Markov Chains and its Applications to Insurance

Faust van der Molen
3471551

Universiteit Utrecht

Mathematics
Bachelor Thesis

Supervised by: Dr. Karma Dajani
June 15, 2017

Abstract

In this thesis we discuss the Markov model for insurance. We start by introducing the reader to Markov chains and deriving some results which we need for the remainder of the thesis. After that we shift our focus to insurance. First we introduce a mathematical way to write down quantities determined in some basic insurance contracts. Next we derive some results about the expected prospective reserves. Then we will implement the Markov chain model in the mathematical language of insurance. Lastly we will take a small detour into the world of statistics, to see how we would obtain estimates for intensities of transition, introduced in the Markov model.

Contents

Chapter 1

Introduction

When thinking about life insurance, usually one or two things come to mind with most people. The first thing that comes to mind is a salesperson trying to sell you a life insurance that you did not know you wanted, but the salesman insists you should buy.

Either that, or they think of some exciting book or TV-series in which a person dies under suspicious circumstances. After an investigation the police find out that the husband or wife of the deceased gets a large sum of money because of a life insurance contract, making things even more suspicious... However let us not delve too deep into that.

So most people know life insurance has something to do with getting money conditional on survival to some period, which is indeed true. Even the ancient Romans had forms of life insurance. Roman soldiers would have a part of their salaries set aside to cover funeral expenses from a mutual fund [3]. While Greece is known for some great mathematicians, the Roman Empire did not produce any mathematicians of note [1]. We may assume that there was no interesting mathematical model underlying the Romans life insurance. The field of statistics was not practiced until hundreds of years later anyway.

One of the first to do a proper study concerning life insurance was Edmund Halley. He produced a paper in 1693, in which he wrote mortality tables, as well as ways to calculate the price of annuities based on the expected life length [2]. In the years between then and now, of course a lot has been said and written on the topic. In this thesis we will mostly rely on a collection of papers from Ragnar Norberg called *Basic Life Insurance Mathematics* [5].

We will write about life insurance from the perspective of a Markov chain model. The first chapter will strictly concern theory on Markov chains. After that, the second chapter starts with some basic things about life insurance, like how to write payment streams, and the definition of basic contract in a mathematical sense. When we know enough about the basics of life insurance, we proceed by implementing the Markov model into the equations we found in the equations for the reserves. Lastly we consider some statistics to be able estimate all needed quantities that arise from assuming the Markov model.

Chapter 2

Markov Chains

In preparation to the next sections, where we make generous use of the properties of Markov chains, we will first describe them in a little more detail. We start with a basic example to make the concept clear, after which we discuss some of the properties and variations of Markov chain models. If you are already familiar with Markov chains, you may skip ahead to Section 2.2.

2.1 An Introductory Example

Consider a student writing an important paper. Suppose we divide the amount of time per day spent on studying in three categories: Studying hard, studying a little and not even bothering. Since students can be short of memory we assume that the amount of studying the student does the next day only depends on the the amount he does today. Figure 2.1 below displays the three states and the arrows indicate the chance of being in that state the next day.

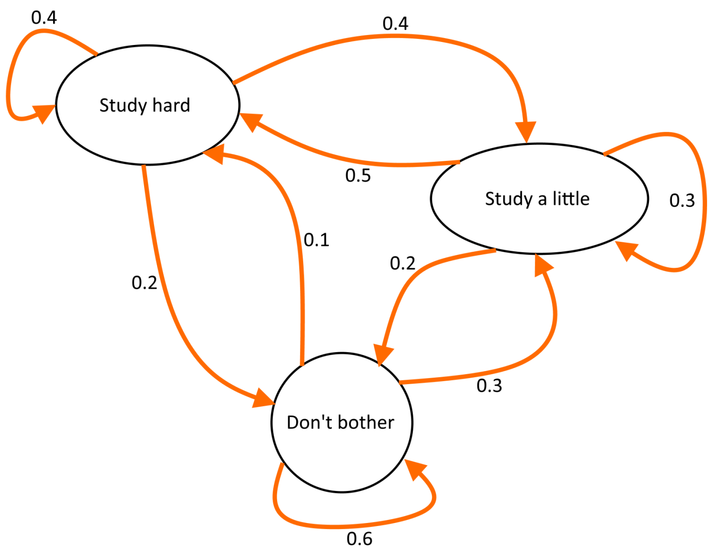


Figure 2.1: An example of a Markov process

Note that in this example we can go from every state to every other state, but this does not have to be the case. We could for example also have added a state 'Dropped out of school, never to return,' in which the probability of going to any state other than itself is 0.

From the graph we can deduce that, for example, if the student is studying a little today, he will study hard with probability 0.5 tomorrow.

The Markov property is just that. The conditional probability of a future state only depends on the present state and not on any other state in the past. Before we make this notion a little more precise we have to distinguish between discrete and continuous time Markov chains.

2.2 The Markov Property

Discrete time Markov chains, such as the one on the previous page, can jump between states only on certain moments of time $t \in \mathbb{N}$. Since the definition of the Markov property for discrete and continuous time are almost analogous we will consider the continuous case directly. For the rest of this chapter we follow the approach of Chapter 7.2 of [5].

We define a stochastic process to be a collection of random variables that describes an evolution over time. For example we write $\{Z(t)|t \in T\}$ for the stochastic process Z and set of possible times T. The process is called a Markov process, if the process possesses the Markov property.

Definition 2.2.1 *Consider a finite state space $\mathcal{Z} = \{0, 1, ..., r\}$ with $j_k \in \mathcal{Z}$, and a stochastic process $Z(t_k)$ on some probability space $(\Omega, \mathcal{H}, P)$ with values in $\mathcal{Z}$. Further let $t_k \in \mathcal{T} \subset \mathbb{R}$ for all k, with $\mathcal{T}$ some ordered index set. We define the Markov property as follows*

$$P[Z(t_k) = j_k \mid Z(t_{k-1}) = j_{k-1}] = P[Z(t_k) = j_k \mid Z(t_h) = j_h, h = 0, 1, ..., k-1], \tag{2.1}$$

where $Z(t_0) = j_0$ by definition and $0 = t_0 < t_1 < \cdots < t_k$.

The Markov property is also often referred to as 'memorylessness', as the future and past are independent when the present is known.

As we will see, it is convenient to introduce some notation for writing the conditional probabilities. From now on, we shall write

$$p_{jk}(t, u) = P[Z(u) = k \mid Z(t) = j]. \tag{2.2}$$

Which we call the transitional probability of transferring from state j at time t to state k at time u. This allows us for example to easily write down (and calculate)

$$P[Z(t_h) = j_h, h = 1, ..., p|Z(t_0) = j_0] = \prod_{h=1}^{p} p_{j_{h-1}j_h}(t_{h-1}, t_h), \tag{2.3}$$

with j_k and t_k as defined above and $Z(t)$ having the Markov property. Furthermore we introduce a compact way to write the probability of going from some state j to a subset of the state space $\mathcal{K} \subset \mathcal{Z}$,

$$p_{j\mathcal{K}}(t,u) = P[Z(u) \in \mathcal{K} \mid Z(t) = j] = \sum_{k \in \mathcal{K}} p_{jk}(t,u). \tag{2.4}$$

Before we move on to the next section, we would like to take the time to review the Chapman-Kolmogorov equation. The equation states that the probability of going from state i at time s to state k at time $u \geq s$ is equal to summing over all possible states the process can go to in the time between. To show this, we first note that the probability of being in any state $j \in \mathcal{Z}$ at a fixed time $t \geq 0$ is 1, or in other words $\sum_{j \in \mathcal{Z}} P[Z(t) = j] = 1$ and that these events $\{Z(t) = j\}$ are disjoint. In that case we may write

$$\begin{aligned}
&P[Z(u) = k \mid Z(s) = i] = P[Z(u) = k, \bigcup_{j \in \mathcal{Z}} Z(t) = j \mid Z(s) = i] \\
&= \sum_{j \in \mathcal{Z}} P[Z(u) = k, Z(t) = j \mid Z(s) = i] \\
&= \sum_{j \in \mathcal{Z}} P[Z(t) = j \mid Z(s) = i] P[Z(u) = k \mid Z(s) = i, Z(t) = j].
\end{aligned}$$

The second line follows from the disjointness of the events $\{Z(t) = j\}$, so that they may be summed. The third line can be easily checked by writing out the conditional probabilities.

Next suppose Z is Markov and that $0 \leq s \leq t \leq u$, then the above equation can be rewritten using the transition probability notation as

$$p_{ik}(s,u) = \sum_{j \in \mathcal{Z}} p_{ij}(s,t) p_{jk}(t,u). \tag{2.5}$$

This equation is known as the *Chapman-Kolmogorov equation.*

2.3 From Transition Probability to Intensity of Transition

Determining the transition probabilities for a discrete time Markov model is pretty straightforward. For a discrete time Markov chain Z with $t \in T = \{t_k | k = 0, 1, ..., q\}$ such that $0 = t_0 < t_1 < \cdots < t_q$ it boils down to letting $p_{jk}(t_{h-1}, t_h) \geq 0$ in such a way that $\sum_{k \in \mathcal{Z}} p_{jk}(t_{h-1}, t_h) = 1$, $h \in T - \{t_0\}$. However, this does not translate well to the continuous model. In the continuous model we work with infinitely small time intervals, so we need to modify our approach when specifying the model.

Norberg gives two equivalent definitions of the so called *intensities of transitions*, allowing smoothness assumptions, that solve this problem [5].

Definition 2.3.1 *Consider a Markov process with state space $\mathcal{Z}$ and states $j, k \in \mathcal{Z}$. Then we define the intensities of transition as*

$$\mu_{jk}(t) = \lim_{h \downarrow 0} \frac{p_{jk}(t, t+h)}{h} \iff$$
$$p_{jk}(t, t+dt) = \mu_{jk}(t)dt + o(dt),$$

where $\frac{o(dt)}{dt} \to 0$ *as* $dt \to 0$*. We suppose that these functions* $\mu_{jk}(t)$ *are piecewise continuous.*

It is important to note that these intensities in general do not represent the transition probabilities. Only in the case where the intensities are approximately constant and $<< 1$ for all $k \neq j$ on an interval dt, then $\mu_{jk}(t)dt \approx p_{jk}(t, t+dt)$. So this is only true when the intensities do not necessarily depend on t, which is an assumption we cannot often make in applications, as we will see in the next section.

To calculate the probability of staying in the same state we first show a few steps in between. We start by looking at the intensity of transition from state j to $\mathcal{K} \subset \mathcal{Z}$, with $j \notin \mathcal{K}$. We define this as follows

$$\mu_{j\mathcal{K}}(t) = \lim_{u \downarrow t} \frac{p_{j\mathcal{K}}(t,u)}{u-t} = \sum_{k \in \mathcal{K}} \mu_{jk}(t). \tag{2.6}$$

If we let $\mathcal{K} = \mathcal{Z} - j$ in the above definition, we write

$$\mu_{j\cdot}(t) = \sum_{k;k \neq j} \mu_{jk}(t), \tag{2.7}$$

to get the intensity of transition out of state j. It is obvious that $p_{j\mathcal{Z}}(t,u) = \sum_{k \in \mathcal{Z}} p_{jk}(t,u) = 1$, for you have to be in some state of the state space. The complement of staying in state j, is going to any other state $k \neq j$. Combining this with the above definition gives

$$p_{jj}(t, t+dt) = 1 - \mu_{j\cdot}(t)dt + o(dt) \tag{2.8}$$

for the transistion probability of staying in the same state.

2.4 The Kolmogorov Differential Equations

Now that we have the intensities of transition, we can use them to write down how the transition probabilities behave over time. To do this, we could just use the Chapman-Kolmogorov equation and insert our definition of the transition probability there. However, we will also give a different, more intuitive argument. The structure of this argument will be used later on as well.

For the transition probability $p_{jk}(t,u)$ we are going to divide the interval (t,u) up into two; $(t, t+dt)$ and $[t+dt, u)$. In the first interval there are two possibilities. With probability $p_{jj}(t,t+dt) = 1 - \mu_{j\cdot}(t)dt + o(dt)$ the process Z may remain in state j after which it goes to state k with probability $p_{jk}(t+dt, u)$. Secondly Z can switch to any other state g with probability $\mu_{jg}(t)dt + o(dt)$, after which the probability of ending up in state k is given by $p_{gk}(t+dt, u)$. Combining the above we get

$$p_{jk}(t,u) = (1 - \mu_{j\cdot}(t)dt)p_{jk}(t+dt,u) + \sum_{g;g \neq j} \mu_{jg}(t)dt p_{gk}(t+dt,u) + o(dt). \tag{2.9}$$

Now let $d_t p_{jk}(t,u) = p_{jk}(t+dt,u) - p_{jk}(t,u)$ then we get

$$d_t p_{jk}(t,u) = \mu_{j\cdot}(t)dt p_{jk}(t+dt,u) - \sum_{g;g\neq j} \mu_{jg}(t)dt p_{gk}(t+dt,u) + o(dt). \quad (2.10)$$

These equations are called the *Kolmogorov backward differential equations* and along with the conditions $p_{jk}(u,u) = \delta_{jk}$, the Kronecker delta, uniquely determine the functions of the transition probabilities.

