

Electronic Thesis and Dissertation Repository

8-17-2016 12:00 AM

Actuarial Modelling with Mixtures of Markov Chains

Yuzhou Zhang, *The University of Western Ontario*

Supervisor: Bruce L. Jones, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Statistics and Actuarial Sciences

© Yuzhou Zhang 2016

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Statistics and Probability Commons](#)

Recommended Citation

Zhang, Yuzhou, "Actuarial Modelling with Mixtures of Markov Chains" (2016). *Electronic Thesis and Dissertation Repository*. 4026.

<https://ir.lib.uwo.ca/etd/4026>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

Abstract

Multi-state models are widely used in actuarial science because that they provide a convenient way of representing changes in people's statuses. Calculations are easy if one assumes that the model is a Markov chain. However, the memoryless property of a Markov chain is rarely appropriate.

This thesis considers several mixtures of Markov chains to capture the heterogeneity of people's mortality rates, morbidity rates, recovery rates, and ageing speeds. This heterogeneity may be the result of unobservable factors that affect individuals' health. The focus of this thesis is on investigating the behaviours of intensities of the observable transitions in the mixture models and assessing the applicability of the models.

We first investigate the disability process. Using a mixture model allows the future of the process to be dependent on by its history. We use mixtures of Markov chains with appropriate assumptions to investigate how the intensities of these processes depend on their histories.

We next explore an approach of using mixtures of Markov chains to model the dependence of two lifetimes. The mixture models allow the history of survivorship to affect future survival probabilities, which indicates a non-Markov behaviour. We discuss a simple mixture of two four-state Markov chains and a generalized mixture model.

Finally, we model the physiological ageing process by using mixtures. The traditional physiological ageing process assumes a homogeneous ageing speed. In fact, the ageing speed of each individual is characterized by his/her own health status. Using mixture models allows the process to reflect health status differences.

Keywords: Disability process, multi-state model, joint-life, dependent, lifetimes, intensity function, physiological age, heterogeneity.

Many of ideas contained in this thesis arose from discussions with my supervisor, Dr. Bruce Jones.

I authorize the Western University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I fully authorize the Western University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Yuzhou (Nikita) Zhang

Acknowledgements

Firstly I am heartily thankful for all the guidance from my supervisor, Bruce Jones. His immense knowledge offers me many ideas and suggestions, his encouragement motivates me to explore and learn, and his patient and careful scrutiny of all my work is the reason for this thesis to exist. I am truly grateful to be a student of Dr. Jones.

Sincere gratitude is also due to other members of my thesis committee: Jaap Spreeuw, David Stanford, Xiaoming Liu, and Geoff Wild, for their willingness and precious time to attend to my defence. In particular, I would like to thank my external examiner, Jaap Spreeuw, for acting on my examining committee even though there is a time zone difference. My examiners, David Stanford and Xiaoming Liu, have provided me many valuable ideas and support throughout my four years of graduate studies.

I would like to acknowledge the Natural Sciences and Engineering Research Council of Canada for providing me with the financial means to complete this thesis.

Also, my appreciation is to my dearest friends in Canada. Their support and confidence in me inspire me all the time.

Finally, I want to thank my beloved family, especially, my husband, Xiaohua Cheng, for their understanding and endless love.

Contents

Abstract	ii
Acknowledgements	iv
List of Figures	viii
List of Tables	x
1 Introduction	1
1.1 Background	1
1.2 Multi-State Models in Actuarial Science	2
1.3 Research Motivations	8
1.4 Outline of Thesis	10
2 Actuarial and Mathematical Background	12
2.1 The Force of Mortality	12
2.1.1 The Law of Mortality	13
2.2 Markov chains	14
2.3 Semi-Markov Processes	15
2.4 Mixtures of Markov chains	16
3 Modelling Disability Using Mixtures of Markov Chains	18
3.1 Introduction	18
3.2 A Simple Mixture of Markov Chains	19

3.2.1	The Example Histories	21
3.2.2	Calculation of Transition Intensities for Mixture Models	22
3.3	Empirical Investigations of Mixture Models	27
3.3.1	The Simple Model	27
3.3.2	A More Realistic Model	30
3.4	Conclusions and Discussion	35
4	Mixture of Markov Chains for the Joint-life Model	37
4.1	Introduction	37
4.2	Data Used in Our Analysis	39
4.3	Mixtures of Markov Chains Approach	41
4.3.1	Simple Mixture of Two Four-state Markov Chains	42
4.3.2	A Six-state Markov Model with Unobservable States	43
4.3.3	Transition Intensities	46
	Simple Mixture of Two Four-state Markov chains	47
	Six-state Model	48
4.4	Estimation Method	50
4.5	Fitting and Checking the Model	54
4.5.1	Specific Model Assumptions	54
4.5.2	Estimation Results	55
4.5.3	Model Testing	56
	Likelihood Ratio Tests	57
	Behaviour of Integrated Force of Mortality Checking	59
	Comparison of The Two Models	59
4.6	Conclusion and Discussion	62
5	Modelling a Physiological Aging Process by Mixture of Markov Chains	63
5.1	Introduction	63

