\tau = j\}}\, \mathbf{1}_{\{S_n \leq \tau < S_{n+1}\}}\, B_j(S_n, \mathrm{d}\tau) \\ & + \sum_{(j,k) \in J} \int_{(t,\infty)} v(t,\tau)\, b_{jk}(\tau, U_\tau)\, \mathrm{d}N_{jk}(\tau)\,. \end{aligned} \tag{4.2}$$

$A(t)$ gives the amount that an insurer needs at time t in order to meet all future obligations in respect of the contract, but, because of its randomness, we do not know its value. However, if we have a homogeneous portfolio of stochastically independent policies, then the law of large numbers yields that the average present value per policy is close to its mean. This motivates the following definition.

Definition 4.1. The *prospective reserve at time t in state (i,u)* is defined by

$$V_i(t,u) := \mathbb{E}\big(A(t) \,\big|\, (X_t, U_t) = (i,u)\big)\,,$$

given that the expectation exists.

$V_i(t,u)$ is the amount that the insurer needs on portfolio average in order to meet all future obligations. At present time t we do not only know the value of (X_t, U_t) but also the complete history of $(X_\tau, U_\tau)_{\tau \geq 0}$ up to time t. But since we assumed that $(X_\tau, U_\tau)_{\tau \geq 0}$ is Markovian, in the above definition we may just condition on (X_t, U_t).

Theorem 4.2. *The prospective reserves $V_i(t,u)$, $i \in \mathcal{S}$, $0 \leq u \leq t$, equal almost sure the unique solution of the integral equation system*

$$\begin{aligned} V_i(t,u) = & \int_{(t,\infty)} v(t,\tau)\,(1 - \overline{p}_{ii}(t,\tau,u))\, B_i(t-u, \mathrm{d}\tau) \\ & + \sum_{j: j \neq i} \int_{(t,\infty)} v(t,\tau)\,(1 - \overline{p}_{ii}(t,\tau - 0,u))\Big(b_{ij}(\tau, \tau - t + u) + V_j(\tau,0)\Big) q_{ij}(t-u, \mathrm{d}\tau)\,. \end{aligned} \tag{4.3}$$

A proof can be found in Helwich (2008). The integral equation system (4.3) is also denoted as Thiele integral equations (of type 1). The theorem provides a way for the calculation of $V_i(t,u)$. Generally, we can use the following method.

Algorithm 4.3. (1) At first solve the integral equation system (4.3) for $u = 0$,

$$\begin{aligned} V_i(t,0) = & \int_{(t,\infty)} v(t,\tau)\,(1 - \overline{p}_{ii}(t,\tau,0))\, B_i(t, \mathrm{d}\tau) \\ & + \sum_{j: j \neq i} \int_{(t,\infty)} v(t,\tau)\,(1 - \overline{p}_{ii}(t,\tau - 0,0))\Big(b_{ij}(\tau, \tau - t) + V_j(\tau,0)\Big) q_{ij}(t, \mathrm{d}\tau)\,. \end{aligned}$$

If there is no closed form solution, we have to use numerical methods in order to obtain at least approximations on a fine time grid. In practice we will always have some time $T < \infty$ such that $V_i(t,0) = 0$ for all $t \geq T$ and $i \in \mathcal{S}$. We start from that T and make small backward steps till we arrive at $t = 0$.

(2) On the basis of the solution from (1), we can now calculate the integrals in (4.3), at least approximately by using numerical integration. Recall that the probability $\overline{p}_{ii}(t,\tau,0)$ has the explicit representation (2.3).

We now demonstrate this algorithm for some examples. We will see that in many practical examples the calculation method can be considerably simplified.

Example 4.4 (disability insurance)**.** We continue with Example 3.2. Suppose that the transition intensity matrix $\mu(t,u)$ exists. As there are no payments in state d = dead, and since the transitions (d,a) and (d,i) are not possible, we obtain $V_d(t,u) = 0$ for all $0 \leq u \leq t$. Furthermore, we have $V_a(t,u) = 0$ and $V_i(t,u) = 0$ for all $t \geq T$, because there are no payments after stopping time T. Thus, we get

$$\begin{aligned} V_a(t,u) &= \int_{(t,T]} v(t,\tau)\,(1-\overline{p}_{aa}(t,\tau,u))\,B_a(t-u,\mathrm{d}\tau) \\ &\quad + \int_t^T v(t,\tau)\,(1-\overline{p}_{aa}(t,\tau,u))\,V_i(\tau,0)\,\mu_{ai}(\tau,\tau-t+u)\,\mathrm{d}\tau\,, \\ V_i(t,u) &= \int_{(t,T]} v(t,\tau)\,(1-\overline{p}_{ii}(t,\tau,u))\,B_i(t-u,\mathrm{d}\tau) \\ &\quad + \int_t^T v(t,\tau)\,(1-\overline{p}_{ii}(t,\tau,u))\,V_a(\tau,0)\,\mu_{ia}(\tau,\tau-t+u)\,\mathrm{d}\tau\,. \end{aligned} \tag{4.4}$$

Inserting the definitions of B_a and B_i, for $u = 0$ we get

$$\begin{aligned} V_a(t,0) &= \sum_{k=0,k>t}^{n_2-1} v(t,k)\,(1-\overline{p}_{aa}(t,k,0))\,(-p) \\ &\quad + \int_t^T v(t,\tau)\,(1-\overline{p}_{aa}(t,\tau,0))\,V_i(\tau,0)\,\mu_{ai}(\tau,\tau-t)\,\mathrm{d}\tau\,, \\ V_i(t,0) &= \int_{(t,T]\cap(n_1,n_2)\cap(t+f,t+f+m)} v(t,\tau)\,(1-\overline{p}_{ii}(t,\tau,0))\,r\,\mathrm{d}\tau \\ &\quad + \int_t^T v(t,\tau)\,(1-\overline{p}_{ii}(t,\tau,0))\,V_a(\tau,0)\,\mu_{ia}(\tau,\tau-t)\,\mathrm{d}\tau\,. \end{aligned}$$