Remember that we defined the intensities as piecewise continuous functions, so that we may divide the above equation by dt and take the limit for $dt \to 0$ to obtain

$$\frac{\partial}{\partial t} p_{jk}(t,u) = \mu_{j\cdot}(t) p_{jk}(t,u) - \sum_{g;g\neq j} \mu_{jg}(t) p_{gk}(t,u). \quad (2.11)$$

The equation is called backwards because we start by looking back at the very beginning t and some very small time $t+dt$ later. As the name suggests, there is a forward equation as well and with argumentation similar to the backwards equation it is given by

$$d_t p_{ij}(s,t) = \sum_{g;g\neq j} p_{ig}(s,t)\mu_{gj}(t)dt - p_{ij}(s,t)\mu_{j\cdot}(t)dt. \quad (2.12)$$

Now that we have a suitable method of finding the transitional probabilities, we want to make a (maybe obvious) distinction between $p_{jj}(t,u)$, and staying in state j the entire time. The former allows the process to go to any other state during time (t,u). It just needs to start and end up in state j. For the latter we define

$$p_{\overline{jj}}(t,u) = P[Z(\tau) = j, \tau \in (t,u] \mid Z(t) = j]. \quad (2.13)$$

Using the same reasoning as in the above, we condition on what happens in the small interval $(t, t+dt)$. Now the only possibility is staying in state j with probability $1 - \mu_{j\cdot}(t)dt + o(dt)$. Next, multiply by $p_{\overline{jj}}(t+dt,u)$, as this corresponds to staying in state j for the rest of the time $(t+dt,u)$ as well. We then get

$$p_{\overline{jj}}(t,u) = (1 - \mu_{j\cdot}(t)dt)p_{\overline{jj}}(t+dt,u) + o(dt), \quad (2.14)$$

and given that $p_{\overline{jj}}(u,u) = 1$ we conclude that

$$p_{\overline{jj}}(t,u) = e^{-\int_t^u \mu_{j\cdot}(s)ds}. \quad (2.15)$$

In conclusion of this theoretical part about Markov chains, we are going to give one more different way of writing the transitional probabilities. Our starting point is the backward differential equation which we derived on the previous page. From this we will deduce the backwards integral equations. We start with multiplying by integrating factor $e^{\int_t^u \mu_{j\cdot}(s)ds}$ to get

$$e^{\int_t^u \mu_{j\cdot}(s)ds} \frac{\partial}{\partial t} p_{jk}(t,u) = e^{\int_t^u \mu_{j\cdot}(s)ds} \mu_{j\cdot}(t) p_{jk}(t,u) - e^{\int_t^u \mu_{j\cdot}(s)ds} \sum_{g;g\neq j} \mu_{jg}(t) p_{gk}(t,u). \quad (2.16)$$

We know that $\frac{\partial}{\partial t}e^{\int_t^u \mu_{j\cdot}(s)ds} = \frac{\partial}{\partial t}e^{-\int_u^t \mu_{j\cdot}(s)ds} = -e^{\int_t^u \mu_{j\cdot}(s)ds}\mu_{j\cdot}(t)$, so that the above equation reduces to:

$$\frac{\partial}{\partial t}(e^{\int_t^u \mu_{j\cdot}(s)ds}p_{jk}(t,u)) = -e^{\int_t^u \mu_{j\cdot}(s)ds}\sum_{g;g\neq j}\mu_{jg}(t)p_{gk}(t,u). \tag{2.17}$$

Next we integrate over $(t,u]$ to obtain the following expression on the left side:

$$\begin{aligned}&\int_t^u \frac{\partial}{\partial t}e^{\int_\tau^u \mu_{j\cdot}(s)ds}p_{jk}(\tau,u)d\tau = \frac{\partial}{\partial t}\left(-\int_u^t e^{-\int_u^\tau \mu_{j\cdot}(s)ds}p_{jk}(\tau,u)d\tau\right)\\ &= -\, e^{\int_t^u \mu_{j\cdot}(s)ds}p_{jk}(t,u) + \delta_{jk}\end{aligned}$$

The Kronecker-delta follows from the fact that $\int_u^u e^{\int_u^\tau \mu_{j\cdot}(s)ds}p_{jk}(\tau,u)d\tau = e^0 p_{jk}(u,u) = \delta_{jk}$. Also note that we used the Leibniz integral rule for changing the order of integration/derivation for continuous functions.

For the final step we subtract the Kronecker-delta from both sides and then multiply by $-e^{-\int_t^u \mu_{j\cdot}(s)ds} = -p_{\overline{jj}}(t,u)$, which gives us the expression we want:

$$p_{jk}(t,u) = \int_t^u p_{\overline{jj}}(t,\tau)\sum_{g;g\neq j}\mu_{jg}(\tau)p_{gk}(\tau,u)d\tau + \delta_{jk}p_{\overline{jj}}(t,u). \tag{2.18}$$

As it is with the differential equations, there is also a forward equivalent of the integral equation, derived in the same way, which reads

$$p_{ij}(s,t) = \delta_{ij}p_{\overline{ii}}(s,t) + \sum_{g;g\neq j}\int_s^t p_{ig}(s,\tau)\mu_{gj}(\tau)p_{\overline{jj}}(\tau,t)d\tau. \tag{2.19}$$

We note that intuitively these equations make a lot of sense.

Let us look at the backwards equation for $p_{jk}(t,u)$. The final term represents the event that the process stays uninterruptedly in state j, while the first term represents the event that the process stays in state j right up until time τ after which it goes to some other state g for the first time at time τ. We sum over all possible states and integrate τ over $(t,u]$ to allow switching states at all moments in time between t and u.

Now that we have a firm theoretical basis on Markov chains, we will do a short introduction on life insurance and then continue with the application of Markov chains in insurance.

Chapter 3

Life Insurance

In this chapter we will first cover how to write the payment streams occurring as a result of insurance contracts in a mathematical sense. After that we discuss some basic insurance contracts and their expected value and variance. Then we introduce the Markov model for insurance and see what the consequences are in terms of the expected reserves. Lastly we deduce a result for the higher order moments of the reserves and add some more detail to the model, using a Markov chain for the interest rate as well. To do this, we make use of Chapters 2, 4 and 7 of [5].

3.1 Benefits, Premiums and Interest

A contract in insurance consists of some arrangement of the policy holder paying a sum of money to the insurer and receiving benefits at certain events, to be paid out by the insurer. It will be useful for the rest of this chapter to make this more precise.

From now on we assume that a contract will be valid for a certain period of time starting at 0 and ending at n. Furthermore, we let $\{A_t\}_{t\geq 0}$ be the payment function. We define this function to be the incomes minus the outgoes from the insurer's perspective. We assume this function is right-continuous, piecewise differentiable and of finite variation. The function A_t defined as in the equation below can be interpreted as the total amount paid by either the insurer or the policy holder in the interval $[0, t]$. This amount may be positive or negative.

$$A_t = A_0 + \int_0^t dA_\tau. \tag{3.1}$$

This elegant notation, introduced in [5], is short for $\int_0^t dA_\tau = \int_0^t a_\tau d\tau + \sum_{0<\tau\leq t} \Delta A_\tau$, where a_τ denotes the continuous stream of payments and the summation represents the lump sum payments. The term ΔA_τ is defined as $\Delta A_\tau = A_\tau - A_{\tau-} = A_\tau - \lim_{s\uparrow\tau} A_s$.

The money in possession of the insurer may be invested in such a way that the insurer earns interest. Let the function $S(s,t)$ denote the value of one unit at time t that is invested at time $s \leq t$ and let us write S_t as a shorthand for $S(0,t)$.

If we wish to disallow for a situation where one withdraws the amount $S(s,t)$ at time t and directly reinvests the same amount in the same fund to make profits, we must make the following assumption:

$$S(s,u) = S(s,t)S(t,u). \tag{3.2}$$

This is the principle of no-arbitrage. It easily follows from the above that

$$S(s,t) = \frac{S_t}{S_s}, \tag{3.3}$$

and if we choose s to be 0, we get $S_0 = 1$ which complies with our definition. Given these constraints we will assume that the general accumulation function for $t \leq u$ is of the form

$$S(t,u) = e^{\int_t^u r_\tau d\tau}. \tag{3.4}$$

Whereas if $t \leq s$ we write the general discount function as

$$S(s,t) = S^{-1}(t,s) = e^{-\int_s^t r_\tau d\tau}. \tag{3.5}$$

This allows us to calculate the value at time t of a payment dA_τ made in the (infinitesimal) small time interval around τ. Let us write $e^{\int_s^t r_\tau d\tau} = e^{\int_s^t r}$ to avoid cumbersome notation. Then the value of dA_τ at time t can be written using the discount functions given above

$$S(\tau,t)dA_\tau = e^{\int_\tau^t r} dA_\tau = e^{\int_0^t r} e^{-\int_0^\tau r} dA_\tau. \tag{3.6}$$

Now we look for an expression for the value at time t of the entire contract, but before we do that, we introduce some more notation. Suppose we wanted to write equation (3.1) as only an integral, we would somehow need to include A_0, the payments made at time 0, in the integral. We can do this by taking the left limit of the integral, so then

$$A_t = A_0 + \int_0^t dA_\tau = \lim_{s \uparrow 0} \int_s^t dA_\tau, \tag{3.7}$$

which we will write for short as

$$\int_{0-}^t dA_\tau := \lim_{s \uparrow 0} \int_s^t dA_\tau. \tag{3.8}$$

From (3.6) we know the value at time t of a payment dA_τ. If we want to know the value at time t of all payments made during the contract, we just sum over all time intervals in $[0,n]$. Then if we use the above notation, the value at time t of all payments during the contract becomes:

$$e^{\int_0^t r} \int_{0-}^n e^{-\int_0^\tau r} dA_\tau = U_t - V_t, \tag{3.9}$$

where the right hand side is a decomposition in $U_t = \int_{0-}^t e^{\int_\tau^t r} dA_\tau$, also called the cash balance, and $V_t = -\int_t^n e^{-\int_t^\tau r} dA_\tau$ the future liability, or prospective reserve.

3.2 Forms of Insurance

In life insurance every policy holder may get a contract tailored entirely to his or her needs and preferences. However, there are three general forms of insurance which we will consider in this chapter. These contracts are the (pure) endowment insurance, the life assurance and the life annuity. We can describe them in a little more detail now that we introduced some notation for payment streams and interest rates.

Pure Endowment Insurance
An insurance that pays out a sum of money if the policy holder survives a prespecified number of years n, is called a n-year pure life endowment insurance. We assume for the sake of simplicity that the interest rate is fixed at r, and that the amount of money paid out is 1 unit. Now let T_x be a random variable for the amount of years a person aged x has yet to live. Also let I_t be the indicator function that returns 1 if the policy holder aged x survives until time t, or $[T_x > t]$. Then we can write the present value of the contract at time 0 as:

$$PV^{e;n} = e^{-rn} I_n, \tag{3.10}$$

where the left hand side topscript $e;n$ signifies that this present value concerns an n-year $e$ndowment insurance.

We let ${}_tp_x$ denote the probability of a person aged x surviving until time t (see Appendix A). Then ${}_np_x$ is the probability of surviving until the end of the contract. We then define ${}_nE_x := E[PV^{e;n}]$, and we write the expected value as

$${}_nE_x = E[e^{-rn} I_n] = e^{-rn} E[I_n] = e^{-rn} P[T_x > n] = e^{-rn} {}_np_x. \tag{3.11}$$

To get a measure of risk involved in these types of insurances it is useful to not only calculate the expected value, but the variance as well. To get the variance, we need the second moment of $PV^{e;n}$. We calculate the q-th non-central moment using $I_n^q = I_n$, then

$$E[(PV^{e;n})^q] = E[(e^{-rn} I_n)^q] = e^{-qrn} E[I_n] = e^{-qrn} {}_np_x = {}_nE_x^{(qr)}, \tag{3.12}$$

where the topscript in ${}_nE_x^{(qr)}$ denotes the force of interest. Using this knowledge, we can write the variance as:

$$V[PV^{e;n}] = E[(PV^{e;n})^2] - {}_nE_x^2 = {}_nE_x^{(2r)} - {}_nE_x^2. \tag{3.13}$$

Life Assurance
For the life assurance we discern two forms: The n-year term insurance, and the n-year endowment insurance. For an n-year term insurance a lump sum payment is made in the event of death within n years. The present value at time 0 of such an insurance is:

$$PV^{ti;n} = e^{-rT_x}(1 - I_n). \tag{3.14}$$

So it is equal to 0 if the policy holder survives past the duration of the contract, and e^{-rT_x} if he does not survive. To get the expected value we need to know the density of T_x, because we want to use the rule $E[g(X)] = \int g(x)f(x)dx$, where g is a function, X a random variable and $f(x)$ the density function of X. The density of T_x is given by $f(t|x) = {}_tp_x\mu_{x+t}$, for more details see Appendix B. The expected value is then given by:

$$E[PV^{ti;n}] = \bar{A}^1_{x:\overline{n}|} = \int_0^n e^{-r\tau} f(\tau|x)d\tau = \int_0^n e^{-r\tau}{}_\tau p_x\mu_{x+\tau}d\tau, \tag{3.15}$$

where $\bar{A}^1_{x:\overline{n}|}$ is actuarial notation for a term insurance for a x aged person with n year contract length and $\mu_{x+\tau}$ denotes the force of mortality, also called the hazard rate, at age $x+\tau$. From now on we will write some actuarial notation without mentioning it explicitly. We expect the reader to turn to the Appendix for explanations of unknown symbols.

Moving on to the variance, we find

$$E[(PV^{ti;n})^q] = \bar{A}^{(qr)}{}^1_{x:\overline{n}|} \text{ and so } V[PV^{ti;n}] = \bar{A}^{(2r)}{}^1_{x:\overline{n}|} - (\bar{A}^1_{x:\overline{n}|})^2 \tag{3.16}$$

similar to the case of a pure life endowment insurance.