5.2	The Physiological Aging Model	65
5.3	A Discrete Mixture of Markov Processes	68
5.3.1	Simple Mixture of Two Markov Processes	69
5.3.2	Mixture of Three Markov Processes	72
5.3.3	A Mixture Involving The Mover-stayer Model	77
5.4	A Continuous Mixture of Markov Processes	80
5.4.1	Simple Continuous Mixture of Markov processes	80
5.4.2	Embedded Autoregressive Process	82
5.5	Conclusion and Future Research	85
6	Conclusion	86
6.1	Concluding Remarks	86
6.2	Discussion and Further Research	88
	Bibliography	90
	A Appendix	101
	Curriculum Vitae	104

List of Figures

1.1	The alive-dead model	3
1.2	The joint-life model	4
1.3	The multiple decrement model	4
1.4	Disability process	5
1.5	The phase-type ageing model	5
3.1	Simple mixture of two Markov chains	20
3.2	Example histories for four individuals	22
3.3	The morbidity rate given \mathcal{H}_t^1 (solid) and \mathcal{H}_t^2 (dashed)	29
3.4	The recovery rate given \mathcal{H}_t^3 (solid) and \mathcal{H}_t^4 (dashed)	29
3.5	The mortality rate of an active individual given \mathcal{H}_t^1 (solid) and \mathcal{H}_t^2 (dashed)	30
3.6	The mortality rate of a disabled individual given \mathcal{H}_t^3 (solid) and \mathcal{H}_t^4 (dashed)	30
3.7	Mixture of three-state Markov chain with time-dependent intensities	31
3.8	The morbidity rate of an individual given \mathcal{H}_t^1 (solid) and \mathcal{H}_t^2 (dashed), and the morbidity rate from each Markov chain (dotted)	33
3.9	The recovery rate of a disabled individual given \mathcal{H}_t^3 (solid) and \mathcal{H}_t^4 (dashed), and the recovery rate from each Markov chain (dotted)	34
3.10	The mortality rate of an active individual given \mathcal{H}_t^1 (solid) and \mathcal{H}_t^2 (dashed), and the mortality rate from each Markov chain (dotted)	34
3.11	The mortality rate of a disabled individual given \mathcal{H}_t^3 (solid) and \mathcal{H}_t^4 (dashed), and the mortality rate from each Markov chain (dotted)	35
4.1	Four-state with mixture model	43

4.2	Six-state model	44
4.3	Mixture version of the six-state model	45
4.4	Baseline Cumulative Hazard Function. The top left panel shows the cumulative function of $\lambda^{01}(t \mathcal{H}_t^1)$. The bottom left panel shows the cumulative function of $\lambda^{23}(t \mathcal{H}_t^2)$. The top right panel shows the cumulative function of $\lambda^{02}(t \mathcal{H}_t^1)$. The bottom right panel shows the cumulative function of $\lambda^{13}(t \mathcal{H}_t^3)$	60
4.5	The force of mortality for a surviving spouse of a fixed age at various times since the first death.	61
5.1	The physiological ageing model	65
5.2	Age-specific force of mortality for Swedish Cohort (1911)	68
5.3	Mixture of two Markov processes	70
5.4	Phase distributions of mixture of two Markov processes	71
5.5	Mixture of three Markov processes	73
5.6	Embedded discrete-time process	74
5.7	Phase distributions of the mixture of three Markov processes	76
5.8	Transitions of a mover	77
5.9	Phase distributions of the mixture of mover-stayer processes	79
5.10	Phase distributions of the continuous mixture of Markov processes	81
5.11	Phase distributions of autoregressive process mixture model with assumptions on ϕ	83
5.12	Phase distributions of autoregressive process mixture model with $\phi = 0.95$	84
5.13	Phase distributions of autoregressive process mixture model with assumptions on σ	84

List of Tables

4.1	Summary of male data	40
4.2	Summary of female data	41
4.3	Estimates from mixture of two four-state Markov chains	55
4.4	Estimates from six-state Markov chain with unobservable states	56
5.1	Parameter Estimates for Swedish Cohort (1911)	68

Chapter 1

Introduction

1.1 Background

Actuarial science deals with risk, uncertainty and finance. Dickson, Hardy and Waters (2009) mentioned that *“the first actuaries were employed by life insurance companies in the early eighteenth century to provide a scientific basis for managing the companies’ assets and liabilities. The liabilities depended on the number of deaths occurring amongst the insured lives each year.”* This description of responsibility of early actuaries provides evidence that many insurance products are based upon the living status and health status of insureds. As a popular actuarial model, the multi-state model provides a convenient view of these statuses.