Since the mapping $t \mapsto (V_a(t,0), V_i(t,0))$ is continuous at non-integer times t, we can approximate this mapping in between integer times by a piecewise constant function on a fine time grid. At integer times $t = n$ we have $V_a(n-,0) = V_a(n+,0) - v(t,n)\,(1-\overline{p}_{aa}(t,n,0))\,p$ and $V_i(n-,0) = V_i(n+,0)$. We start from $(V_a(T,0), V_i(T,0)) = (0,0)$ and use a backward approximation scheme with sufficiently small steps. When reaching $t = 0$, we have an approximation of $t \mapsto (V_a(t,0), V_i(t,0))$ for all $t \geq 0$, and in a second step we can now approximate the $V_a(t,u)$ and $V_i(t,u)$ by using numerical integration in formula (4.4). Sometimes actuaries further simplify the model by assuming that there is no reactivation (i,a) possible. In this case, $V_i(t,u)$ can be directly calculated from

$$V_i(t,u) = \int_{(t,T]} v(t,\tau)\,(1-\overline{p}_{ii}(t,\tau,u))\,B_i(t-u,\mathrm{d}\tau)\,,$$

and then $V_a(t,u)$ can be directly calculated from (4.4), if necessary by using numerical integration.

Example 4.5 (critical illness insurance)**.** We continue with Example 3.3. Again suppose that the transition intensity matrix $\mu(t,u)$ exists. Generally, we can use Algorithm 4.3 similarly to Example 4.4. In case the deferred period f is zero, we have $V_i(t,u)=0$ and $V_d(t,u)=0$ for all $0\le u\le t$, and it suffices to solely look at $V_a(t,u)$,

$$\begin{aligned} V_a(t,u) = &\sum_{k=0,k>t}^{T-1} v(t,k)\,(1-\overline{p}_{aa}(t,k,u))\,(-p) \\ &+ \int_t^T v(t,\tau)\,(1-\overline{p}_{aa}(t,\tau,u))\,b_{ai}(\tau,\tau-t+u)\,\mu_{ai}(\tau,\tau-t+u)\,\mathrm{d}\tau \\ &+ \int_t^T v(t,\tau)\,(1-\overline{p}_{aa}(t,\tau,u))\,b_{ad}(\tau,\tau-t+u)\,\mu_{ad}(\tau,\tau-t+u)\,\mathrm{d}\tau\,. \end{aligned} \tag{4.5}$$

Since the right hand side does not contain any prospective reserves anymore, we just have to (numerically) integrate the right hand side in order to obtain $V_a(t,u)$.

Example 4.6 (long-term care insurance)**.** We continue with Example 3.4. Our model with state space $\mathcal{S}=\{a=\text{active/healthy}, c^I=\text{need for basic care}, c^{II}=\text{need for medium care}, c^{III}=$ need for comprehensive care, $l=\text{lapsed/canceled}, d=\text{dead}\}$ is a so called *hierarchical* model since the transition probability matrix $p(s,t)$ has a triangular form. As a consequence, for each state i in the ordered set $[a,c^I,c^{II},c^{III},l,d]$ formula (4.3) depends only on prospective reserves of states that follow state i in $[a,c^I,c^{II},c^{III},l,d]$. We start with state d for which we get $V_d(t,u)=0$ for all $0\le u\le t$. Then we continue with state l for which we also have $V_d(t,u)=0$ for all $0\le u\le t$. At next we calculate $V_{c^{III}}(t,u)$,

$$V_{c^{III}}(t,u) = \int_{(t,\infty)} v(t,\tau)\,(1-\overline{p}_{ii}(t,\tau,u))\,r^{III}\,\mathrm{d}\tau\,.$$

Then we go on with c^{II} and calculate $V_{c^{II}}(t,u)$ on the basis of $V_d(t,0)$, $V_l(t,0)$, and $V_{c^{III}}(t,0)$. The last steps are the calculations of $V_{c^I}(t,u)$ and $V_a(t,u)$. While for the general method according to Algorithm 4.3 we have to solve an integral equation system, we can here simply use consecutive (numerical) integration in order to obtain the prospective reserves.

Example 4.7 (German private health insurance)**.** We continue with Example 3.5. As there are no payments in state l and state d, and since we cannot jump back to state a, we have $V_l(t,u)=0$ and $V_d(t,u)=0$ for all $0\le u\le t$, and it suffices to solely look at $V_a(t,u)$,

$$V_a(t,u) = \sum_{k=0,k>t}^{\infty} v(t,k)\,(1-\overline{p}_{aa}(t,k,u))\,(K_{x+k}-p)\,.$$

According to the definition of $B_a(s,t)$ in Example 3.5, all medical expenses are paid as lump sums yearly in advance. This simplification is motivated by the discrete time approach (b) of section 2.1, where state changes may only occur at the turn of the years. German private health insurers usually use such a discrete time approach and additionally assume that the random pattern of states is a Markovian process according to section 2.2. This Markov assumption is controversial because statistical data shows that withdrawal probabilities decrease with the time elapsed since inception of the contract. This effect can be explained by the fact that policyholders partly lose their ageing provision (the prospective reserve) if they switch insurance companies. The aging provision rises with increasing contract time and with it the loss due to lapse. Nevertheless, German private health insurers use the Markovian approach, and for integer times t and k with $t < k$ we get

$$\begin{aligned} 1 - \overline{p}_{aa}(t,k,u) &= P(T(t) > k \,|\, X_t = a, U_t = u) \\ &= P(T(t) > k \,|\, X_t = a) \\ &= \prod_{j=t}^{k-1} P(T(j) > j+1 \,|\, X_j = a) \\ &= \prod_{j=t}^{k-1} {}_1p_{x+j}^{aa} \,. \end{aligned}$$