An n-year endowment insurance is different from the n-year term insurance in the fact that the insurance also pays out if the policy holder survives until the end of the contract. The quantities of interest are derived in the same way as the previous two examples and are:

$$PV^{ei;n} = e^{-r(T_x \wedge n)} \tag{3.17}$$

$$E[PV^{ei;n}] = \bar{A}_{x:\overline{n}|} = \int_0^n e^{-r\tau}{}_\tau p_x\mu_{x+\tau}d\tau + e^{-rn}{}_np_x = \bar{A}^1_{x:\overline{n}|} + {}_nE_x \tag{3.18}$$

$$E[(PV^{ei;n})^q] = \bar{A}^{(qr)}_{x:\overline{n}|} \tag{3.19}$$

$$V[PV^{ei;n}] = \bar{A}^{(2r)}_{x:\overline{n}|} - \bar{A}^2_{x:\overline{n}|} \tag{3.20}$$

Life Annuity

A life annuity is a form of insurance that pays out 1 unit at fixed intervals, e.g. per year or per month. There are several different types of life annuity contracts possible. The version we consider is the continuous version of the n-year temporary life annuity. This contract pays out n years, given that the policy holder is alive. The associated present value at time 0 is

$$PV^{a;n} = \bar{a}_{\overline{T_x \wedge n}|} = \int_0^{T_x \wedge n} e^{-r\tau}d\tau = \frac{1 - e^{-r(T_x \wedge n)}}{r}. \tag{3.21}$$

The formula for the expected value is again derived with the rule $E[g(X)] = \int g(x)f(x)dx$ and $f(t|x) = {}_tp_x\mu_{x+t}$ as density function for T_x.

$$\begin{aligned} E[PV^{a;n}] = \bar{a}_{x\overline{n}|} &= \int_0^n \frac{1 - e^{-r\tau}}{r} f(\tau|x)d\tau + \frac{1 - e^{-rn}}{r}{}_np_x \\ &= \int_0^n \bar{a}_{\overline{\tau}|}{}_\tau p_x\mu_{x+\tau}d\tau + \bar{a}_{\overline{n}|}{}_np_x. \end{aligned} \tag{3.22}$$

Let us translate that to words. The last term covers the case where $T_x \geq n$. The first term under the integral sign $\bar{a}_{\overline{\tau}|}$ is the value of a person surviving up to time τ. The terms ${}_\tau p_x \mu_{x+\tau}$ denote the conditional probability density function of mortality. By integrating τ over the interval $(0, n)$ we sum the probability of surviving over all infinitesimally small intervals $(\tau, \tau + d\tau)$ and multiply by the corresponding present value. This precisely gives us our expected value. Norberg argues that this expected value may be written much more elegantly as

$$\bar{a}_{x\overline{n}|} = \int_0^n e^{-r\tau} {}_\tau p_x d\tau. \tag{3.23}$$

This can be proven by integrating $\bar{a}_{\overline{n}|n} p_x$ by parts, or by arguing that the expected present value at time 0 equals $e^{-r\tau} d\tau {}_\tau p_x$ for payments in the small interval of $(\tau, \tau + d\tau)$. Then the integral is realized by summing over all small intervals.

Note that this reasoning can also be applied to give the the expected values for life assurances 'directly'.

The variance and expected value of the q-th moment are given below. When the equation for the moments is known, it is easy to find the variance. However, finding the equation for the moments takes a bit more time.

$$E[(PV^{a;n})^q] = \frac{q}{r^{q-1}} \sum_{p=1}^{q} (-1)^{p-1} \binom{q-1}{p-1} \bar{a}^{(pr)}_{x\overline{n}|} \tag{3.24}$$

$$V[PV^{a;n}] = \frac{2}{r}(\bar{a}_{x\overline{n}|} - \bar{a}^{(2r)}_{x\overline{n}|}) - \bar{a}^2_{x\overline{n}|} \tag{3.25}$$

We start by noting that $PV^{a;n} = \frac{1-PV^{ei;n}}{r}$ so that we can write

$$E[(PV^{a;n})^q] = E[(\frac{1 - PV^{ei;n}}{r})^q] = \frac{1}{r^q} E[(1 - PV^{ei;n})^q]. \tag{3.26}$$

We can expand the term inside the expectation to get

$$\begin{aligned} \frac{1}{r^q} E[(1 - PV^{ei;n})^q] &= \frac{1}{r^q} E\Big[\sum_{p=0}^{q} \binom{q}{p} (1)^{q-p} (-PV^{ei;n})^p\Big] \\ &= \frac{1}{r^q} E\Big[\sum_{p=0}^{q} \binom{q}{p} (-1)^p (PV^{ei;n})^p\Big] \end{aligned} \tag{3.27}$$

As the expectation only depends on what happens in the terms $(-PV^{ei;n})^p$, we may get it into the summation.

$$\frac{1}{r^q} E\Big[\sum_{p=0}^{q} \binom{q}{p} (-1)^p (PV^{ei;n})^p\Big] = \frac{1}{r^q}\Big(\sum_{p=0}^{q} \binom{q}{p} (-1)^p E[PV^{ei;n})^p]\Big) \tag{3.28}$$

We already know from the previous page that $E[PV^{ei;n})^p] = \bar{A}^{(pr)}_{x:\overline{n}|}$ and it is easy to check that $\bar{A}^{(pr)}_{x:\overline{n}|} = 1 - (pr)\bar{a}^{(pr)}_{x\overline{n}|}$. Inserting this in the equation gives:

$$\frac{1}{r^q}\Big(\sum_{p=0}^{q} \binom{q}{p} (-1)^p [1 - (pr)\bar{a}^{(pr)}_{x\overline{n}|}]\Big) \tag{3.29}$$

Now we are getting close. We need to make two more steps to complete the calculation. The first step is note that $(-1+1)^q = \sum_{p=0}^{q}(-1)^p(1)^{q-p} = 0^q = 0$. Secondly we need $\binom{q}{p}\binom{p}{r} = \binom{q}{r}\binom{q-r}{p-r}$, where in our case $r = 1$, so that $\binom{q}{r}$ reduces to just q. So now we can write

$$\frac{1}{r^q}\Big(\sum_{p=0}^{q}\binom{q}{p}(-1)^p - \sum_{p=0}^{q}\binom{q}{p}(-1)^p(pr)\bar{a}^{(pr)}_{x\overline{n}|}\Big) \tag{3.30}$$

$$=\frac{1}{r^{q-1}}\Big(0 - \sum_{p=0}^{q}\binom{q}{p}\binom{p}{1}(-1)^p\bar{a}^{(pr)}_{x\overline{n}|}\Big) \tag{3.31}$$

$$= \frac{q}{r^{q-1}}\sum_{p=1}^{q}(-1)^{p-1}\binom{q-1}{p-1}\bar{a}^{(pr)}_{x\overline{n}|}, \tag{3.32}$$

which is exactly the expression we were looking for. The summation starts at $p = 1$, since for $p = 0$ we have $\binom{q-1}{-1} = 0$.

We have given expressions for the moments of these basic insurances. Below is a table of the first three moments, taken from [5], which can be calculated using the above. The variation can be interpreted as a measure of riskiness of the insurance. Values in this table were calculated using $x = 30$, $n = 30$, $r = \ln(1.045)$ and μ according to the G82M mortality law in [5].

	Pure Endowment	Term Insurance	Endowment Insurance	Life Annuity
Expected Value	0.2257	0.06834	0.2940	16.04
Variation	0.4280	2.536	0.3140	0.1308
Skewness	-1.908	2.664	4.451	-4.451

Table 3.1: Moments of the forms of insurance discussed before

3.3 Reserves and Thiele's Differential Equation

Just as a bank needs to have enough reserves to ensure they can accommodate the withdrawals from accounts of their customers, insurance companies need to have certain reserves as well to pay their policy holders benefits. The previous section displays that the variation among different types of insurance introduces some form of risk. While insurance companies try to combat this risk by diversification among the policy holders, they simply cannot have an infinite amount of customers to let the risk tend to zero. For that reason an insurance company always must have reserves.

We discern two types of reserves. First there is the retrospective reserve, which equals the past incomes. The second reserve is called the prospective reserve, based on expected expenses. For both reserves we can ask ourselves the question of how large they should be.

An obvious answer to this question is that the expected present value of the incomes should be at least as large as the expected expenses. This is also known in actuarial science as the *equivalence principle*. We will consider a contract that is a combination of the insurance forms discussed in the previous section. For this generalized contract we will write down some formulas for the reserve. Then we will take a look at Thiele's differential equation. At that point we are ready to start implementing the Markov chain model in insurance in the next section.

The setup for this generalized contract is as follows. We assume the policy holder to be aged x upon initiating the n-year contract. Since we assume a general setup, the benefits of the contract comprises a term insurance of b_t to be payed in the event of death during the contract and a pure endowment insurance of b_n payable if the policy holder survives the duration of the contract. For the premium payments, the contract specifies a lump sum payment of π_0 and an annuity that is to be payed continuously at π_t per time unit, conditional on survival at time $t \in (0, n)$.

Suppose that we wanted to calculate an appropriate premium function π, then by the equivalence principle, we would have to solve

$$\pi_0 + \int_0^n e^{-\int_0^\tau r} {}_\tau p_x \pi_\tau d\tau = \int_0^n e^{-\int_0^\tau r} {}_\tau p_x \mu_{x+\tau} b_\tau d\tau + b_n e^{\int_0^n r} {}_n p_x, \qquad (3.33)$$

as this equation denotes the present expected value of incomes to equal the expected present values of the benefits.

During the contract more information naturally becomes available, most importantly concerning the survival of the policy holder. As time progresses we can continuously asses the renewed conditional expectations. In fact insurance laws require the equivalence principle to be applicable not only at time 0, but at any time $t \in (0, n)$. This gives rise to an equation for the prospective reserve V_t, given by

$$V_t = \int_t^n e^{-\int_t^\tau r} {}_{\tau-t} p_{x+t} (\mu_{x+\tau} b_\tau - \pi_\tau) d\tau + b_n e^{-\int_t^n r} {}_{n-t} p_{x+t}. \qquad (3.34)$$

Using this equation we will derive Thiele's differential equation. First we have to do some rewriting. If we let ${}_{\tau-t} p_{x+t} = e^{-\int_t^\tau \mu_{x+s} ds}$, then we can reformulate this as

$$V_t = \int_t^n e^{-\int_t^\tau r_s + \mu_{x+s} ds} (\mu_{x+\tau} b_\tau - \pi_\tau) d\tau + e^{-\int_t^n r_s + \mu_{x+s} ds} b_n. \qquad (3.35)$$

It is convenient to rewrite the integral part using:

$$\int_t^n e^{-\int_t^\tau q_s ds} d\tau = \int_t^n e^{-\int_0^\tau q_s ds + \int_0^t q_s ds} d\tau = e^{\int_0^t q_s ds} \int_t^n e^{-\int_0^\tau q_s ds} d\tau, \qquad (3.36)$$

where $q_s = r_s + \mu_{x+s}$. Now let us take the derivative of V_t with respect to t and use the above equation to get

$$\frac{d}{dt} V_t = \frac{d}{dt} \int_t^n e^{-\int_t^\tau q_s ds} (\mu_{x+\tau} b_\tau - \pi_\tau) d\tau + \frac{d}{dt} e^{-\int_t^n q_s ds} b_n \qquad (3.37)$$

$$= \frac{d}{dt} \left[e^{\int_0^t q_s ds} \int_t^n e^{-\int_0^\tau q_s ds} (\mu_{x+\tau} b_\tau - \pi_\tau) d\tau \right] + q_t e^{-\int_t^n q_s ds} b_n. \qquad (3.38)$$

We apply the chain rule to write out the derivative

$$q_t e^{\int_0^t q_s ds} \int_t^n e^{-\int_0^\tau q_s ds}(\mu_{x+\tau} b_\tau - \pi_\tau) d\tau + q_t e^{-\int_t^n q_s ds} b_n - (\mu_{x+t} b_t - \pi_t). \tag{3.39}$$

The terms with $q_t = r_t + \mu_{x+t}$ in front of it add precisely up to V_t so that we finally get Thiele's differential equation:

$$\frac{d}{dt} V_t = (r_t + \mu_{x+t}) V_t + \pi_t - b_t \mu_{x+t}. \tag{3.40}$$

While we already had a means of calculating the reserve directly, without using Thiele's equation, it is a useful tool to monitor the development of the reserve over time and we will revisit it a few times in the part to come.

3.4 Introducing Markov Chains

The material covered up to now introduced the notion of Markov chains and stated some notation and formulas often used in life insurance. This was done with the unification of both subjects in mind. We start with a very easy model with only two states and will gradually build to some more interesting cases.

Two state model
We consider the model where the policy holder can be either alive or dead. This represented in the graph below by an 'alive' state 0 and a 'dead' state 1, which is absorbing: You can not get out of the state. His life length is given by T, a positive random variable with cumulative distribution function $F(t) = P[T \leq t]$ and survival function $\bar{F}(t) := 1 - F(t)$. More details on this survival function can be found in Appendix B.