We can find the description of multi-state models in Jones (1994): *“we assume that, at any time, an individual is in one of a number of states. The individual’s presence in a given state or movement (transition) from one state to another may have some financial impact. Our task then is to quantify this impact.”* Therefore, in a multi-state model, transitions between states represent the changes of status, and the transition rates (intensities) represent speeds of the changes.

Recently, multi-state models have played a very important role in actuarial modelling because they are intuitive and easy to analyse using numerical techniques. In general, Markov models and semi-Markov models are commonly used due to their mathematical convenience. However, like a coin has two sides, this mathematical convenience simplifies the calculations

but the resulting model may not always be realistic. Therefore, my research aims at capturing the realism that traditional Markov or semi-Markov models miss.

The purpose of this chapter is to introduce some background to motivate the study of Actuarial Modelling with Mixtures of Markov Chains, and it is organized as follows. Section 1.2 discusses some multi-state models used in actuarial science. After introducing sufficient background knowledge of actuarial multi-state models, Section 1.3 presents the motivation of my research and briefly discusses the research I have done. Finally, an outline of my thesis contents are presented in Section 1.4.

1.2 Multi-State Models in Actuarial Science

A multi-state process is a stochastic process whose index set represents time and whose value at any time is in a finite set called the state space. A multi-state model is a mathematical representation of a multi-state process, which often involves simplifying assumptions to make the model tractable.

Multi-state models offer a natural way to solve certain actuarial problems. Seal (1977) reviewed the history of multi-state models, noting that the development of multi-state models can be traced back to the model of smallpox by Bernoulli (1766). In Bernoulli's model, one state consisted of individuals who have never had smallpox, the other state consisted of individuals who contracted smallpox and either die almost immediately or survive and never suffer the disease again. After years of research, the numerical techniques for multi-state models were improved and multi-state models are now commonly adopted to solve actuarial problems.

To understand the multi-state model, we start from the simplest two-state model with state space $\{0, 1\}$. As shown in Figure 1.1, if an individual is alive, the process is in state 0. When the individual dies, the process enters state 1 and stays there afterwards. Therefore, for a whole life insurance, premiums are payable when the process is in state 0 and the death benefit is paid at the time of transition to state 1. In the case of a simple life annuity, benefits are payable

while the process is in state 0 and cease once the transition to state 1 is made. This simple model captures all the life contingent information for the calculation of actuarial values that only depend on whether a single individual is alive or dead at a given age.

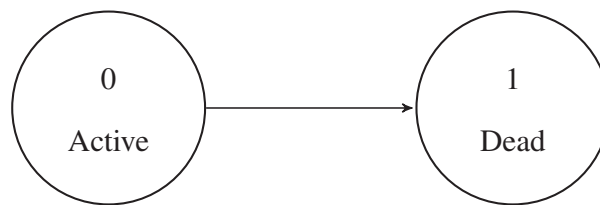


Figure 1.1: The alive-dead model

Based on the simple alive-dead model, additional states are considered to analyse more complicated models. To investigate the dependence of individuals' lifetimes, a multiple life model was considered by Norberg (1989), Wolthuis (2003), and Spreeuw and Wang (2008) as an application of the multi-state model. Norberg (1989) introduced a four-state model as shown in Figure 1.2. This model can be used for joint life insurances or joint life and last survivor annuity policies issued to (x) and (y) . To understand this model, consider an example of a married couple with husband age x and wife age y . If both are alive, the process is in state 0, once the husband dies, the process makes a transition to state 1 and stays there until it makes another transition to state 3 when the wife dies. Similarly, if the wife dies first, the process makes a transition to state 2 first and makes another transition to state 3 when the husband dies. This four-state model inspires us to further investigate the dependence of the lifetimes of a married couple.

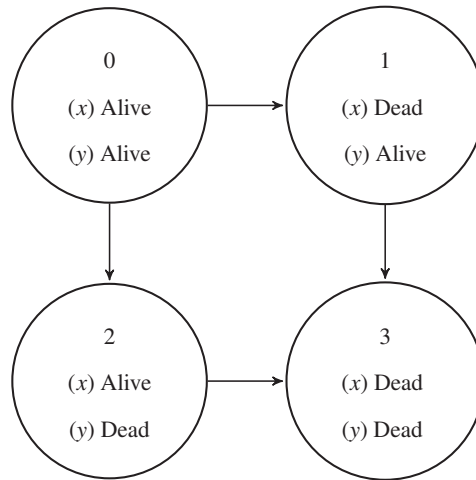


Figure 1.2: The joint-life model

Another well-known generalization of the alive-dead model is the multiple decrement model. A multiple decrement model is characterized by an initial state and several possible exit states. As shown in Figure 1.3, there are l possible decrements. Starting in state 0 (*active/insured*), the process may move to any one of the absorbing states $1, \dots, l$. The multiple decrement model is frequently used in actuarial applications.