The probabilities ${}_1p_{x+j}^{aa} = 1 - {}_1p_{x+j}^{al} - {}_1p_{x+j}^{ad} =: 1 - \omega_{x+j} - q_{x+j}$ are estimated from statistical data. Referring to approach (b) in section 2.1, we define the cumulative transition intensity matrix $q(s,t)$ by $q_{al}(m,n) - q_{al}(m,n-) := \omega_{x+n-1}$, $q_{ad}(m,n) - q_{ad}(m,n-) := q_{x+n-1}$, and $q(m,n-) - q(m,n-1) = 0$ for all integer times $n \geq 1$.

Remark 4.8 (Discrete time approach)**.** Following approach (b) in section 2.1, let transitions between different states only happen at integer times. Then $q(s,\mathrm{d}\tau)$ is zero at non-integer times. Suppose that also the $B_i(s,\mathrm{d}\tau)$ are zero at non-integer times, which means that payments during sojourns in a state are paid either yearly in advance or yearly in arrears. As transitions shall only occur at the turn of the years, that does not mean a loss of generality. Let $T < \infty$ be the contract term. Then we can rewrite (4.3) to

$$\begin{aligned} V_i(n,m) = &\sum_{\tau=n+1}^{T} v(n,\tau)\,(1 - \overline{p}_{ii}(n,\tau,m))\big(B_i(\tau,n-m) - B_i(\tau-,n-m)\big) \\ &+ \sum_{j:j\neq i}\sum_{\tau=n+1}^{T} v(n,\tau)\,(1-\overline{p}_{ii}(n,\tau-0,m))\Big(b_{ij}(\tau,\tau-n+m) + V_j(\tau,0)\Big) \\ &\qquad\qquad \times \big(q_{ij}(n-m,\tau) - q_{ij}(n-m,\tau-)\big) \end{aligned}$$

for all integer times $0 \leq m \leq n$. We obtained a recursion formula that can be easily solved backwards starting from some $n = T$ and making time steps of -1 till we arrive at $t = 0$. Recall that the probabilities $\overline{p}_{ii}(n,\tau,m)$ can be easily calculated with the help of (2.3).

Remark 4.9 (Absolute continuity approach)**.** If not only the cumulative transition intensities are differentiable with respect to their second argument (the transition intensity matrix

μ exists), but also the $B_j(s,t)$ are differentiable in the second argument with derivatives $b_j(t, t-s)$, then we can alternatively obtain the prospective reserves $V_i(t,u)$ by solving the system of partial differential equations

$$\begin{aligned} \frac{\partial}{\partial t} V_i(t,u) = & -\frac{\partial}{\partial u} V_i(t,u) - b_i(t,u) + \varphi(t)\, V_i(t,u) \\ & - \sum_{j:j\neq i} \Big(b_{ij}(t,u) + V_j(t,0) - V_i(t,u) \Big) \mu_{ij}(t,u) \,. \end{aligned} \tag{4.6}$$

See Helwich (2008, Theorem 4.11). If there are no durational effects, then $\frac{\partial}{\partial u} V_i(t,u)$ is always zero and (4.6) reduces to a system of ordinary differential equations.

5 Premiums

Definition 5.1 (equivalence premium)**.** The equivalence principle states that premiums should be chosen in such a way that at time zero (the beginning of the contract) the expected present value of future premiums should equal the expected present value of future benefits. In mathematical terms that is just the case if and only if

$$B_{X_0}(0,0) + V_{X_0}(0,0) = 0 \,.$$

(Recall that we assumed X_0 to be deterministic, and that $V_{X_0}(0,0)$ just contains the payments strictly after time zero and, thus, we have to add $B_{X_0}(0,0)$.)

The typical way for the calculation of the equivalence premium is as follows. First we choose the benefits. Second, we choose a premium scheme such as a lump sum premium at beginning of the contract, a yearly constant premium, or a regular premium that increases each year by some specified factor. The last step is then to find a scaling factor for the premiums such that the equivalence principle is met.

The equivalence premium is also called net premium. In practice the real market premiums (gross premiums) additionally comprise a risk load, acquisition costs, administrative costs, profit margins, and taxes. The risk load is needed since the real sum $B_{X_0}(0,0) + A(0)$ that an insurer needs at the beginning of the contract (in order to meet all obligations in respect of the contract) can deviate from the expected one,

$$B_{X_0}(0,0) + A(0) \overset{?}{\approx} \mathbb{E}\big(B_{X_0}(0,0) + A(0)\big) = B_{X_0}(0,0) + V_{X_0}(0,0) \,.$$

We suppose for the moment that the valuation basis, that is the probability distribution of $(X_\tau)_{\tau\geq 0}$ and the discounting factor, is known at the time of inception of the contract. Then, because of the law of large numbers, $B_{X_0}(0,0) + V_{X_0}(0,0)$ is a good approximation for $B_{X_0}(0,0) + A(0)$ in case the insurer has a large portfolio of independent insured. We only have a so-called *unsystematic risk* that is diversifiable by increasing the size of the portfolio. In reality, however, the valuation basis may undergo significant and unforeseeable changes within the time horizon of the contract, thus exposing the insurer to a *systematic risk* that is non-diversifiable. There are two ways in which actuaries deal with the systematic risk:

(a) Adding *explicit risk loads* on the premium. The equivalence premium is increased by an amount that reflects the uncertainty of $V_{X_0}(0,0)$. For example, the risk load is chosen proportional to the variance or the standard deviation, or is derived from quantiles.