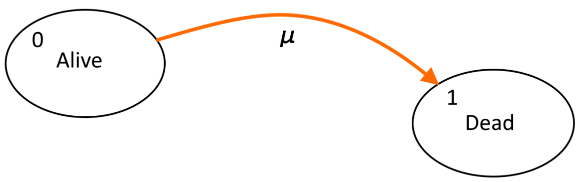


Figure 3.1: A two state Markov Chain

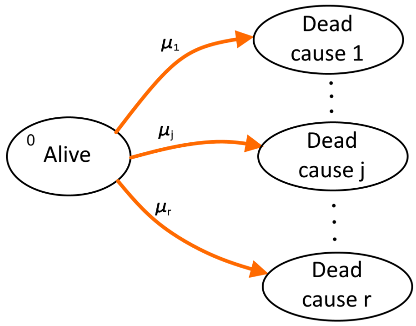
We can easily define the state process Z. For an n-year contract it is given by the indicator function

$$Z(t) = \mathbf{1}[T \leq t], \; t \in [0, n], \tag{3.41}$$

that tells us which state the process is in, at any time $t \in [0, n]$. The process clearly has the Markov property, as the transition probability only depends on the current state. The transition probabilities are determined by $p_{00}(s,t) = \bar{F}(t)/\bar{F}(s) = e^{-\int_s^t \mu}$, or the probability of surviving up to time t, given being alive at s. Applying the Chapman-Kolmogorov equation to this model gives $p_{00}(s,u) = p_{00}(s,t) p_{00}(t,u)$, which is also trivial.

Multiple causes of death

The previous example was rather simple, but it was a good way to get used to how we will analyze these kinds of models. The next model we consider is one with various absorbing states, signifying death from r different causes (disease, accident, etc.). Since all intensities have their origin in state 0 we let $\mu_{0j} = \mu_j$ in this example.



To get the transition probability of staying alive we remember from Section 2.3 that we may sum all the intensities of leaving state 0 to get the total mortality intensity.

$$\mu(t) = \sum_{j=1}^{r} \mu_j(t) \tag{3.42}$$

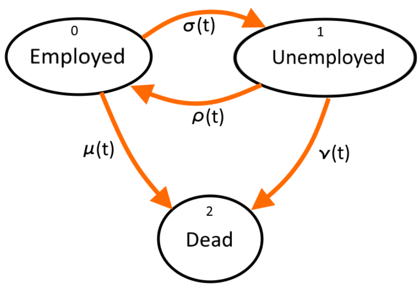
This implies that the transitional probability reduces to the same $p_{00}(s,t) = e^{-\int_s^t \mu}$ as in the previous example. This model however allows us to study the causes of death in more detail, to get a better mortality law in the aggregate, as argued by Norberg. The rest of the transition probabilities also easily follow from the integral equations given in Section 2.4 and are

$$p_{0j}(t,u) = \int_t^u e^{-\int_s^t \mu} \mu_j(\tau) d\tau. \tag{3.43}$$

Norberg comments that the mortality intensities μ_j are a basic but very powerful tool for this model. Since overall mortality is the sum of the different mortality intensities, an increase in one of the intensities means a decrease for all transitional probabilities of other causes of death. Comparing this to the real world, medical advances have decreased the chances of dying from all sort of diseases. So the fact that people nowadays increasingly die from cancer and heart diseases can be explained by the diminishing intensities of the other causes of death. He concludes that the transition probabilities are just the result of the interactions between the more basic intensities.

Multiple alive states

Now consider a model where the payment scheme depends on the state when alive. In the figure we take an unemployment insurance as example, but the model is applicable to many other insurances. A person can be employed (state 0), unemployed (state 1) and dead (state 2). Movement between the states 0 and 1 is



possible, but state 2 is absorbing. This results in the image on the right. For a person who is employed at time s we consider the Kolmogorov forward differential equations. As the transition probabilities sum up to one, we only have to calculate two of the three, to get the third one

$$\frac{\partial}{\partial t} p_{00}(s,t) = p_{01}(s,t)\rho(t) - p_{00}(s,t)\big[\mu(t) + \sigma(t)\big] \tag{3.44}$$

$$\frac{\partial}{\partial t} p_{01}(s,t) = p_{00}(s,t)\sigma(t) - p_{01}(s,t)\big[\nu(t) + \rho(t)\big]. \tag{3.45}$$

They can be solved by setting $p_{00}(s,s) = 1$ and $p_{01}(s,s) = 0$.

Now that we have reviewed some basic models and have seen how to calculate transition probabilities we are going to expand a bit to a more general multi-state contract.

3.5 A More General Model

In this section we revisit Thiele's differential equation, as well as the valuation of the reserve in the Markov chain model. The setup is the same as before, however we now introduce an indicator process I_j and a counting process N_{jk}. The notation $Z(\tau-) := \lim_{t\uparrow\tau} Z(t)$ is used to define the moment of the changing of states, reminiscent of the notation introduced right after equation (3.1). The processes are then defined by

$$I_j(t) = \mathbf{1}[Z(t) = j] \tag{3.46}$$

$$N_{jk}(t) = |\big\{\tau; Z(\tau-) = j, Z(\tau) = k, \tau \in (0,t]\big\}|. \tag{3.47}$$

So $I_j(t)$ indicates whether the process Z is in state j at time t. $N_{jk}(t)$ then counts the number of jumps from state j to state k, $(k \neq j)$, during the time interval $(0,t]$.

In terms of insurance a change in state often means a change in the payments to be made. We want to accommodate that when the state changes, a lump sum is payed, or that during a certain time in state k an annuity is payed. We use $B(t)$ to denote the benefits minus the premiums, note that this is $-A(t)$ specified in equation (3.1). In that equation we (implicitly) used $dA_t = a_t dt + \Delta A_t$. We want to write and use $dB(t)$ in the same way, while altering it to the Markov model. We define the contractual agreements on payments to be made while in state k as

$$dB_k(t) = b_k(t)dt + B_k(t) - B_k(t-), \tag{3.48}$$

where $b_k(t)$ represents the continuous stream of payments in state k, while $B_k(t) - B_k(t-)$ captures endowment payments to be paid at time t.

The payment function is not complete yet. We also need to include payments made when switching states. We denote b_{kl} as the amount that is payed when jumping from state k to state l. We can use the counting process $N_{kl}(t)$ for counting the number of jumps up to time t. Moreover if a jump takes place in the time interval $(t, t+dt)$ from state k to l we get $dN_{kl}(t) = N_{kl}(t+dt) - N_{kl}(t) = 1$, so that we can use this as an indicator for a jump taking place.

Since we want $B(t)$ to represent the payment function of a multi-state contract, we should include the possibility to be in different states in $dB(t)$. We achieve this by multiplying by the indicator function and then summing over all possible states k, to get $\sum_k I_k(t)dB_k(t)$. For the payments b_{kl} we need to sum over all states $k \neq l$ for both indices. Combining the above, we arrive at

$$dB(t) = \sum_k I_k(t)dB_k(t) + \sum_{k \neq l} b_{kl}(t)dN_{kl}(t), \tag{3.49}$$

where we assume the functions b_k and b_{kl} to be piecewise continuous.

Not only the payment streams can be rewritten for the Markov model, but the reserves as well. When we first mentioned the prospective reserve, we gave the formula $V(t) = \int_t^n e^{-\int_t^\tau r} dB(\tau)$. In the Markov chain model the expected present value of the reserve at time t is obviously conditional on the state j the contract is in. If we take the expectation in $dB(t)$ above and insert it in the equation for V_t we get the expected prospective reserves conditional on j. To get there we note that

$$E[I_k(\tau)|Z(t) = j] = p_{jk}(t, \tau) \tag{3.50}$$

$$E[dN_{kl}(\tau)|Z(t) = j] = p_{jk}(t, \tau)\mu_{kl}(\tau)d\tau. \tag{3.51}$$

Then if we denote the prospective reserves conditional on j as $V_j(t)$, we get

$$\begin{aligned} V_j(t) =& E\Big[\int_t^n e^{-\int_t^\tau r} dB(\tau)\Big|Z(t) = j\Big] = \int_t^n e^{-\int_t^\tau r} E\Big[dB(\tau)\Big|Z(t) = j\Big] \\ =& \int_t^n e^{-\int_t^\tau r} E\Big[\sum_k I_k(t)dB_k(t) + \sum_{k \neq l} b_{kl}(t)dN_{kl}(t)\Big|Z(t) = j\Big] \\ =& \int_t^n e^{-\int_t^\tau r} \sum_k E\Big[I_k(t)\Big|Z(t) = j\Big]dB_k(t) + \sum_{k \neq l} b_{kl}(t)E\Big[dN_{kl}(t)\Big|Z(t) = j\Big]. \end{aligned} \tag{3.52}$$

And then filling in what we know from equations (3.50) and (3.51), we find

$$\begin{aligned} V_j(t) &= \int_t^n e^{-\int_t^\tau r} \sum_k p_{jk}(t, \tau)dB_k(\tau) + \sum_{l; l \neq k} b_{kl}(\tau)p_{jk}(t, \tau)\mu_{kl}(\tau)d\tau \\ &= \int_t^n e^{-\int_t^\tau r} \sum_k p_{jk}(t, \tau)\Big(dB_k(\tau) + \sum_{l; l \neq k} b_{kl}(\tau)\mu_{kl}(\tau)d\tau\Big). \end{aligned} \tag{3.53}$$

This can also be seen by using the same argumentation as with the intuitive derivation of the Kolmogorov differential equation, where we look at what can happen in every small interval of length $d\tau$ and then sum over the entire interval $(t, n]$ to get the integral in question.

Now using the Chapman-Kolmogorov equation we can divide the equation into a part concerning payments in $(t, u]$ and a part in $(u, n]$, with $t < u < n$. The equation then becomes

$$\begin{aligned} V_j(t) = &\int_t^u e^{-\int_t^\tau r} \sum_k p_{jk}(t, \tau)\Big(dB_k(\tau) + \sum_{l; l \neq k} b_{kl}(\tau)\mu_{kl}(\tau)d\tau\Big) \\ &+ e^{-\int_t^u r} \sum_k p_{jk}(t, u)V_k(u). \end{aligned} \tag{3.54}$$

Just like before we can obtain Thiele's differential equation just by taking the derivative with respect to t. We can also get the differential equation by conditioning on what happens in the small interval $(t, t + dt)$, again similar to the derivation of the Kolmogorov differential.

$$\begin{aligned} V_j(t) = b_j(t)dt + \sum_{k;k\neq j} \mu_{jk}(t)dt b_{jk}(t) \\ +(1 - \mu_{j\cdot}(t)dt)e^{-r(t)dt}V_j(t + dt) + \sum_{k;k\neq j} \mu_{jk}(t)dt e^{-r(t)dt}V_k(t + dt). \end{aligned} \tag{3.55}$$

Now we subtract $V_j(t + dt)$ on both sides, divide by dt and let $dt \to 0$. We arrive at Thiele's backward differential equation for the statewise prospective reserves:

$$\frac{d}{dt}V_j(t) = \big[r(t) + \mu_{j\cdot}(t)\big]V_j(t) - \sum_{k;k\neq j} \mu_{jk}(t)V_k(t) - b_j(t) - \sum_{k;k\neq j} \mu_{jk}(t)b_{jk}(t). \tag{3.56}$$

We must make a note about continuity now. We assumed the functions b_j, b_{jk}, μ_{jk} and r to be piecewise continuous, so there is no problem at points where all functions are continuous. However in practice there are only finitely many possible points of discontinuity, so Norberg argues that they are not a problem in numerical procedures solving the equation. The conditions for solving the equations are:

$$V_j(t_d-) = \big[B_j(t_d) - B_j(t_d-)\big] + V_j(t_d), \tag{3.57}$$

where $j \in \mathcal{Z}$ and $t_d \in \mathcal{D}$, the set discontinuity points of B_j.

From this we can conclude that Thiele's differential is a generalized version of Kolmogorov's differential equation, where the transition probability $p_{jk}(t, u)$ is the prospective reserve for a contract in state j at time t with a lump sum of 1 payment if the contract is in state k at time u.

Using the above differential equation we can decompose the premium payments in two parts. We will do the decomposition first and then we will interpret the results. First we isolate $-b_j(t)dt$ on the left hand side,

$$\begin{aligned} -b_j(t)dt =& dV_j(t) - \big[r(t) + \mu_{j\cdot}(t)\big]V_j(t)dt + \\ & \sum_{k;k\neq j} \mu_{jk}(t)V_k(t)dt + \sum_{k;k\neq j} \mu_{jk}(t)b_{jk}(t)dt. \end{aligned} \tag{3.58}$$

Once we remember that $\mu_{j\cdot}(t) = \sum_{k;k\neq j} \mu_{jk}(t)$ we can put all the summations under one summation sign, so then

$$-b_j(t)dt = dV_j(t) - r(t)dtV_j(t)dt + \sum_{k;k\neq j} R_{jk}(t)\mu_{jk}(t)dt. \tag{3.59}$$

Here $R_{jk}(t) = b_{jk}(t) + V_k(t) - V_j(t)$ denotes the so called sum at risk. This consists of the amount to be payed when transitioning from state j to k at time t and the difference between the prospective reserve given the transition to state k. So now the last term of the equation signifies the part of the premium payment that goes to covering off the risk of going to another state, whereas the first two terms can be seen as the payment needed to have the appropriate reserve for the current state.

To conclude this section we will take the time to motivate why we have gone through the trouble of constructing differential equations from the explicit expressions for the reserve. In the examples covered so far we never allowed the payment functions to be dependent on the reserves, however in practice such situations may apply. Think about a clause in the contract allowing repayment of part of the reserve upon prematurely ending the contract. In such cases we cannot use the direct equation and need the differential equations to solve the problem.