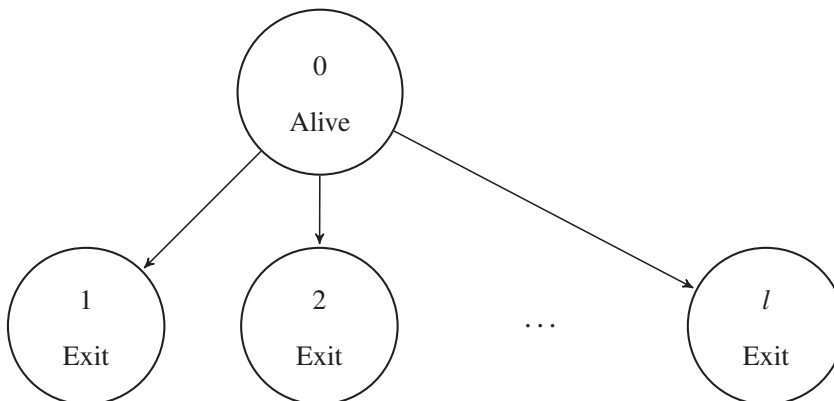


Figure 1.3: The multiple decrement model

Notice that it is not necessary to assume the process will never transit back to a state once it leaves. In modelling disability, a three-state model is used. As shown in Figure 1.4, if the individual is healthy/active, the process is in state 1. Once the individual gets sick/diabled,

the process makes a transition to state 2. Also, the individual can recover from the disabled status. Therefore, probabilities of being in a given state at a future time are more complicated to determine in this model.

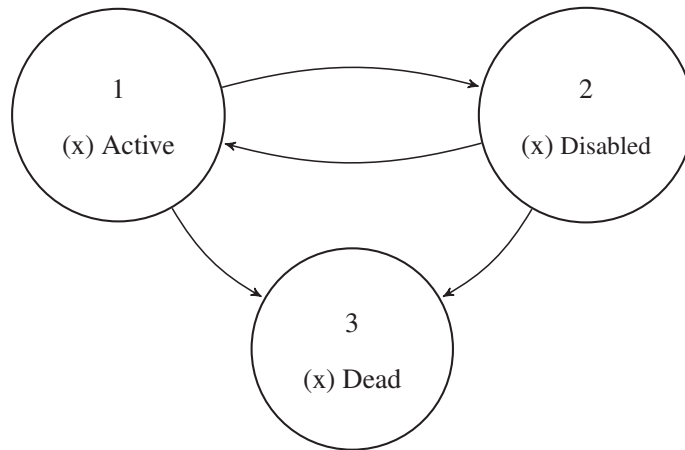


Figure 1.4: Disability process

To model the heterogeneity of human mortality, a phase-type physiological ageing model can be used. The model captures the fact that people can age physiologically differently. As shown in Figure 1.5, the model involves n states that represent physiological ages. For each state, there is a possibility to leave and enter an absorbing state representing death or to age physiologically. In this model, the time of death follows a phase-type distribution and the mortality rate depends on the actual physiological health condition of an individual instead of his/her calendar age.

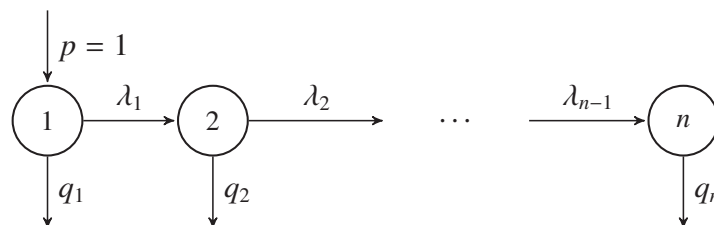


Figure 1.5: The phase-type ageing model

More examples of and probabilistic calculations using multi-state models can be found in

Dickson, Hardy and Waters (2009, ch.8). Many actuarial applications of multi-state models can be found in Jones and Wu (2014).

Multi-state models may be characterized in terms of their transition intensities. A transition intensity is the instantaneous rate of transition between two states. It is normally a function of time, but it may also depend on aspects of the past history of the process. One may have prior knowledge of how the intensities should behave, and specific parameters can be estimated from data about the process of interest. Methods of estimating the intensities for multi-state models are introduced by Waters (1984).

When a multi-state process consists of a collection of random variables indexed by a continuous time parameter, the multi-state model is a continuous time model. It is convenient to assume that a multi-state model is a continuous-time Markov chain, whose transition intensities depend only on the current state and time. Thus, the previous history of an individual does not affect the conditional probabilities of the future states. This property is known as the memoryless property. For a semi-Markov model, the memoryless property does not hold because transition intensities depend on the time since entering the current state.