(b) Adding *implicit risk loads* on the premium. In practice this is done by calculating the equivalence premium on a conservative so-called *first order basis*, which represents a provisional worst-case scenario for the future development.

For that part of the systematic risk that is traded on a financial market, e.g. interest rate risk, we can also derive a risk load from market prices observed in reality. In practice, we often have a mixture of explicit and implicit risk loads. They must be chosen very carefully, since the insurer has no right to increase premiums later on if it turns out that the first order basis was too optimistic. An exception are German private health insurances, where the regulatory regime allows to adapt premiums under specified circumstances. These circumstances are given if the average yearly medical expenses or the mortality rates change significantly. However, a German health insurer still needs carefully chosen risk loads since the regulatory regime does not allow premium adaptions at the discretion of the insurer but imposes sharp restrictions.

6 Calculation on the safe side

For all kinds of actuarial calculations, the actuary first has to specify the probability distribution of the random pattern of states. The insurance literature offers various statistical methods for the estimation of (past) transition rates from data. For the interested reader we suggest to consult Haberman and Pitacco (1999, section 4) and Milbrodt and Helbig (1999, section 3.F). However, history shows that the probability distribution of the random pattern of states can change significantly within the contract period. For example, the world life expectancy more than doubled over the past two centuries (see Oeppen and Vaupel (2002)), and disability and recovery rates are affected by the rapid progress in medical treatment and fast changing requirements in the world of work. Since many health insurance policies have rather long contract periods, these changes can have a considerable effect on actuarial calculations, exposing the insurer to a systematic biometric risk. In this section we discuss several methods that address this risk.

6.1 Valuation basis of first order

Health insurance pursued on a similar technical basis to that of life insurance is calculated either with first order valuation bases or with second-order valuation bases. First order bases include a safety margin whereas second-order ones do not contain any margin and are assumed to be close to reality. Appropriate first order bases are essential to the health insurance business if implicit risk loads according to item (b) in section 5 are used. The first order basis justifies the use of expected present values (see Definitions 4.1 and 5.1) without explicit safety loading. In general it is not obvious how to include a safety margin in the transition probabilities. Should we rather overestimate or underestimate mortality rates, morbidity rates, reactivation rates, and so on?

According to Helwich (2008, chapter 4.D), the prospective reserves $V_i(t,u)$ can also be seen as the unique solution of the following Thiele integral equation system (of type 2)

$$\begin{aligned} V_i(t,u) &= \int_{(t,\infty)} B_i(t-u,\mathrm{d}\tau) - \int_{(t,\infty)} V_i(t-,u+\tau-t)\,\varphi(\tau)\,\mathrm{d}\tau \\ &\quad + \sum_{j:j\neq i}\int_{(t,\infty)} R_{ij}(\tau,u+\tau-t)\,q_{ij}(t-u,\mathrm{d}\tau)\,, \end{aligned} \tag{6.1}$$

where $R_{ij}(\tau,v)$ is the so-called sum at risk associated with a possible transition from state i to state j at time τ,

$$R_{ij}(\tau,v) := b_{ij}(\tau,v) + V_j(t,0) + B_j(\tau,\tau) - V_i(\tau,v) - (B_i(\tau-v,\tau) - B_i(\tau-v,\tau-))\,.$$

(We use the convention $B_i(\tau,\tau-) := 0$.) Now suppose that we have another set of actuarial assumptions $\varphi^*(t)$ and $q^*(s,t)$ with corresponding prospective reserves $V_i^*(t,u)$ and sums at risk $R^*_{ij}(t,u)$. Let $W_i(t,u) := V_i^*(t,u) - V_i(t,u)$. From (6.1) we get

$$\begin{aligned} W_i(t,u) = &- \int_{(t,\infty)} W_i(\tau-,u+\tau-t)\,\varphi(\tau)\,\mathrm{d}\tau \\ &- \int_{(t,\infty)} V_i^*(\tau-,u+\tau-t)\,(\varphi^*(\tau)-\varphi(\tau))\,\mathrm{d}\tau \\ &+ \sum_{j:j\neq i}\int_{(t,\infty)} \big(W_j(\tau,0) - W_i(\tau,u+\tau-t)\big)\,q_{ij}(t-u,\mathrm{d}\tau) \\ &+ \sum_{j:j\neq i}\int_{(t,\infty)} R^*_{ij}(\tau,u+\tau-t)\,(q^*_{ij}-q_{ij})(t-u,\mathrm{d}\tau)\,. \end{aligned}$$

For the sum of the second and fourth integral we write

$$\begin{aligned} &- \int_{(t,r]} V_i^*(\tau-,u+\tau-t)\,(\varphi^*(\tau)-\varphi(\tau))\,\mathrm{d}\tau \\ &+ \sum_{j:j\neq i}\int_{(t,r]} R_{ij}(\tau,u+\tau-t)\,(q^*_{ij}-q_{ij})(t-u,\mathrm{d}\tau) \\ &=: \int_{(t,r]} C_i(t-u,\mathrm{d}\tau) = C_i(t-u,r) - C_i(t-u,t)\,. \end{aligned}$$