3.6 Higher Order Moments of the Reserves

Now that we have a better understanding of the reserves, we will look at a theorem on the moments of the present values.

The setup we use is that of a Markov model with the generalized contract introduced in Section 3.5. Let $\mathcal{D}$ again be the set of discontinuity points. We are now interested in the higher order moments of the reserve $V(t)$. Since we have assumed the model to be Markov, the state-wise conditional moments fully determine the moments of V(t). So we need to calculate

$$V_j^{(q)}(t) = E[V(t)^q | Z(t) = j]. \tag{3.60}$$

These functions $V_j^{(q)}$ are determined by the following differential equations

$$\begin{aligned} \frac{d}{dt} V_j^{(q)}(t) =& (qr(t) + \mu_{j\cdot}(t)) V_j^{(q)}(t) - q b_j(t) V_j^{(q-1)}(t) \\ & - \sum_{k;k \neq j} \mu_{jk}(t) \sum_{p=0}^{q} \binom{q}{p} (b_{jk}(t))^p V_k^{q-p}(t), \end{aligned} \tag{3.61}$$

on $(0,n) \backslash \mathcal{D}$ with the conditions

$$V_j^{(q)}(t-) = \sum_{p=0}^{q} \binom{q}{p} (B_j(t) - B_j(t-))^p V_j^{q-p}(t), \tag{3.62}$$

with $t \in \mathcal{D}$.

The proof for this is in the same vain as before. We condition on what happens in a small time interval and continue from there.

Before we start we introduce the notation $V(t,u)$ which corresponds to the present value at time t of the payment stream during (t,u). Also let $V(t) = V(t,n)$. Then if $t < u < n$ we get

$$V(t) = V(t,u) + e^{-\int_t^u r} V(u), \tag{3.63}$$

as we have to discount $V(u)$ to time t. Using this notation and binomial expansion we can now write

$$V(t)^q = \sum_{p=0}^{q} \binom{q}{p} V(t,u)^p \left(e^{-\int_t^u r} V(u) \right)^{q-p}. \tag{3.64}$$

The conditions already follow from this equation if we insert $t - dt$ and t as values for t and u and letting dt tend to zero.

For the differential equation, we are looking for the state-wise conditional moments. So let us take the expectation conditional on being in state j. We also put $u = t + dt$, so that we get

$$V_j^{(q)}(t) = \sum_{p=0}^{q} \binom{q}{p} E\Big[V(t, t+dt)^p \Big(e^{-r(t)dt} V(t+dt)\Big)^{q-p} \Big| Z(t) = j\Big]. \quad (3.65)$$

Now if we condition on either staying in state j or transitioning to some other state k, then we get

$$\begin{aligned} V_j^{(q)}(t) = \sum_{p=0}^{q} \binom{q}{p} (1 - \mu_{j\cdot}(t)dt)(b_j(t)dt)^p e^{-(q-p)r(t)dt} V_j^{(q-p)}(t+dt) + \\ \binom{q}{p} \sum_{k;k\neq j} \mu_{jk}(t)dt(b_j(t)dt + b_{jk}(t))^p e^{-(q-p)r(t)dt} V_k^{(q-p)}(t+dt). \quad (3.66) \end{aligned}$$

To get the differential later on, we will divide by dt and let it tend to zero. This means that we can ignore terms that are multiplied by dt^k with $k > 1$. Looking at the first line the term $(b_j(t)dt)^p$ means that all terms reduce to zero for $p \leq 2$. For $p = 0, 1$ we are left with

$$\begin{aligned} &(1 - \mu_{j\cdot}(t)dt)e^{-qr(t)dt} V_j^{(q)}(t+dt) + \\ &q(1 - \mu_{j\cdot}(t)dt)(b_j(t)dt)e^{-(q-1)r(t)dt} V_j^{(q-1)}(t+dt) \longrightarrow \\ &(1 - \mu_{j\cdot}(t)dt)e^{-qr(t)dt} V_j^{(q)}(t+dt) + qb_j(t)dte^{-(q-1)r(t)dt} V_j^{(q-1)}(t+dt), \quad (3.67) \end{aligned}$$

keeping only the terms not containing higher orders of dt.

The second line can be reduced using binomial expansion on $dt(b_j(t)dt + b_{jk}(t))^p$. It is immediately clear that we lose all terms with $b_j(t)dt$, so that the second line becomes:

$$\binom{q}{p} \sum_{k;k\neq j} \mu_{jk}(t)dt(b_{jk}(t)^p e^{-(q-p)r(t)dt} V_k^{(q-p)}(t+dt). \quad (3.68)$$

Now denote $V_j^{(q)}(t)^*$ as the terms of $V_j^{(q)}(t)$ with maximum order dt of one. Then combining the above we get

$$\begin{aligned} V_j^{(q)}(t)^* =& (1 - \mu_{j\cdot}(t)dt)e^{-qr(t)dt} V_j^{(q)}(t+dt) + qb_j(t)dte^{-(q-1)r(t)dt} V_j^{(q-1)}(t+dt) \\ &+ \sum_{p=0}^{q} \binom{q}{p} \sum_{k;k\neq j} \mu_{jk}(t)dt(b_{jk}(t))^p e^{-(q-p)r(t)dt} V_k^{(q-p)}(t+dt). \quad (3.69) \end{aligned}$$

Now we subtract $V_j^{(q)}(t+dt)$ from both sides. The first term on the right then can be written as

$$-\mu_{j\cdot}(t)dte^{-qr(t)dt} V_j^{(q)}(t+dt) + (e^{-qr(t)dt} - 1)V_j^{(q)}(t+dt) \quad (3.70)$$

When we divide by dt and let it go to zero, we get

$$-\mu_{j\cdot}(t)V_j^{(q)}(t) - qr(t)V_j^{(q)}(t). \quad (3.71)$$

To see why, we just write out the e-power

$$\frac{(e^{-qr(t)dt}-1)}{dt}=\frac{1}{dt}\Big(1-qr(t)dt+\frac{(qr(t)dt)^2}{2!}-\frac{(qr(t)dt)^3}{3!}\cdots-1\Big). \tag{3.72}$$

So it is clear that this equals $-qr(t)$, when letting dt tend to zero.

The rest of $V_j^{(q)}(t)^*$ becomes a little more readable as well when apply these steps to the other terms in the equation. We end up with

$$\begin{aligned}-\frac{d}{dt}V_j^{(q)}(t)=&-(\mu_{j\cdot}(t)+qr(t))V_j^{(q)}(t)+qb_j(t)V_j^{(q-1)}(t)\\&+\sum_{p=0}^{q}\binom{q}{p}\sum_{k;k\neq j}\mu_{jk}(t)(b_{jk}(t))^pV_k^{(q-p)}(t).\end{aligned} \tag{3.73}$$

Now multiply by -1 and change the order of the summation to get the desired result.

3.7 Modeling the Interest Rate

In the former part of this thesis we considered the interest rate $r(t)$ to be constant, or 'some' function of time t. In conclusion of this chapter we want to show that expanding the model by using a Markov chain for the interest rate as well is fairly straightforward, to show the flexibility of the model. We will assume that the interest rate can be modeled by a continuous time Markov chain Y on a finite state space $\mathcal{J}^Y$, with intensities λ_{ef}, where $e,f\in\mathcal{J}^Y$. Then we can write the interest rate as

$$r(t)=\sum_e I_e^Y(t)r_e, \tag{3.74}$$

where r_e denotes the force of interest in state e, and $I_e^Y(t)=\mathbf{1}[Y(t)=e]$ an indicator.

Suppose Y and Z are independent processes, then (Y,Z) is a Markov chain on $\mathcal{J}^Y\times\mathcal{J}^Z$, where $\mathcal{J}^Z$ is the state space of the process Z.

In equation (3.52) we first introduced the statewise reserves $V_j(t)$, conditional on the state the contract is in. Now we want to do the same, but conditional on the state of (Y,Z), so we take the interest rate into account as well. We denote $V_{ej}(t)$ as the statewise reserves conditional on $[Y(t)=e,Z(t)=j]$. Remember $V(t)=\int_t^n e^{-\int_t^\tau r(s)ds}dB(\tau)$, so that

$$V_{ej}(t)=E\Big[\int_t^n e^{-\int_t^\tau r(s)ds}dB(\tau)\Big|Y(t)=e,Z(t)=j\Big] \tag{3.75}$$

Similar to equation (3.50) we have:

$$E[I_f^Y(t)\big|Y(t)=e,Z(t)=j]=E[I_f^Y(t)\big|Y(t)=e]=\lambda_{ef}, \tag{3.76}$$

by the independence of Y and Z. Using this and the equation for the interest rate, we can write the conditional reserves as

$$
\begin{aligned}
V_{ej}(t) &= E\Big[\int_t^n e^{-\int_t^\tau r(s)ds} dB(\tau)\Big| Y(t)=e, Z(t)=j\Big] \\
&= E\Big[\int_t^n e^{-\int_t^\tau \sum_f I_f^Y(s) r_f ds} dB(\tau)\Big| Y(t)=e, Z(t)=j\Big] \\
&= \int_t^n e^{-\int_t^\tau \sum_f \lambda_{ef}(s) r_f ds} E\Big[dB(\tau)\Big| Y(t)=e, Z(t)=j\Big]. \qquad (3.77)
\end{aligned}
$$

By the independence of Y and Z we have $E[dB(\tau)|Y(t) = e, Z(t) = j] = E[dB(\tau)|Z(t) = j]$, so now we can proceed in exactly the same way as in Section 3.5, to find

$$
\begin{aligned}
\frac{d}{dt}V_{ej}(t) = \big[r_e + \mu_{j\cdot}(t) + \lambda_{e\cdot}\big]V_{ej}(t) &- \sum_{k;k\neq j}\mu_{jk}(t)V_{ek}(t) \\
- b_j(t) - \sum_{k;k\neq j}\mu_{jk}(t)b_{jk}(t) &- \sum_{f;f\neq e}\lambda_{ef}(t)V_{fj}(t). \qquad (3.78)
\end{aligned}
$$

And following the same steps as in Section 3.6 we would also find the differential equations for higher moments of the reserves to be determined by

$$
\begin{aligned}
\frac{d}{dt}V_{ej}^{(q)}(t) &= \big[qr_e + \mu_{j\cdot}(t) + \lambda_{e\cdot}\big]V_{ej}^{(q)}(t) - qb_j(t)V_{ej}^{(q-1)}(t) \\
-\sum_{k;k\neq j}\mu_{jk}(t)&\sum_{p=0}^{q}\binom{q}{p}(b_{jk}(t))^p V_{ek}^{q-p}(t) - \sum_{f;f\neq e}\lambda_{ef}(t)V_{fj}^{(q)}(t), \qquad (3.79)
\end{aligned}
$$

on $(0, n)\backslash\mathcal{D}$ with the conditions

$$
V_{ej}^{(q)}(t-) = \sum_{p=0}^{q}\binom{q}{p}(B_j(t) - B_j(t-))^p V_{ej}^{q-p}(t), \qquad (3.80)
$$

with $t \in \mathcal{D}$.

It is clear that allowing the interest rate to be determined by a Markov chain does not complicate the calculations much. It does however make the calculations more realistic. That is what we did in this chapter as well. We started with some basic forms of insurance, defining the payment stream and introducing Markov chains. Gradually we allowed the model to become more general. Now that we have a nice model, a natural question to ask would be how to determine the parameters needed to do the actual numerical calculations. This question we try to answer in the next chapter.

Chapter 4

Parameter Estimation

In this chapter we try to find estimates for the intesities of transition. We start with a quick refresher on Maximum Likelihood. In the sections after that we make use of Chapter 11 of [5], to show some results for the distribution of the estimates of μ.

4.1 Refresher on Maximum Likelihood

The method of maximum likelihood is assumed to be known among the readers, so we will not discuss the method in great detail. Instead we present a short summary on the theory and the results we use, making use of [6].

Let $X_1, ..., X_n$ be continuous random variables with joint density function $f(x_1, ..., x_n|\theta)$. The likelihood function is then defined as a function of θ, given some observed values $X_i = x_i$

$$\text{lik}(\theta) = f(x_1, ..., x_n|\theta). \tag{4.1}$$

If we assume the X_i's to be i.i.d. (independent and identically distributed), then we may write the joint density as the product of the marginal densities. So the likelihood function becomes

$$\text{lik}(\theta) = \prod_{i=1}^{n} f(X_i|\theta). \tag{4.2}$$

In practice it is usually easier to maximize the natural logarithm of the likelihood function. As the logarithm is a monotonic function, this transformation gives the same maximum. Taking the log of the likelihood function, we get

$$\ln\big[\text{lik}(\theta)\big] = \ln\Big[\prod_{i=1}^{n} f(X_i|\theta)\Big] = \sum_{i=1}^{n} \ln[f(X_i|\theta)]. \tag{4.3}$$

Often we use $l(\theta)$ instead of writing $\ln\big[\text{lik}(\theta)\big]$. While the latter notation shows exactly what happens, the shorter version is often more convenient.

With the method of maximum likelihood our objective is to maximize the likelihood function. Put differently, we search for a value of θ that makes the observations the most likely or probable. To find this maximum, we calculate

$$\frac{d}{d\theta} l(\theta)\Big|_{\hat{\theta}} = 0. \tag{4.4}$$

To see if the value is a maximum, we need to check if the second derivative is negative as well. If this is the case, we denote the value we find by $\hat{\theta}_{ML}$. The hat displays that this concerns an estimate for the true (unobserved) value θ.