Many researchers have studied applications of Markov chains. Hoem (1969) first proposed using Markov chains in an actuarial context and showed how they can be used to calculate life insurance premiums and reserves. Hoem (1972) considered time inhomogeneous semi-Markov models and discussed their applications to several actuarial and demographic processes. In doing so, Hoem points out where it is important to allow transition intensities to depend on the time spent in the current state. Ramlau-Hansen (1988) used a Markov model to examine how the stochastic approach to life contingencies provides insight in life insurance calculations. Specifically, Ramlau-Hansen showed how one can analyse the variability in the emergence of profit in life insurance. Ramlau-Hansen (1991) examined several methods of distribution of surplus with a Markov chain framework. Norberg (1995) derived ordinary differential equations to obtain higher moments of present values by considering continuous time Markov chains. Applications of Markov chains to continuing care retirement commu-

nities (CCRC) were discussed in Jones (1995, 1996, 1997a and b) and Rickayzen and Walsh (2002). Markov chains have been used in critical illness models by Macdonald, et al (2005 a and b). Ji, et al (2012) modelled reverse mortgages by a semi-Markov model. For disability insurance, Tolly and Manton (1991) proposed Markov models involving various risk factors for morbidity and mortality. The use of Markov and semi-Markov models for disability insurance calculations was discussed by Waters (1984), (1990), and a review was provided by Haberman and Pitacco (1999). D'Amico, et al (2009) applied a semi-Markov process to an extended disability model with additional states. D'Amico, et al (2013) used a discrete time inhomogeneous semi-Markov process to model disability insurance. Maegebier (2013) introduced a trivariate Markov renewal reward process to capture the dependence between the durations in two successive states in the disability process. Hassan Zaden, et al (2014) proposed a phase-type disability model with several ageing phases and several disability stages. For modelling the dependence of lifetimes, Norberg (1989) introduced a four-state continuous-time Markov chain. Discussions on assumptions of the force of mortality in Norberg's model can be found in Wolthuis (2003). Spreeuw and Wang (2008) and Spreeuw and Owadally (2013) extended Norberg's work by introducing additional states. Ji, et al (2011) extended Norberg's work by using a semi-Markov model. Mortality models involving time-homogeneous Markov chains have been proposed by Lin and Liu (2007) as a physiological ageing process. In this model, the time until death follows a phase-type distribution. Lin and Liu (2012) developed the physiological ageing process to a more flexible model by assuming that the underlying ageing process follows a subordinated Markov process. This model fits historical mortality data and projected mortality data well and meets many model selection criteria. Govorun, et al (2016) extended the phase-type ageing model by considering the impact of an individual's health history using health cost data.

1.3 Research Motivations

In actuarial modelling, the theory of Markov chains provides mathematical tools for solving problems in multi-state models. However, although the memoryless property makes calculations easier, it is not always reasonable to assume that future changes of state are independent of past changes. Actually, Pitacco (1995a) recognized the lack of realism in Markov chains. Considering an in-force insurance policy, Pitacco mentioned three more complicated but more realistic assumptions on the dependence of future transition intensities:

1. '*Duration-since-initiation*' dependence means the dependence is on the age at policy issue. In this case, issue-select transition intensities should be used;
2. '*Duration-in-current-state*' dependence means the dependence is on the time since the latest transition to that state. In this case, the model is a semi-Markov model;
3. '*Health story*' of the insured means the dependence is on the total time spent in each state since policy issue.

Nowadays, many actuarial models implement the first two assumptions. However, a more realistic model may consider the whole history of the process.

Why is it important to model the realism? To answer this question, we need to be aware that in a population, an individual's mortality, morbidity, and longevity are affected by his/her own risk factors. Many researchers in the medical, epidemiological, and statistical fields have studied these risk factors. Kestenbaum (2000), Horiuchi and Wilmoth (1998), and Deaton and Lubosky (2003) investigated the pattern of mortality changes by age. Trovato and Lalu (1998), and Valling (1995) investigated the pattern of mortality changes due to sex differences. Williams and Collins (1996), Kallan (1997), Hummer (1996), Vaillant and Mukamal (2001), Deaton and Paxson (1999), Elo and Preston (1997), Knox and Tomlin (1997), Wolfson, et al (1990), Lobmayer and Wilkinson (2000), Judge (1995), Sorlie, et al (1995), Preston and Elo (1995), Pappas, et al (1993), and Rogers (1995) studied the risk factors related to socio-economic and demographic status, such as race, education, income, occupation, religion, and

marital status. Attanasio and Emmerson (2001) noted that education is a risk factor associated with morbidity more than mortality. Adler, et al (1994), Allison, et al (1999), Godtfredsen, et al (1998), Prescott, et al (1998), Tessier, et al(2000), Landi, et al (1999), Wei, et al (1999), and Lantz, et al (1998) studied health behaviours such as smoking, obesity, and alcohol intake. In addition, some health indicators, such as blood pressure, blood sugar level and cholesterol level, are studied respectively by Rastas, et al (2006), Bagshaw, et al (2009), and Anderson, et al (1987). Brown and McDaid (2003) identified several risk factors affecting mortality rates by reviewing the literature. They also pointed out the importance of classifying risks affecting mortality rates. Kwon and Jones (2006) grouped these risk factors in three categories:

- *socioeconomic/demographic risk factors: age, sex, education, income, occupation, marital status, etc;*
- *behavioural risk factors: smoking, alcohol intake, physical activity, dietary habit, body mass index, etc;*
- *health indicators: blood pressure, cholesterol level, blood sugar level, etc.*

Heterogeneity in mortality rates significantly affects the pricing and valuation of insurance products. As concluded in Rothschild and Stiglitz (1976), if insurers offer the same price to both low-risk and high-risk individuals, the low-risk individuals will not buy the product, and most insureds will be high-risk individuals. Consequently, the insurer makes a loss. Therefore, it is important to model this heterogeneity.

Many models are developed to capture the heterogeneity in mortality rates due to observable risk factors, such as the multiple decrement model that is discussed by Waters and Wilkie (1988) and included in substantial text books referred in Seal (1977). Vaupel, Manton and Stallard (1979) introduced the well-known frailty model that can capture the heterogeneity in mortality rates due to an unobservable frailty factor.

We consider a different approach to capture the heterogeneity of mortality rates in reality by using mixtures of Markov chains to model future transitions conditional on past informa-

tion of an individual. We start by investigating mixtures of disability processes, because it is intuitive that the morbidity rate, recovery rate, and mortality rate of a healthier individual are different from a frailer individual. By using a mixture model, we are able to model these rates conditional on the history of illness of an individual. We also investigate mixtures of joint-life models. As the dependence of two lifetimes is an important topic in actuarial research, we are interested in modelling changes in the mortality rate for the survivor after the spouse dies. By using a mixture model, we are able to model the mortality rate for a spouse conditional on the survivorship information of the couple.

Notice that heterogeneity also exists in the ageing of individuals as their health is affected by many risk factors. Numerous medical and health studies investigate the reasons for this heterogeneity. These include Liochev (2015), Mitnitski and Rockwood (2016), and Kirkwood (2015). To investigate the heterogeneity in the ageing speed of individuals, we use mixtures of physiological ageing processes having fixed mortality rates at specific physiological ages.

This approach of using mixture models allows transition intensities to depend on additional information about the history of the process. The focus on modelling intensities is because a multi-state model can be characterized by the transition intensities, and actuarial calculations can be performed using transition intensities. However, some quantities such as the probability of being in a Markov chain can be investigated for other research purposes.

1.4 Outline of Thesis

In order to provide the necessary background for the later chapters, Chapter 2 reviews models of the force of mortality and definition of Markov chains. Main ideas of constructing a mixture model of Markov chains are also presented. Chapter 3 investigates mixtures of two three-state Markov chains, where each Markov chain models a disability process. The applicability of this mixture model is also discussed. Chapter 4 investigates mixtures of four-state Markov chains, where each Markov chain is a joint-life model. The dependence of the two lifetimes is

investigated by two different approaches of constructing the mixture models. Chapter 5 extends the physiological ageing process introduced by Lin and Liu (2007) by using several mixture approaches to capture the heterogeneity in ageing speeds. Finally, Chapter 6 concludes the thesis and discusses potential future research related to my work.

Chapter 2

Actuarial and Mathematical Background

2.1 The Force of Mortality

The force of mortality is fundamental to modelling lifetimes. Since the two-state model with states “Alive” and “Dead” is used to represent lifetimes, the force of mortality has the identical meaning as the transition intensity in this model.

In life contingencies, the force of mortality is normally denoted by μ_x for (x) (an individual at age x). Suppose T_x is a continuous random variable representing the future lifetime of (x) , the force of mortality at age x is defined as

$$\mu_x = \lim_{h \rightarrow 0^+} \frac{1}{h} P(T_0 \leq x + h \mid T_0 > x).$$

This means the force of mortality μ_x is the instantaneous rate of dying at age x . With the survival function $S_x(t) = P(T_x > t)$ and related probability density function $f_x(t)$, we have

$$\begin{aligned} \mu_x &= \lim_{h \rightarrow 0^+} \frac{1}{h} \frac{S_0(x) - S_0(x+h)}{S_0(x)} \\ &= \frac{S'_0(x)}{S_0(x)} \\ &= \frac{f_0(x)}{S_0(x)}. \end{aligned}$$

Hence,

$$\mu_{x+t} = \frac{f_x(t)}{S_x(t)} \quad \text{and} \quad S_x(t) = \exp\left(-\int_0^t \mu_{x+s} ds\right).$$

This means when the force of mortality function is known, the lifetime distribution is known. Therefore, it is important to determine a force of mortality function that can capture the empirical mortality rates of people.