Suppose that the cumulative transition intensities q^*_{ij} and q_{ij} have representations of the form

$$\begin{aligned} q^*_{ij}(s,t) &= \int_{(s,t]} \lambda^*_{ij}(\tau,\tau-s)\,\mathrm{d}\Lambda_{ij}(\tau)\,, \\ q_{ij}(s,t) &= \int_{(s,t]} \lambda_{ij}(\tau,\tau-s)\,\mathrm{d}\Lambda_{ij}(\tau)\,. \end{aligned}$$

For example, if the transition intensity matrices μ^* and μ exist, then let the Λ_{ij} be Lebesgue-Borel measures and define $\lambda^*_{ij} := \mu^*_{ij}$ and $\lambda_{ij} := \mu_{ij}$. In the discrete time model according to approach (b) in section 2.1 and Remark 4.8, we define the Λ_{ij} as counting measures and $\lambda_{ij}^*(n,m) := p_{jk}^*(n-1,n,n-1-m,\infty)$ and $\lambda_{ij}(n,m) := p_{jk}(n-1,n,n-1-m,\infty)$. By interpreting the C_i as synthetic annuity payments, we can see the integral equation system for the $W_i(t,u)$ as a Thiele integral equation system of the form (6.1) of a policy with cumulative annuity payments C_i and zero transition benefits under the valuation basis (φ^*, q^*). It has the unique solution (see Helwich (2008), chapter 4)

$$\begin{aligned} W_i(t,u) = &-\sum_{j\in\mathcal{S}} \int_{(t,\infty)} \int_{[0,\infty)} v(t,\tau)\, V_j^*(\tau-,l)\, (\varphi^*(\tau) - \varphi(\tau))\, p_{ij}(t,\tau,u,\mathrm{d}l)\, \mathrm{d}\tau \\ &+ \sum_{(j,k)\in J} \int_{(t,\infty)} \int_{[0,\infty)} v(t,\tau)\, R_{jk}^*(\tau,l)\, (\lambda_{jk}^*(\tau,l) - \lambda_{jk}(\tau,l))\, p_{ij}(t,\tau,u,\mathrm{d}l)\, \mathrm{d}\Lambda_{jk}(\tau)\,. \end{aligned}$$

From this equation we can derive a sufficient condition for a valuation basis of first order to be on the safe side.

Lemma 6.1. *If*

$$V_j^*(\tau-,l) \gtreqless 0 \Longleftrightarrow \varphi(\tau) \gtreqless \varphi^*(\tau) \quad \textit{and} \quad R_{jk}^*(\tau,l) \gtreqless 0 \Longleftrightarrow \lambda_{jk}^*(\tau,l) \gtreqless \lambda_{jk}(\tau,l) \tag{6.2}$$

for all $j,k \in \mathcal{S}$, $j \neq k$, $0 \leq l \leq \tau$, *then we have* $V_i^*(t,u) \geq V_i(t,u)$ *for all* $i \in \mathcal{S}$ *and* $0 \leq u \leq t$. *If the transition intensity matrices* μ^* *and* μ *exist, the second condition in* (6.2) *is equivalent to*

$$R_{jk}^*(\tau,l) \gtreqless 0 \Longleftrightarrow \mu_{jk}^*(\tau,l) \gtreqless \mu_{jk}(\tau,l)\,,$$

and in the discrete time model according to approach (b) *in section* 2.1 *and Remark* 4.8, *the second condition in* (6.2) *has the form*

$$R_{jk}^*(n,m) \gtreqless 0 \Longleftrightarrow p_{jk}^*(n-1,n,n-1-m,\infty) \gtreqless p_{jk}(n-1,n,n-1-m,\infty)$$

for integer times $0 \leq m \leq n$.

Example 6.2 (critical illness insurance)**.** In case the deferred period f is zero, we have $V_i(t,u) = 0$ and $V_d(t,u) = 0$ for all $0 \leq u \leq t$. The only relevant transitions are (a,i) and (a,d). The corresponding sums at risk are

$$\begin{aligned} R_{ai}(\tau,v) &= b_{ai}(\tau,v) - V_a(\tau,v) - (B_a(\tau-v,\tau) - B_a(\tau-v,\tau-))\,, \\ R_{ad}(\tau,v) &= b_{ad}(\tau,v) - V_a(\tau,v) - (B_a(\tau-v,\tau) - B_a(\tau-v,\tau-))\,. \end{aligned}$$

If the illness benefit l_i and the death benefit l_d are equal, one can show that both $R_{ai}(t,u)$ and $R_{ad}(t,u)$ are never negative, and the second condition in (6.2) can be written in the form

$$\mu_{ai}^*(\tau,l) \geq \mu_{ai}(\tau,l)\,, \quad \mu_{ad}^*(\tau,l) \geq \mu_{ad}(\tau,l)\,, \quad 0 \leq l \leq \tau\,.$$

If $l_i \neq l_d$, the sign of $R_{ai}^*(\tau,l)$ or $R_{ad}^*(\tau,l)$ can get negative for some τ and l. In this case it is still unclear how to choose the first order basis μ^*. Although we can calculate $R_{ai}^*(\tau,l)$ and $R_{ad}^*(\tau,l)$ for fixed μ^*, a change of μ^* – in order to meet (6.2) – at the same time leads to a change of $R_{ai}^*(\tau,l)$ and $R_{ad}^*(\tau,l)$ and, thus, we have the problem that we simultaneously change both sides of condition (6.2).