Properties of the ML-estimator
In [8], Verbeek gives a useful list of properties of the estimator $\hat{\theta}_{ML}$. Firstly he notes that the maximum likelihood estimator is consistent, since it can be shown that the estimator converges in probability to the true value θ, so

$$\text{plim } \hat{\theta} = \theta. \tag{4.5}$$

The next property of $\hat{\theta}_{ML}$ is that it's asymptotically efficient, e.g. it has the smallest variance of all consistent estimators $\hat{\theta}$ of θ, so

$$Var(\hat{\theta}_{ML}) \leq Var(\hat{\theta}). \tag{4.6}$$

The estimator $\hat{\theta}_{ML}$ is also asymptotically normally distributed, such that

$$\sqrt{n}(\hat{\theta}_{ML} - \theta) \rightarrow \mathcal{N}(0, V), \tag{4.7}$$

where V is the asymptotic covariance matrix. This matrix V is defined by its inverse, the Fisher Information matrix

$$V^{-1} = -E\Big[\frac{\partial^2}{\partial\theta\partial\theta^T} l(\theta)\Big]. \tag{4.8}$$

Asymptotic theorems
In our analysis we will need two well known theorems concerning asymptotic results. The first is the Law of Large Numbers as stated in [6]. Let X_i, $i = 1, 2, ...$ be independently distributed random variables., with $E[X_i] = r$. Furthermore we write the sample average as $\bar{X}_n = \frac{1}{n}\sum_{i=i}^{n} X_i$. The Weak Law of Large Numbers then states that for any $\epsilon > 0$ we have

$$\lim_{n\to\infty} P(|\bar{X}_n - r|) > \epsilon \rightarrow 0. \tag{4.9}$$

Based on the same assumptions, but harder to prove, is the Strong Law of Large Numbers, which states:

$$P(\lim_{n\to\infty} \bar{X}_n = r) = 1. \tag{4.10}$$

We say that X_n almost surely converges to r.

Lastly we state the Central Limit Theorem, as given in [6]. We now consider a sequence of random variables Y_i, with $E[Y_i] = 0$ and $Var[Y_i] = \sigma^2$. We assume that the random variables Y_i have a common distribution F and that the moment generating function is defined in a neighborhood of zero. Now let $S_n = \sum_{i=1}^{n} Y_i$. then the CLT implies

$$\lim_{n\to\infty} P(\frac{S_n}{\sigma\sqrt{n}} \leq x) = \Phi(x). \tag{4.11}$$

Loosely put this theorem says that if we take the sum of independently distributed, zero mean random variables, their sum tends toward the normal distribution.

4.2 The Density and Distribution with Censoring

In the ideal situation we fully observe the length of a life, however this is not always the case. We will consider some different cases of censoring, but before we do that, we look at what happens to the denisty and distribution functions when using censored data.

The setup is the same as before, with a non-negative random variable T denoting the life length of an individual. the cumulative distribution function is given by

$$F(t) = 1 - e^{-\int_0^t \mu(s)ds}, \tag{4.12}$$

with $\mu(t)$ (piecewise) continuous, so that we may write the probability density function as

$$f(t) = \mu(t)(1 - F(t)). \tag{4.13}$$

During the life of an individual we do not know what his or her life length will be. If we observe he is alive at time t, while born at time 0, we only know his life length is at least t. These kinds of observations are called *right-censored*. Denote this truncated life length by J and let the observations be made continually at z years from birth. The distribution of J is

$$P[J \leq t] = \begin{cases} F(t), & 0 < t < z, \\ 1, & t \leq z. \end{cases} \tag{4.14}$$

The probability density function for $0 < t < z$ is just $f(t)$. However when $t = z$, the density function is different from the case without censoring. All probability that was in the part $t \geq z$ now falls into the single point $t = z$. We have

$$\int_0^z f(t)dt = \int_0^z \mu(t)e^{-\int_0^t \mu(s)ds} = 1 - e^{-\int_0^z \mu(s)ds} = F(z). \tag{4.15}$$

So for the density function $g(t)$ we now conclude that the case $t = z$ should have probability $1 - F(z)$. So for g(t) we write

$$g(t) = \begin{cases} \mu(t)(1 - F(t)), & 0 < t < z, \\ 1 - F(z), & t = z. \end{cases} \tag{4.16}$$

If we want to write this in a single line, we take

$$d(t) = \mathbf{1}_{(0,z)}(t) = \begin{cases} 1, & 0 < t < z, \\ 0, & t \geq z, \end{cases} \tag{4.17}$$

so that the density function $g(t)$ now may be written as

$$g(t) = \mu(t)^{d(t)}(1 - F(t)), 0 < t \leq z. \tag{4.18}$$

Suppose now that $\mu(t) = \mu$ is constant. In that case the distribution function reduces to

$$F(t) = 1 - e^{-\mu t}. \tag{4.19}$$

Remembering that $E[X] = \int [1 - F_X(x)]dx$, we can easily calculate the expected value of an exponentially distributed life length U

$$E[U] = \int_0^\infty e^{-\mu t} dt = \frac{e^{-\mu t}}{\mu}\Big|_{t=0}^{\infty} = \frac{1}{\mu}. \tag{4.20}$$

Now let U_i, $i = 1, 2, ...$ be independent and identically distributed according to (4.19). We wish to estimate μ using censored life lengths $T_i = U_i \wedge z_i$. Let N_i be the indicator function of dying before time z_i, then

$$N_i = \mathbf{1}[T_i < z_i] \tag{4.21}$$

And let $N = \sum_i N_i$ be the total number of deaths occurred in the sample. Furthermore we introduce $W = \sum_i T_i$, the total time spent alive in the sample, up to the observations z_i. We will write the likelihood function as the product of censored densities $g(t)$ introduced in the previous section. Then we get

$$\text{lik}(\mu) = \prod_i \mu^{N_i} e^{-\mu T_i} = \mu^N e^{-\mu W}. \tag{4.22}$$

Taking the log of likelihood function then results in

$$l(\mu) = \ln\left[\mu^N e^{-\mu W}\right] = N \ln \mu - \mu W. \tag{4.23}$$

We need to take the first derivative of $l(\mu)$ and set it equal to 0 to obtain

$$\frac{\partial l(\mu)}{\partial \mu} = \frac{N}{\mu} - W = 0 \implies \hat{\mu}_{ML} = \frac{N}{W}. \tag{4.24}$$

To verify that this is indeed a maximum, we take the second derivative

$$\frac{\partial^2 l(\mu)}{\partial \mu^2} = -\frac{N}{\mu^2}, \tag{4.25}$$

which is clearly non-positive. We conclude that the maximum likelihood estimator is equal to the so-called occurrence-exposure rate (OE-rate). Norberg states that this estimator does not depend on the censoring scheme. The distribution of $\hat{\mu}_{ML}$ however does. That is why we will consider some different cases of censoring, ranging from no censoring, to uniform censoring and lastly random censoring.

4.3 The Distribution of $\hat{\mu}_{ML}$ Under Different Censoring Schemes

No Censoring

We observe n lives completely, so translated to the situation in the previous section this boils down to $z_i = \infty$. Then $T_i = U_i$, $i = 1, ..., n$ and $N = \sum_{i=1}^n N_i = n$. The likelihood function looks pretty similar

$$\text{lik}(\mu) = \mu^n e^{-\mu W}. \tag{4.26}$$

Equivalent to the previous section we get

$$\hat{\mu}_{ML} = \frac{n}{W}. \tag{4.27}$$

Now W is the sum of n i.i.d. exponential random variables. We shall demonstrate that W is Gamma distributed with shape parameter n and scale parameter μ. Then W would have density

$$f_W(w) = \frac{\mu^n w^{n-1} e^{-\mu w}}{\Gamma(n)}. \tag{4.28}$$

Let X and Y be i.i.d. exponentially distributed with parameter λ and let $X + Y = Z$. Then the density function of Z, $f_Z(z)$ is given by calculating the convolution of the densities of X and Y, so

$$f_Z(z) = \int_0^z \lambda e^{-\lambda t} \lambda e^{-\lambda(z-t)} dt = \lambda^2 \int_0^z e^{-\lambda z} = \lambda^2 z e^{-\lambda z}. \tag{4.29}$$

Then Z is Gamma(2, λ) distributed, noting that $\Gamma(n) = (n-1)!$, so $\Gamma(2) = 1$.

Looking at the density function of W it is easy to see that the exponential distribution is just a special case of the Gamma distribution, with scale parameter $n = 1$. Taking this into account, we can now show by induction that if $X \sim \text{Gamma}(n, \lambda)$ and $Y \sim \text{Gamma}(1, \lambda)$ then $X + Y = Z \sim \text{Gamma}(n+1, \lambda)$. We proceed in the same way, calculating the convolution

$$\begin{aligned} f_Z(z) &= \int_0^z \frac{\lambda^n t^{n-1} e^{-\lambda t}}{\Gamma(n)} \lambda e^{-\lambda(z-t)} dt = \lambda^{n+1} \int_0^z \frac{t^{n-1} e^{-\lambda t}}{\Gamma(n)} e^{-\lambda(z-t)} dt \\ &= \frac{\lambda^{n+1}}{\Gamma(n)} e^{-\lambda z} \int_0^z t^{n-1} dt = \frac{\lambda^{n+1}}{\Gamma(n)} e^{-\lambda z} \frac{t^n}{n} \Big|_{t=0}^{t=z} \\ &= \frac{\lambda^{n+1} t^n e^{-\lambda z}}{\Gamma(n+1)}. \end{aligned} \tag{4.30}$$

Proving that $Z \sim \text{Gamma}(n + 1, \lambda)$. Now given that W is the sum of n i.i.d. exponential(μ) random variables, we may conclude that $W \sim \text{Gamma}(n, \mu)$.

Having established the distribution of W, we can now move on the calculating the mean and variance of $\hat{\mu}_{ML}$. First we calculate the expected value of W^k, $k > -n$, so k may also be negative

$$E[W^k] = \int_0^\infty w^k f_W(w) dw = \int_0^\infty \frac{\mu^n w^{n+k-1} e^{-\mu w}}{\Gamma(n)} dw. \tag{4.31}$$

This could be solved by applying partial integration, however there is a quicker way. We need to realize that $\int f_X(x)dx = 1$ for any density function of a random variable X. So we try to get a probability density function under the integral sign

$$\begin{aligned} \int_0^\infty \frac{\mu^n w^{n+k-1} e^{-\mu w}}{\Gamma(n)} dw &= \frac{1}{\mu^k} \int_0^\infty \frac{\mu^{n+k} w^{n+k-1} e^{-\mu w}}{\Gamma(n)} dw \\ \frac{\Gamma(n+k)}{\Gamma(n)\mu^k} \int_0^\infty \frac{\mu^{n+k} w^{n+k-1} e^{-\mu w}}{\Gamma(n+k)} dw &= \frac{\Gamma(n+k)}{\Gamma(n)\mu^k}. \end{aligned} \tag{4.32}$$

Using this, we find the expected value of $\hat{\mu}_{ML}$. Taking $k = -1$ in $E[W^k]$ we get

$$E[\hat{\mu}_{ML}] = nE[W^{-1}] = \frac{n\mu}{n-1}, \tag{4.33}$$

We conclude the estimator is biased.

The bias of the estimator is given by $E[\hat{\mu}_{ML} - \mu]$, so in our case the bias is

$$E[\hat{\mu}_{ML} - \mu] = \frac{n\mu}{n-1} - \mu = \frac{-\mu}{n-1}, \tag{4.34}$$

which goes to zero, as n grows to infinity. So indeed our estimator is consistent.

The variance of the estimator is given by

$$Var[\hat{\mu}_{ML}] = E[\hat{\mu}_{ML}^2] - E[\hat{\mu}_{ML}]^2 = n^2E[W^{-2}] - \Big(\frac{n\mu}{n-1}\Big)^2, \tag{4.35}$$

which by equation(4.32) can be written as

$$n^2E[W^{-2}] - \Big(\frac{n\mu}{n-1}\Big)^2 = n^2\frac{\mu^2}{(n-1)(n-2)} - \Big(\frac{n\mu}{n-1}\Big)^2. \tag{4.36}$$

A bit of manipulation then gives

$$Var[\hat{\mu}_{ML}] = \frac{n^2\mu^2}{(n-1)^2(n-2)}. \tag{4.37}$$

We have been able to obtain these results without using any asymptotic results. This makes the above results very strong *if* the assumptions hold. Sadly, the special case of no censoring is not very realistic. Let us find out what happens when we use uniform censoring.

Uniform Censoring

In this scenario we let the moment of observation be equal for all observations in the sample, so $z_i = z$.

Since the censoring only has an impact on the distribution of $\hat{\mu}_{ML}$, but not on the value of $\hat{\mu}_{ML}$, we can still write

$$\hat{\mu}_{ML} = \frac{N}{W} = \frac{\sum_{i=1}^n N_i}{\sum_{i=1}^n T_i}. \tag{4.38}$$

We want to check if the estimator is consistent. To do so, we need to know $E[N_i]$ and $E[T_i]$. Using equation (4.14) we have

$$E[T_i] = \int_0^\infty 1 - P[T_i \leq t]dt = \int_0^z 1 - F(t)dt = \int_0^z e^{-\mu t}dt = \frac{1-e^{-\mu z}}{\mu}. \tag{4.39}$$

The calculation for $E[N_i]$ is even more basic

$$E[N_i] = E[\mathbf{1}[T_i < z]] = F(z) = 1 - e^{-\mu z}. \tag{4.40}$$

We did this exercise to verify that $E[N_i] = \mu E[T_i]$.