2.1.1 The Law of Mortality

Since the first life table was constructed empirically in the 1600's, researchers have investigated mathematical expressions to describe the mortality rate of people. De Moivre (1725) suggested a mortality law with one parameter ω representing the limiting age; the force of mortality function is represented as

$$\mu_{x+t} = \frac{1}{\omega - x - t}, \quad 0 < t \leq \omega - x.$$

The most influential and popular mortality law was developed by Gompertz (1825). The Gompertz' law suggests the force of mortality involves two positive constant parameters B interpreted as the general mortality level and c representing the exponential coefficient of mortality growth. So the force of mortality is expressed as

$$\mu_{x+t} = Bc^{x+t}.$$

Gompertz' law describes an exponential ageing of individuals. Many applications of Gompertz' law establish that it captures the pattern of a wide range of mortality data between the ages of 30 and 80. To improve Gompertz' law so that the pattern of mortality at younger ages can be better captured, Makeham (1860) simply extended Gompertz' law by adding one age

independent parameter A , resulting in the force of mortality

$$\mu_{x+t} = A + Bc^{x+t}.$$

Reviews on the development of mortality models are presented by Wetterstrand (1981), Olshansky and Carnes (1997), and Pitacco (2003).

2.2 Markov chains

The theory of Markov chains was introduced in 1907 by Andrei Andreyevich Markov, and named in his honor. Since then, Markov chains have been popularly used in various fields.

Definition If for all $n \geq 0$ and $i_0, i_1, \dots, i_{n+1} \in \mathcal{S}$, the process $\{X_n, n \geq 0\}$ has both properties below:

- $P(X_0 = i_0) = \alpha_{i_0}$,
- $P(X_{n+1} = i_{n+1} \mid X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = P(X_{n+1} = i_{n+1} \mid X_n = i_n)$,

the process $\{X_n, n \geq 0\}$ is a Markov chain with a state space \mathcal{S} , whose initial distribution is given by $\{\alpha_i, i \in \mathcal{S}\}$ and transition probability matrix is $\mathbf{P} = (p_{ij}, i, j \in \mathcal{S})$.

In later chapters of this thesis, we use a continuous time Markov chain which can be denoted by $\{X(t), t \geq 0\}$. Therefore, for $i, j \in \mathcal{S}$,

$$P(X(t) = j \mid X(s) = i, X(u) = x(u) \text{ for } 0 \leq u \leq s) = P(X(t) = j \mid X(s) = i) = P^{ij}(s, t).$$

The instantaneous rate of transition from state i to state j at time t is called the transition intensity, and is denoted $\mu^{ij}(t)$. For $i, j \in \mathcal{S}$, the transition intensity matrix is normally denoted

$$\mathbf{Q}(t) = \begin{cases} \mu^{ij}(t) & i \neq j \\ -\sum_{j=0, j \neq i}^n \mu^{ij}(t) & i = j \end{cases}.$$

The transition probability matrix is given by

$$P(s, t) = \lim_{r \rightarrow \infty} \prod_{k=1}^r (I - Q(u_k)(u_k - u_{k-1})),$$

where $0 = u_0 < u_1 < u_2 < \dots < u_r = t - s$, and as $r \rightarrow \infty$, $\max(u_i - u_{i-1}) \rightarrow 0$.

To calculate the transition probability matrix, Kolmogorov's forward equations can be solved numerically. These equations are

$$\frac{\partial P^{ij}(s, t)}{\partial t} = - \sum_{k=0, k \neq j}^n (P^{ij}(s, t)\mu^{jk}(t) - P^{ik}(s, t)\mu^{kj}(t)), \quad i, j \in \mathcal{S}.$$

If the transition intensities are constant, then $\mu^{ij}(t) = \mu^{ij}$ and $Q = (\mu^{ij}, i, j \in \mathcal{S})$, where the (i, i) entry of Q is $-\sum_{j=0, j \neq i}^n \mu^{ij}$. In this case, we can use the matrix exponential $P(t) = \exp(Qt)$, where the (i, j) entry of $P(t)$ is $P(X(t+s) = j | X(s) = i)$. The matrix exponential of a square matrix A is defined as

$$\exp(A) = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

Many methods have been proposed for numerical computation of the matrix exponential (see Moler and Van Loan, 2003). The `expm` function in the *Matrix* package in R uses Ward's diagonal Padé approximation with three step preconditioning.