Example 6.3 (German private health insurance)**.** The only relevant transitions are (a, l) and (a, d). The corresponding sums at risk are

$$R_{al}(\tau, v) = R_{ad}(\tau, v) = -V_a(\tau, v) - (B_a(\tau - v, \tau) - B_a(\tau - v, \tau -)) \,.$$

Normally, the mean individual medical expenses K_{x+m} are non-decreasing with respect to age $x + m$, which implies that $V_a(\tau, v)$ is never negative. Hence, $R_{al}(\tau, v) = R_{ad}(\tau, v) \leq 0$, and the second condition in (6.2) has here the form

$$q^*_{x+n} \geq q_{x+n}\,, \quad \omega^*_{x+n} \geq \omega_{x+n}\,, \quad n \geq 0\,.$$

As already mentioned in Example 6.2, if the sign of the $R^*_{jk}(\tau, l)$ in (6.2) can change, then it is still unclear how to find a safe side valuation basis. The problem is that a change of λ^* in order to meet condition (6.2) at the same time changes the condition itself since the $R^*_{jk}(\tau, l)$ are varying as well. This construction problem is formulated in form of an optimization problem in section 6.3.

6.2 Stochastic biometric valuation basis

Though actuaries usually model the sojourn times in the different states stochastically, they frequently rely on deterministic prognoses of the transition probabilities. However, the future transition probabilities are unknown and, thus, random. Among actuaries there is an increasing awareness for the corresponding systematic biometric risk, and there is a recent trend to model transition rates stochastically. The advantage of a stochastic modeling of transition rates is not only in substituting point estimates by confidence estimates, but also in the fact that a stochastic model much better describes the diversity of possible future scenarios than a single scenario like the first order basis.

In order to allow for stochastic transition rates, we need to extent our modeling framework. Let Q be a set of regular cumulative transition intensity matrices q, and let $\widetilde{q} \in Q$ be a stochastic cumulative interest intensity matrix on $(\Omega, \mathfrak{F}, P)$. (Regular means that the matrix elements $q_{jk}(s, \cdot)$, $j \neq k$, are non-decreasing and right-continuous functions that are zero at time s, $q_{jj} = -\sum_{k \neq j} q_{jk}$, $q_{jj}(s, t) - q_{jj}(s, t-) \geq -1$, and in case of $q_{jj}(s, t_0) - q_{jj}(s, t_0-) = -1$ the function $q_{jj}(s, \cdot)$ is constant from t_0 on.) As already mentioned above, for each $q \in Q$ there exists a corresponding semi-Markovian process $(X^q_t, U^q_t)_{t \geq 0}$ with q as its cumulative transition intensity matrix (see Helwich (2008), Remark 2.34). We write $A^q(t)$ for the corresponding present value (cf. formula (4.2)). If $(q, A) \mapsto P((X^q_t, U^q_t)_{t \geq 0} \in A)$ is a Markov kernel (Christiansen (2007) showed that in the Markovian case according to section 2.2 we basically always have this kernel property), then there exists a random pattern of states $(\widetilde{X}_t, \widetilde{U}_t)_{t \geq 0}$ such that

$$P\big((\widetilde{X}_t, \widetilde{U}_t)_{t \geq 0} \in C \,\big|\, \widetilde{q} = q\big) = P\big((X^q_t, U^q_t)_{t \geq 0} \in C\big)$$

almost sure for all measurable events C and $q \in Q$. We now consider $(\widetilde{X}_t)_{t \geq 0}$ to be the random pattern of states that gives the current state of the policyholder and write $\widetilde{A}(t)$ for the corresponding present value. Then one can show that

$$\mathbb{E}\big(\widetilde{A}(t) \,\big|\, (\widetilde{X}_t, \widetilde{U}_t) = (i, u), \widetilde{q} = q\big) = \mathbb{E}\big(A^q(t) \,\big|\, (X^q_t, U^q_t) = (i, u)\big)$$

almost sure for all $i \in \mathcal{S}$, $u \geq 0$, and $q \in Q$. Thus, the prospective reserve of the extended model

$$\widetilde{V}_i(t,u) := \mathbb{E}\big(\widetilde{A}(t)\,\big|\,(\widetilde{X}_t,\widetilde{U}_t)=(i,u),\widetilde{q}=q\big)$$

– now a stochastic quantity – can be represented by just replacing q in (4.3) or (6.1) with the stochastic cumulative transition intensity matrix $\widetilde{q}$. In other words, we can calculate $\widetilde{V}_i(t,u)$ pathwise with the help of (4.3) or (6.1). The actuary is then interested in the probability distribution of $\widetilde{V}_i(t,u)$ in order to

- quantify the systematic mortality risk,
- place premiums and reserves on the safe side,
- estimate future surplus and losses,
- determine solvency reserves,
- value a portfolio of insurance contracts.

First we need to specify the probability distribution of $\widetilde{q}$, and then we can get the probability distribution of $\widetilde{V}_i(t,u)$ either by analytical methods (in general very difficult) or by Monte-Carlo simulation. While the second step is mainly a technical problem, the first step is a really controversial matter. So far, systematic changes of transition rates are not very well understood. What is worse, health insurance policies can have quite long contract terms, and so $\widetilde{q}$ has to be modeled for the long-term.

Concerning mortality, the most prominent stochastic mortality model was already introduced in 1992 by Lee and Carter (1992). But it took some more years until the scientific community saw stochastic modeling of future mortality as a central actuarial task; see, for example, the time-dependent Gompertz-approach of Milevsky and Promislov (2001), the stochastic Perks-Modell of Cairns et al. (2006), the extension of the Lee-Carter model of Delwarde, Denuit, and Eilers (2007), or the forward model of Bauer et al. (2008). Apart from the two state model where the only transition is from active to dead, very few studies investigated time trends in transition rates for multistate actuarial models. Noticeable exceptions are Renshaw and Haberman (2000) and Christiansen, Denuit and Lazar (2010). Renshaw and Haberman (2000) considered the sickness recovery and inception transition rates, together with the mortality rates when sick, which form the basis of the UK continuous mortality investigation Bureau's multistate model. These authors identified the underlying time trends from the observation period 1975-1994 using separate Poisson GLM regression models for each transition. Christiansen, Denuit and Lazar (2010) consider a three state disability model, using a multivariate Lee-Carter type model that is fitted as described in Hyndman and Ullah (2007), that is, by means of a functional data approach.