Now let us write

$$\hat{\mu}_{ML} = \frac{\frac{1}{n}\sum_{i=1}^{n} N_i}{\frac{1}{n}\sum_{i=1}^{n} T_i}. \tag{4.41}$$

Then by the Strong Law of Large Numbers, $\hat{\mu}_{ML}$ almost surely converges to μ

$$\hat{\mu}_{ML} \to \frac{\mu E[T_i]}{E[T_i]} = \mu \text{ , as } n \to \infty. \tag{4.42}$$

Using the CLT we would also like to say something about the asymptotic distribution of $\hat{\mu}_{ML}$. To do so, we look at the random variable $Y_i = N_i - \mu T_i$. Then $E[Y_i] = 0$ and $Var[Y_i] = E[Y_i^2] = 1 - e^{-\mu z}$. To see why, we have to calculate a few intermediate steps. We start by writing out $E[Y_i^2]$

$$E[Y_i^2] = E\big[(N_i - \mu T_i)^2\big] = E[N_i^2] - 2\mu E[T_i N_i] + \mu^2 E[T_i^2]. \tag{4.43}$$

The first term is an indicator function, so $E[N_i^p] = E[N_i]$ for any $p \geq 1$. Thus $E[N_i^2] = F(z)$.

The last term can be found using the rules for taking expectation of a function of a random variable

$$E[T_i^2] = 2\int_0^z t(1 - F(t))dt = 2\int_0^z te^{-\mu t}dt. \tag{4.44}$$

We find the integral using integration by parts

$$2\int_0^z te^{-\mu t}dt = 2\frac{te^{-\mu t}}{\mu}\Big|_{t=0}^{z} + 2\int_0^z \frac{e^{-\mu t}}{\mu}dt = \frac{2 - 2\mu z e^{-\mu z} - 2e^{-\mu z}}{\mu^2}. \tag{4.45}$$

So then the third term of the equation for the variance becomes

$$\mu^2 E[T_i^2] = 2 - 2\mu z e^{-\mu z} - 2e^{-\mu z} = 2F(z) - 2\mu z e^{-\mu z}. \tag{4.46}$$

Now we only need to find an expression for $2\mu E[T_i N_i]$. Let $I(t) = \mathbf{1}[T_i > t]$ denote the indicator function of survival up to time t, then we may write $N_i = 1 - I(z)$ and $T_i = \int_0^z I(t)dt$. So then we obtain the expression

$$T_i N_i = \int_0^z I(t)dt - \int_0^z I(t)I(z)dt = T_i - \int_0^z I(t)I(z)dt. \tag{4.47}$$

Now we note that $I(t)I(z) = 1$ if and only if $T_i > t$ and $T_i > z$, and 0 otherwise. This realization allows us to write

$$T_i N_i = T_i - zI(z). \tag{4.48}$$

Now we take the expectation and multiply by 2μ to find

$$2\mu E[T_i N_i] = 2\mu E[T_i] - 2\mu z(1 - F(z)) = 2F(z) - 2\mu z e^{-\mu z}. \tag{4.49}$$

Plugging this in the equation for the variance of Y_i we finally arrive at

$$E[Y_i^2] = F(z) = 1 - e^{-\mu z}. \tag{4.50}$$

Now to apply the Central Limit Theorem, we need to rewrite equation (4.41). We subtract μ and multiply by $\sqrt{n}$

$$\sqrt{n}(\hat{\mu}_{ML} - \mu) = \sqrt{n}\Big(\frac{\frac{1}{n}\sum_{i=1}^{n} N_i}{\frac{1}{n}\sum_{i=1}^{n} T_i} - \mu\Big) = \frac{\sum_{i=1}^{n}\big[N_i - \mu T_i\big]}{\sqrt{n}\frac{1}{n}\sum_{i=1}^{n} T_i}. \tag{4.51}$$

So we are now in the right form to use the Central Limit Theorem for equation (4.11), because the $\frac{1}{n}\sum_{i=1}^{n} T_i$ in the denominator converges to $E[T_i]$, the variance of $Y_i = N_i - \mu T_i$. Given that asymptotically $\sqrt{n}(\hat{\mu}_{ML} - \mu) \sim \mathcal{N}(0, 1)$, it is easy to see that asymptotically

$$\hat{\mu}_{ML} \sim \mathcal{N}(\mu, \frac{\mu^2}{n(1 - e^{-\mu z})}. \tag{4.52}$$

Random Censoring

Until now, we have assumed that all observations were made at the same time z. In the context of insurance, this is a rare situation. It would be more appropriate to consider the case of random sampling, where the times of the observation are given by i.i.d. random variables. In [5] Norberg proves some results for the case of general censoring, where the time of observation z_i is different among all individuals (but not random). He proves that $\sum_{i=1}^{n} E[T_i] \to \infty$ is a sufficient condition for $\hat{\mu}_{ML}$ to be consistent and asymptotically normal distributed. This condition can be explained as that the censoring must not be too severe, as the expected number of deaths needs to go to infinity as well.

Now let us assume the above condition, and that the censoring times are given by independent observations Z_i of some distribution function H, with density h independent of μ. Then instead of working with i.i.d. pairs of observations (N_i, T_i), we now have the i.i.d. triplet (N_i, T_i, Z_i), making the setup almost the same as before. The likelihood function in this case is given by

$$\text{lik}(\mu) = \prod_i \mu^{N_i} e^{-\mu T_i} h(Z_i). \tag{4.53}$$

Note that we are allowed to just multiply by $h(Z_i)$, because of the assumed independence. Because $h(Z_i)$ is independent of μ we also have that the estimate in equation (4.24) is still valid. The difference is found in analyzing the distribution of $\hat{\mu}_{ML}$.

In the previous part of this section we made use of the expression $E[N_i] = 1 - e^{-\mu z}$ for a few calculations. This is the only change we need to make in our analysis. So we need to write

$$E[N_i] = 1 - E[e^{-\mu Z_i}] \tag{4.54}$$

$$E[T_i] = \frac{1 - E[e^{-\mu Z_i}]}{\mu}. \tag{4.55}$$

Since the triplets (N_i, T_i, Z_i) are i.i.d., we can still use the CLT and LLN to obtain the asymptotic distribution of $\hat{\mu}_{ML}$. Then in a similar way as before, we get

$$\hat{\mu}_{ML} \sim \mathcal{N}(\mu, \frac{\mu^2}{n(1 - E[e^{-\mu Z_i}])}. \tag{4.56}$$

So clearly only the distribution of the ML-estimate depends on the censoring scheme. If we assume $\sum_{i=1}^{n} E[T_i] \to \infty$, the estimator is always consistent.

4.4 Estimating Intensities in the Markov Model

Of course it would not only be interesting to estimate a mortality law, equivalent to the two state Markov model introduced in section3.4. We want to estimate a more general model as well. To do so, we recall the setup of Section 3.5, where we introduced a more general Markov model for insurance policies. Consider a Markov process Z, with state space $\mathcal{Z} = 0, 1, ..., J$. Furthermore remember $I_g(t)$ and $N_{gh}(t)$, an indicator function and counting process, defined as in Section 3.5.

We observe the state of the insurance policy constantly during some time interval $[t_0, t_{end}]$. Let g_i, $i = 0, 1, ..., n-1$ denote the subsequent states the process visits. Then we can write a realization of the process as:

$$X(\tau) = \begin{cases} g_0, & t_0 \leq \tau < t_1, \\ g_1, & t_1 + dt_1 \leq \tau < t_2, \\ \vdots & \vdots \\ g_{n-2}, & t_{n-2} + dt_{n-2} \leq \tau < t_{n-1}, \\ g_{n-1}, & t_{n-1} + dt_{n-1} \leq \tau \leq t_{end}. \end{cases} \tag{4.57}$$

We will try to get the estimates for μ_{gh} using Maximum Likelihood, so we need to find some expression for the probability of $X(\tau)$. Remember that the transition probability of going from state j to state j is equal to $p_{jj}(t, u) = e^{-\int_t^u \mu_{j\cdot}}$. Furthermore, recall the definition for the intensities of transition given in Section 2.3, where we wrote $p_{jk}(t, t + dt) = \mu_{jk}(t)dt + o(dt)$. Then we can write to probability of $X(\tau)$ happening as

$$\begin{aligned} &e^{-\int_{t_0}^{t_1} \mu_{g_0\cdot}} \mu_{g_0 g_1}(t_1)dt_1 e^{-\int_{t_1}^{t_2} \mu_{g_1\cdot}} \mu_{g_1 g_2}(t_2)dt_2 \cdots \\ \cdots &e^{-\int_{t_{n-2}}^{t_{n-1}} \mu_{g_{n-2}\cdot}} \mu_{g_{n-2} g_{n-1}}(t_{n-1})dt_{n-1} e^{-\int_{t_{n-2}}^{t_{n-1}} \mu_{g_{n-1}\cdot}}, \end{aligned} \tag{4.58}$$

where we ignored the $o(dt_i)$ terms, for we will be dividing by the product of dt_i's and let them tend to zero. However, first we do some more rewriting

$$\begin{aligned} &\prod_{k=1}^{n-1} \mu_{g_{k-1}g_k}(t_k)dt_k \exp\Big(-\sum_1^n \int_{t_{k-1}}^{t_k} \mu_{g_{k-1}\cdot}\Big) \\ = &\exp\Big(\sum_{k=1}^{n-1} \ln\big[\mu_{g_{k-1}g_k}(t_k)\big] - \sum_1^n \int_{t_{k-1}}^{t_k} \mu_{g_{k-1}\cdot}\Big) dt_1 \cdots dt_{n-1}. \end{aligned} \tag{4.59}$$

Now divide by $\prod_i dt_i$ and let the dt_i's go to zero. Then from this we can construct a general expression for the likelihood of a observation. We need to sum over all possible states, instead of over some given sequence $g_0, g_1, ..., g_{n-1}$. Integrating with respect to a counting function is the same as just summing the values according to [7], so

$$\sum_{k=1}^{n-1} \ln \mu_{g_{k-1}g_k}(t_k) = \int_{t_0}^{t_{end}} \ln \mu_{g_{k-1}g_k}(\tau) dN_{g_{k-1}g_k}(\tau). \tag{4.60}$$

The second part can be rewritten using an indicator function

$$\sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} \mu_{g_{k-1}\cdot} = \sum_{k=1}^{n} \int_{t_0}^{t_{end}} \mu_{g_{k-1}\cdot}(\tau) I_{g_{k-1}}(\tau) d\tau. \tag{4.61}$$

Using the above two equations and summing over all possible states g, we get the following likelihood function

$$\text{lik} \; = \exp\Big(\sum_{g \neq h} \int_{t_0}^{t_{end}} \ln\big[\mu_{gh}(\tau) dN_{gh}(\tau)\big] - \sum_{g} \int_{t_0}^{t_{end}} \mu_{g\cdot}(\tau) I_g(\tau) d\tau\Big). \quad (4.62)$$

The integrals and summations can be written as one and recall $\mu_{g\cdot} = \sum_{g \neq h} \mu_{gh}$, to get

$$\text{lik} \; = \exp\Big(\sum_{g \neq h} \int_{t_0}^{t_{end}} \ln \mu_{gh}(\tau) dN_{gh}(\tau) - \mu_{gh}(\tau) I_g(\tau) d\tau\Big). \quad (4.63)$$

Notice that we have not written an argument for the likelihood function yet. To do so, we will assume that the intensities are of a parametric form and twice differentiable, and can be written as $\mu_{gh}(t, \theta)$, with $\theta = (\theta_1, ..., \theta_s)^T$, with $s \geq 1$ an integer. Suppose we have n observations of the same insurance policy, or in other words, replicates of the process Z, which we assume to be independent. We will not assume equal censoring, so let $X^{(m)}$ denote the policy of the m-th insurance taker. Furthermore assume $I_g^{(m)}(t)$ and $dN_{gh}^{(m)}(t)$ are 0, for t outside of $[t_0^{(m)}, t_{end}^{(m)}]$. The log-likelihood function is then given by

$$l(\theta) = \sum_{g \neq h} \int \ln \mu_{gh}(\tau, \theta) dN_{gh}(\tau) - \mu_{gh}(\tau, \theta) I_g(\tau) d\tau, \quad (4.64)$$

where $N_{gh} = \sum_{m=1}^{n} N_{gh}^{(m)}$ and $I_g = \sum_{m=1}^{n} I_g^{(m)}$.