6.3 Worst-case method

Basically all stochastic models for transition rates to be found in the literature are in some way parametric models that are based on a number of a priori assumptions with scarce empirical evidence, thus exposing the insurer to a considerable model risk, especially in the long term. However, often actuaries do not create a complete stochastic model for $\widetilde{q}$ but only estimate confidence sets $M \subset Q$ for $\widetilde{q} \in Q$. As a result we loose the chance to derive the probability distribution of $\widetilde{V}_i(t,u)$, but we gain robustness with respect to misspecification. Once we have estimated a confidence set M for $\widetilde{q}$, we aim to calculate premiums and reserves in such a way that we are on the safe side with respect to all scenarios in M. From a mathematical perspective, we are looking for

$$\sup_{q \in M} V^q(t,u)\,.$$

It is often convenient to identify also a corresponding worst case scenario

$$q^{WC} = \operatorname*{argmax}_{q \in M} V^q(t,u)\,. \tag{6.3}$$

For example, we can use such a worst-case scenario q^{WC} as a valuation basis of first order, see section 6.1. The maximality of q^{WC} can often be proven with the help of Lemma 6.1. The difficulties are more in the construction of a maximal solution. If we are in the Markovian case of section 2.2, some construction methods can be found in Christiansen (2010a, 2010b).

Example 6.4 (critical illness insurance)**.** Suppose we are in the setting of Example 4.5 with a deferred period of zero, but with a premium that is paid continuously, that is, $B_a(s,t) = -p\,(t-s)\,\mathbf{1}_{[0,T]}(t)$. If the random pattern of states is a Markovian process, then formula (6.1) has the form

$$\begin{aligned} V_a(t) &= \int_t^T -p\,\mathrm{d}\tau - \int_t^T V_a(\tau)\,\varphi(\tau)\,\mathrm{d}\tau + \int_t^T R_{ai}(\tau)\,\mu_{ai}(\tau)\,\mathrm{d}\tau + \int_t^T R_{ad}(\tau)\,\mu_{ad}(\tau)\,\mathrm{d}\tau \\ &= \int_t^T \Big(-p - V_a(\tau)\,\varphi(\tau) + (l_i - V_a(\tau))\,\mu_{ai}(\tau) + (l_d - V_a(\tau))\,\mu_{ad}(\tau)\Big)\mathrm{d}\tau\,, \end{aligned}$$

or, equivalently,

$$\frac{\mathrm{d}}{\mathrm{d}t}V_a(t) = p + V_a(t)\,\varphi(t) - (l_i - V_a(t))\,\mu_{ai}(t) - (l_d - V_a(t))\,\mu_{ad}(t)$$

with starting value $V_a(T) = 0$. Let the confidence set M for q be given by the condition

$$l_{ai}(t) \le \mu_{ai}(t) \le u_{ai}(t)\,, \quad l_{ad}(t) \le \mu_{ad}(t) \le u_{ad}(t)$$

for all $t \le 0$, where $l_{ai}(t), u_{ai}(t), l_{ad}(t), u_{ad}(t)$ are confidence bounds that are estimated from data. Then, according to Christiansen (2010a), any solution of

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}V_a(t) = p + V_a(t)\,\varphi(t) - (l_i - V_a(t))\,\mu_{ai}(t) - (l_d - V_a(t))\,\mu_{ad}(t)\,, \\ &(\mu_{ai}(t), \mu_{ad}(t)) = \operatorname*{argmin}_{(m_{ai},m_{ad}) \in [l_{ai}(t),u_{ai}(t)] \times [l_{ad}(t),u_{ad}(t)]} \Big\{ p + V_a(t)\,\varphi(t) - (l_i - V_a(t))\,m_{ai} - (l_d - V_a(t))\,m \end{aligned}$$

with starting value $V_a(T) = 0$ is a solution of the optimization problem (6.3). We can find a solution of this ordinary differential equation system by using numerical standard techniques.

7 Conclusion

For many different types of health insurances, reasonable actuarial models can be built by describing the medical history of the policyholder by a multistate jump process. Usually the random pattern of states of the policyholder is assumed to be Markovian or semi-Markovian. The probability distribution of the random pattern of states has to be estimated for the future; in doing so it is convenient to work with (cumulative) transition intensities. The past showed that this (cumulative) transition intensities can vary significantly within a contract period, exposing the insurer to a systematic biometric risk. In order to deal with that risk, an actuary needs reliable prognoses of future demographical developments. The prognoses can be in the form of point estimates, confidence estimates, or even stochastic processes for the transition intensities. However, so far very few studies investigated time trends in transition rates for multistate health models.

Acknowledgement

The author thanks Hartmut Milbrodt from the University of Rostock and Hans-Joachim Zwiesler from the University of Ulm for comments that have been very helpful for revising a previous version of the present work.

References

Amsler, M.H., 1968. Les chaines de Markov des assurances vie, invalidité et maladie. Transactions of the 18th International Congress of Actuaries, Vol. 5, 731-746, München.

Bauer, D., Börger, M., Ruß, J., Zwiesler, H.-J., 2008. The volatility of the mortality. Asia-Pacific Journal of Risk and Insurance 3, 184-211.

Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A., Nesbitt, C.J., 1997. Actuarial Mathematics. The Society of Actuaries, Schaumburg, Illinois.

Cairns, A.J.G., Blake, D., and Dowd, K., 2006. A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. Journal of Risk and Insurance 73, 687-718.

Christiansen, M.C., 2007. A joint analysis of financial and biometrical risks in life insurance. Dissertation thesis, University of Rostock, urn:nbn:de:gbv:28-diss2007-0007-6.

Christiansen, M.C., 2010a. Biometric worst-case scenarios for multi-state life insurance policies. Insurance: Mathematics and Economics 47, 190-197.

Christiansen, M.C., 2010b. Making use of netting effects when composing life insurance contracts. Preprint Series 2010-13, University of Ulm.

Christiansen, M.C., Denuit, M.M., Lazar, D., 2010. The solvency II square root formula for systematic biometric risk. Preprint Series 2010-14, University of Ulm.

Davis, M.H.A., 1984. Piecewise-deterministic Markov Processes: A General Class of Non-diffusion Stochastic Models, Journal of the Royal Statistical Society, Series B, 46(3), 353-388.

Delwarde, A., Denuit, M., Eilers, P., 2007. Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: A penalized log-likelihood approach. Statistical Modelling 7, 29-48.

Denuit, M., Robert, C., 2007. Actuariat des Assurances de Personnes: Modélisation, Tarification et Provisionnement. Collection Audit-Actuariat-Assurance, Economica, Paris.

Gatenby, P., Ward, N., 1994. Multiple state modelling. Presented to the Staple Inn Actuarial Society, 1 Feb 1994.

Haberman, S., Pitacco, E., 1999. Actuarial Models for Disability Insurance, Chapman & Hall, Boca Raton.

Helwich, M., 2008. Durational effects and non-smooth semi-Markov models in life insurance. Doctoral dissertation, University of Rostock, urn:nbn:de:gbv:28-diss2008-0056-4.

Hoem, J.M., 1969. Markov chain models in life insurance. Blätter der Deutschen Gesellschaft für Versicherungsmathematik 9, 91-107.

Hoem, J.M., 1972. Inhomogeneous semi-Markov processes, select actuarial tables, and duration-depedence in demography. Population Dynamics, Academic Press New York, 1972, 251-296.

Hyndman, R.J., Ullah, Md.S., 2007. Robust forecasting of mortality and fertility rates: a functional data approach. Computational Statistics and Data Analysis 51, 4942-4956.

Lee, R.D., Carter, L., 1992. Modelling and forecasting the time series of US mortality. Journal of the American Statistical Association 87, 659-671.

Janssen, J., 1966. Application des processus semi-markoviens à un problème d'invalidité. Bulletin de l'Association Royale des Actuaires Français 63, 35-52.

Janssen, J., De Dominicis, R., 1984. Finite non-homogeneous semi-Markov processes: theoretical and computational aspects. Insurance: Mathematics and Economics 3, 157-165.

Milbrodt, H., Helbig, M., 1999. Mathematische Methoden der Personenversicherung. Walter de Gruyter, Berlin.

Milbrodt, H., 2005. Aktuarielle Methoden der deutschen Privaten Krankenversicherung. Verlag Versicherungswirtschaft, Karlsruhe.

Milbrodt, H., Stracke, A., 1997. Markov models and Thiele's integral equations for the prospective reserve. Insurance: Mathematics and Economics 19, 187-235.

Milevsky, M.A., Promislow, S.D. 2001. Mortality derivatives and the option to annuitise. Insurance: Mathematics and Economics 29, 299-318.

Møller, C.M., 1993. A stochastic version of Thiele's differential equation. Scandinavian Actuarial Journal, 1-16.

Möller, H.-G., Zwiesler, H.-J., 1996. Mehrzustandsmodelle in der Berufsunfähigkeitsversicherung. Blätter der DGVFM XXII, 479-499.

Nollau, V., 1980. Semi-Markovsche Prozesse. Akademie-Verlag, Berlin.

Oeppen, J., Vaupel, J.W., 2002. Broken limits to life expectancy. Science 296, 1029-1031.

Pitacco, E., 1995. Actuarial models for pricing disability benefits: Towards a unifying approach. Insurance: Mathematics and Economics 16, 39-62.

Renshaw, A., Haberman, S., 2000. Modelling the recent time trends in UK permanent health insurance recovery, mortality and claim inception transition intensities. Insurance: Mathematics and Economics 27, 365-396.

Rickayzen, B.D., Walsh, D.E.P., 2000. A Model for Projecting the Number of People Who Will Require Long-Term Care in the Future, Parts I-III. Actuarial Research Papers 123, 124 and 125, The City University.

Segerer, G., 1993. The actuarial treatment of the disability risk in Germany, Austria and Switzerland. Insurance: Mathematics and Economics 13, 131-140.

Sverdrup, E., 1965. Estimates and test procedures in connection with stochastic models for deaths, recoveries and transfers between different states of health. Skandinavisk Aktuarietidskrift 48, 184-211.

Waidmann, T.A., Liu, K., .Disability Trends Among Elderly Persons and Implications for the Future. The Journals of Gerontology, Series B, 55(5), 298-307.

Waters, H.R., 1989. Some aspects of the modelling of permanent health insurance. Journal of the Institute of Actuaries 116, 611-624.

Wetzel, C., Zwiesler, H.-J., 2003. Einsatzmöglichkeiten von Markovschen Prozessen bei der Kalkulation von Lebensversicherungsprodukten. Blätter der DGVFM XXVI, 239-253.

Wolthuis, H., 2003. Life insurance mathematics (Tha Markovian model), second edition. Instituut voor Actuariaat en Econometrie, Amsterdam.