Then taking the derivative with respect to θ we get

$$\frac{\partial}{\partial \theta} l(\theta) = \sum_{g \neq h} \int \frac{\partial}{\partial \theta} \Big[\ln \mu_{gh}(\tau, \theta) dN_{gh}(\tau) - \mu_{gh}(\tau, \theta) I_g(\tau) d\tau\Big]. \quad (4.65)$$

Setting this equal to zero yields the equations from which to solve the estimate $\hat{\theta}_{ML}$. A more practical way to write this is by introducing transition times $T_{gh}^{(i)}$. This is the time of the i-th transition, $i = 0, 1, ..., N_{gh}$ from state g to h. Then we can write the ML-equations as

$$\sum_{g \neq h} \sum_{j=1}^{N_{gh}} \frac{\frac{\partial}{\partial \theta_i} \mu_{gh}(T_{gh}^{(j)}, \hat{\theta}_{ML})}{\mu_{gh}(T_{gh}^{(j)}, \hat{\theta}_{ML})} = \sum_{g \neq h} \int \frac{\partial}{\partial \theta_i} \mu_{gh}(\tau, \hat{\theta}_{ML}) I_g(\tau) d\tau, \quad (4.66)$$

for $i = 1, ..., s$

The variance is given by the inverse of the Fisher Information matrix. We need the matrix of second order derivatives of the likelihood function

$$\frac{\partial^2}{\partial \theta \partial \theta^T} l(\theta) = \sum_{g \neq h} \int \frac{\partial^2}{\partial \theta \partial \theta^T} \Big[\ln \mu_{gh}(\tau, \theta) dN_{gh}(\tau) - \mu_{gh}(\tau) I_g(\tau) d\tau\Big]. \quad (4.67)$$

When taking the expectation, it will be convenient to rewrite this using

$$\frac{\partial^2}{\partial \theta \partial \theta^T} \ln \mu_{gh}(\tau, \theta) = \frac{\frac{\partial^2}{\partial \theta \partial \theta^T} \mu_{gh}(\tau, \theta)}{\mu_{gh}(\tau, \theta)} - \frac{\frac{\partial^2}{\partial \theta \partial \theta^T} \mu_{gh}(\tau, \theta)}{(\mu_{gh}(\tau, \theta))^2}. \quad (4.68)$$

Next we note that

$$\frac{\partial}{\partial\theta}\ln\mu_{gh}(\tau,\theta)\frac{\partial}{\partial\theta^T}\mu_{gh}(\tau,\theta)I_g(\tau)=\frac{\frac{\partial}{\partial\theta}\mu_{gh}(\tau,\theta)}{\mu_{gh}(\tau,\theta)}\frac{\partial}{\partial\theta^T}\mu_{gh}(\tau,\theta)I_g(\tau). \tag{4.69}$$

Then the matrix of derivatives is alternatively given by

$$\begin{aligned}\frac{\partial^2}{\partial\theta\partial\theta^T}l(\theta)=&\sum_{g\neq h}\int\frac{\partial^2}{\partial\theta\partial\theta^T}\ln\mu_{gh}(\tau,\theta)\Big[dN_{gh}(\tau)-\mu_{gh}(\tau)I_g(\tau)d\tau\Big]\\&-\frac{\partial}{\partial\theta}\ln\mu_{gh}(\tau,\theta)\frac{\partial}{\partial\theta^T}\mu_{gh}(\tau,\theta)I_g(\tau).\end{aligned} \tag{4.70}$$

Recalling equations (3.50) and (3.51), it is clear that

$$E[dN_{gh}(\tau)-\mu_{gh}(\tau)I_g(\tau)d\tau]=0. \tag{4.71}$$

Let the probability of staying in state g at time t for the censored process $Z^{(m)}$ be $p_g^{(m)}(t)$, so that

$$E[I_g(t)]=\sum_{m=1}^{n}p_g^{(m)}(t). \tag{4.72}$$

Then the Fisher Information matrix, V^{-1} is given by

$$V^{-1}=-E\Big[\frac{\partial^2}{\partial\theta\partial\theta^T}l(\theta)\Big]=\sum_{g\neq h}\int\frac{1}{\mu_{gh}(\tau,\theta)}\Big[\frac{\partial^2}{\partial\theta\partial\theta^T}\mu_{gh}(\tau,\theta)\Big]\sum_{m=1}^{n}p_g^{(m)}(t) \tag{4.73}$$

Given the above, we can calculate estimates for the coefficients in $\mu_{gh}(t,\theta)$ for a general Markov model and we also know the distribution of these estimates. They are given by

$$\hat{\theta}\sim\mathcal{N}(\theta,V). \tag{4.74}$$

4.5 Concluding Example

In conclusion of our analysis we would like to show an example of theory described above.

Consider a sample of n individuals. For each $m\in\{0,1,...,n\}$ we have x_m, the age on entering the study, and y_m, the age when leaving the study. We also have N_m, indicating whether the individual has died (1) or was still living during the study (0). Norberg proceeds assuming the mortality law is the first Gompertz-Makeham mortality law, so that the intensity at age t is given by:

$$\mu(t,\theta)=\alpha+\beta e^{\gamma t}, \tag{4.75}$$

with $\theta=(\alpha,\beta,\gamma)^T$. However, we will look at the second Gompertz-Makeham mortality law as stated in [4]

$$\mu(t,\theta)=\alpha+\beta t+e^{\gamma t+\delta}, \tag{4.76}$$

with $\theta = (\alpha, \beta, \gamma, \delta)^T$.

We need the vector of first derivatives $\frac{\partial}{\partial\theta}\mu(t,\theta)$.

$$\frac{\partial}{\partial\theta}\mu(t,\theta) = (1,\ t,\ te^{\gamma t+\delta},\ e^{\gamma t+\delta})^T \tag{4.77}$$

and also their integral over (x, y)

$$\int_x^y \frac{\partial}{\partial\theta}\mu(t,\theta)dt = \\ \left(y - x, \frac{y^2 - x^2}{2}, \frac{ye^{\gamma y+\delta} - xe^{\gamma x+\delta}}{\gamma} - \frac{e^{\gamma y+\delta} - e^{\gamma x+\delta}}{\gamma^2},\ e^{\gamma y+\delta} - e^{\gamma x+\delta}\right)^T. \tag{4.78}$$

So then according to equation (4.66), the equations to find the ML-estimates are

$$\sum_{m;N_m=1} \frac{1}{\hat{\alpha} + \hat{\beta}y_m + e^{\hat{\gamma}y_m+\hat{\delta}}} = \sum_m y_m - x_m \tag{4.79}$$

$$\sum_{m;N_m=1} \frac{t}{\hat{\alpha} + \hat{\beta}y_m + e^{\hat{\gamma}y_m+\hat{\delta}}} = \sum_m \frac{y_m^2 - x_m^2}{2} \tag{4.80}$$

$$\sum_{m;N_m=1} \frac{te^{\gamma t+\delta}}{\hat{\alpha} + \hat{\beta}y_m + e^{\hat{\gamma}y_m+\hat{\delta}}} = \sum_m \frac{(y_m\gamma - 1)e^{\gamma y_m+\delta} - (x_m\gamma - 1)e^{\gamma x_m+\delta}}{\gamma^2} \tag{4.81}$$

$$\sum_{m;N_m=1} \frac{e^{\gamma t+\delta}}{\hat{\alpha} + \hat{\beta}y_m + e^{\hat{\gamma}y_m+\hat{\delta}}} = \sum_m e^{\gamma y_m+\delta} - e^{\gamma x_m+\delta} \tag{4.82}$$

From these equations we find the estimates $\hat{\theta}_{ML}$.

The variance of the estimator is again given by the inverse information matrix. To find it, we construct a matrix from the vector of derivatives.

$$\frac{\partial}{\partial\theta}\mu(t,\theta)\frac{\partial}{\partial\theta^T}\mu(t,\theta) = \begin{bmatrix} 1 & t & te^{\gamma t+\delta} & e^{\gamma t+\delta} \\ t & t^2 & t^2e^{\gamma t+\delta} & te^{\gamma t+\delta} \\ te^{\gamma t+\delta} & t^2e^{\gamma t+\delta} & t^2e^{2\gamma t+2\delta} & te^{2\gamma t+2\delta} \\ e^{\gamma t+\delta} & te^{2\gamma t+2\delta} & te^{2\gamma t+2\delta} & e^{2\gamma t+2\delta} \end{bmatrix} \tag{4.83}$$

The probabilities $p_0^{(m)}(\tau,\theta)$ are given by

$$p_0^{(m)}(\tau,\theta) = e^{-\int_{x_m}^{\tau}(\alpha+\beta s+e^{\gamma s+\delta})ds}. \tag{4.84}$$

With that in hand, all necessary terms are now known to calculate the Fisher Information matrix as defined in equation (4.73).

Chapter 5

Discussion

In the first chapter of this thesis we introduced Markov chains and showed some results for them. The Kolmogorov differential equations are of particular interest, as we can use them to compute the transition probabilities. The transition probabilities and intensities are then used in the second chapter to derive equations for the expected prospective reserves. However, to fully calculate the quantities in this model, we need more. The intensities of transition are needed in the Kolmogorov differential equations to calculate the transition probabilities. The intensities also appear in the equations for the reserves. That is why we dedicated the last chapter to the estimation of the intensities of transition.

Using this thesis one has all tools in hand to calculate all necessary numbers for basic insurance contracts. That is, if we assume the model to be Markov. One might argue however, that this model is too simplistic. For example if we consider a case in health insurance, where an insured individual has been gravely ill many times. It could be argued that given a current healthy state the probability of falling ill again is higher after having been gravely ill before. Thus that the Markov property does not hold.

This example can then be countered by realizing that we could add many more states to the state space. For example we could have a state 'healthy', 'healthy, with light medical history' and 'healthy, with severe medical history'. However this does make the model less intuitive to use and makes the dimensions of the matrices with the transition coefficients larger as a result.

For further research it would be interesting to see how well the theory discussed in this thesis holds in practice and what changes need to be made to the model to make it (even) better applicable.

Appendix A

Actuarial Notation

Some of the notation used in this text find their origin in a standard defined by the International Actuarial Association. Below you find the part of this standard that is used throughout this work.

$$ {}_tp_x = \bar{F}(t) \tag{A.1} $$
$$ \mu_{x+t} = \mu(x+t) \tag{A.2} $$
$$ {}_nE_x = e^{-rn}\,{}_np_x \tag{A.3} $$

Concerning the expected value of insurances, given a policy holder aged x, we use the following:

$\bar{A}^1_{x:\overline{n}|}$ - The expected value of an n-year term insurance.

$\bar{A}_{x:\overline{n}|}$ - The expected value of an n-year endowment insurance.

$\bar{a}_{x\overline{n}|}$ - The expected value of an n-year temporary life annuity.

$\bar{a}_{\overline{n}|}$ - The present value of an n-year temporary life annuity.

Appendix B

Mortality

Throughout the thesis we make use of a non-negative random variable T, the amount of years remaining in a life, with a cumulative distribution function $F(t) = P[T \leq t]$. We also define the survival function as $\bar{F}(t) = P[T > t] = 1 - F(t)$. We assume that F is continuous, such that $f(t) = \frac{d}{dt}F(t)$. Norberg argues that it is convenient to work with the derivative of $-\ln \bar{F}$, which we call the force of mortality $\mu(t)$

$$\mu(t) = \frac{d}{dt}\Big[-\ln \bar{F}(t)\Big] = \frac{f(t)}{\bar{F}(t)}. \tag{B.1}$$

Here we need $\bar{F}(t) > 0$ for the function to be well defined. Now to get an expression for $\bar{F}(t)$ from this we first integrate over the interval (0,t)

$$\int_0^t \mu(\tau)d\tau = \int_0^t \frac{d}{d\tau}\Big[-\ln \bar{F}(\tau)\Big] = -\ln \bar{F}(t) + \ln \bar{F}(0). \tag{B.2}$$

Now since T is non-negative we have $\bar{F}(0) = 1 - F(0) = 1$. Using this we rewrite the above equation as

$$\bar{F}(t) = e^{-\int_0^t \mu(\tau)d\tau}. \tag{B.3}$$

Next we take a look at the distribution of the remaining life length of a person aged x. We denote this random variable by T_x. It is closely related T, as T_x is distributed as $T - x$ conditional on $T > x$. The distribution is then given by

$$F(t|x) = P[T \leq x + t | T > x] = \frac{F(x+t) - F(x)}{1 - F(x)}. \tag{B.4}$$

The surivival function can be found by taking $\bar{F}(t|x) = 1 - F(t|x)$, or by realizing

$$\bar{F}(t|x) = P[T > x + t | T > x] = \frac{\bar{F}(x+t)}{\bar{F}(x)}. \tag{B.5}$$

Now by equation B.3 we easily find

$$\bar{F}(t|x) = \frac{e^{-\int_0^{x+t} \mu(y)dy}}{e^{-\int_0^x \mu(y)dy}} = e^{-\int_x^{x+t} \mu(y)dy} = e^{-\int_0^t \mu(x+\tau)d\tau}. \tag{B.6}$$

If we want the density of T_x, we take the derivative of $\bar{F}(t|x)$ with respect to t, to get

$$f(t|x) = \frac{d}{dt}\bar{F}(t|x) = \mu(x+t)e^{-\int_0^t \mu(x+\tau)d\tau}. \tag{B.7}$$

Using the notation introduced above, this becomes

$$f(t|x) = {}_tp_x\mu_{x+t}. \tag{B.8}$$

Bibliography

[1] C.B. Boyer and U.C. Merzbach. *A History of Mathematics.* Wiley, 2011.

[2] Anders Hald. *A history of probability and statistics and their applications before 1750*, volume 501. John Wiley & Sons, 2003.

[3] Yann Le Bohec. *The imperial roman army.* Routledge, 2013.

[4] William Matthew Makeham. On the further development of gompertz's law (concluded). *Journal of the Institute of Actuaries (1886-1994)*, 28(4):316–332, 1890.

[5] Ragnar Norberg. Basic life insurance mathematics. *Lecture notes, Laboratory of Actuarial Mathematics, University of Copenhagen*, 2000.

[6] J.A. Rice. *Mathematical statistics and data analysis.* Brooks/Cole Pub. Co., 1988.

[7] R.L. Schilling. *Measures, Integrals and Martingales.* Cambridge University Press, 2017.

[8] M. Verbeek. *A Guide to Modern Econometrics.* Wiley, 2012.