Therefore, by knowing the transition intensity matrix, we can determine the transition probability matrix of a Markov chain with constant transition intensities.

2.3 Semi-Markov Processes

As noted by Hoem (1972), if the transition intensities functions depend on the duration in the current state as well as the current time, the process is semi-Markovian. In this case, we can express the transition intensity matrix as $Q(t, v)$, where v is the duration in the current state.

While we do not actually use semi-Markov models in this thesis, this is the behavior that we refer to when we use the term semi-Markov process.

2.4 Mixtures of Markov chains

Generally, there are many different ways to construct a mixture model, our research is inspired by the approaches below:

1. *A simple discrete mixture of n Markov chains with the same state space \mathcal{S} , but different transition intensities*

Here we are mixing over the processes themselves. If we let the random variable M represent the Markov chain, then for $m = 1, 2, \dots, n$, $P(M = m)$ is the mixing probability associated with the Markov chain with transition intensities $\{\mu_m^{ij}(t) : i, j \in \mathcal{S}, i \neq j\}$. This model was discussed by Aalen (1987).

This model is easy to construct. However, when n is small, this simple mixture model may exhibit a lack of smoothness, in which case, we may consider other approaches of constructing a mixture model.

2. *A mixture of Markov chains that involves randomness in the time scale*

This was studied by Aalen (1988). In this case, our stochastic process X is such that $X(t) = Y(Zt)$, where Y is a Markov chain with transition intensities $\{\mu^{ij}(t) : i, j \in \mathcal{S}, i \neq j\}$, and Z is a positive random variable, referred as the mixing variable. This means that, conditional on Z , the process X is a Markov chain with transition intensities given by $Z\mu^{ij}$. In the case where Z is discrete with support $\{z_1, z_2, \dots, z_n\}$, this approach leads to a special case of the above discrete mixture with $\mu_m^{ij} = z_m\mu^{ij}$ for all i, j pairs. The random time scale approach is rather restrictive, as each Markov Chain in the mixture has the same relationships among its transition intensities.

We recognize the accelerated lifetime idea in this model. That is, different mixing variables Z decelerate or accelerate the speed of moving through the process. This model allows for continuous distributions for the mixing random variable Z , but it may not be flexible enough because the same mixing factor is applied to all the intensities.

3. *A mixture of Markov chains with a different mixing variable associated with each transition intensity*

In this case, we have the collection of random variables $\{Z^{ij} : i, j \in \mathcal{S}, i \neq j\}$. Conditional on these random variables, the process X is a Markov chain with transition intensities $Z^{ij}\mu^{ij}$.

Notice that this model is more general; the first and second approaches are special cases of this approach. The well-known frailty model in survival analysis is a special case of this mixture construction (see Hougaard, 1984). When the Z^{ij} s are discrete random variables, the resulting process is itself a Markov chain.

4. *Another more general mixture of Markov chains involving mixing probabilities that depend on the history of the process*

In this case, $M(t)$ indicates which Markov chain is applicable at time t , and the mixing probability is

$$P(M(t) = m \mid X(u) = x(u), 0 \leq u \leq t).$$

This is quite a flexible mixture model.

Chapter 3

Modelling Disability Using Mixtures of Markov Chains

3.1 Introduction

In order to analyse sickness insurance and disability income benefits, the three-state model shown in Figure 1.4 is used widely. With the state space $\{1, 2, 3\}$, where each state indicates the status of an individual at age x . Extensions on this three-state model are discussed by Haberman and Pitacco (1999), D'Amico, et al (2009), D'Amico, et al (2013), and Maegebier (2013).

In reality, we know that, as a person ages, the force of mortality (transition intensity to death) increases. Also, a disabled individual should have a larger force of mortality than an active individual. In addition, we expect that the intensity of disability incidence will be an increasing function of age that is greater than the force of mortality. It may even depend on past disabilities. Finally, we expect that an individual who has been disabled for a short time has a higher recovery intensity than an individual who has been disabled for a long time. Therefore, the intensity of recovery should decrease with duration of disability.

To construct a tractable model of heterogeneous behaviour in an individual's health status,

we use a mixture of three-state Markov chains, in which transition intensities are conditional on one's sickness history. We do not only consider the semi-Markov behaviour that the force of mortality depends on duration of being disabled, but also consider the possibility that the force of mortality depends on the number of past spells of disability.

The goal of this chapter is to investigate how the transition intensities of a mixture of Markov chains depend on an individual's sickness history. In order to do so, we investigate two mixture models with different assumptions about the transition intensities. We also explore the suitability of mixtures of Markov chains for modelling the disability process. Section 3.2 constructs a simple mixture of two Markov chains with constant transition intensities and determines future transition intensities given an individual's history. Section 3.3 investigates a modification of the simple mixture of two Markov chains by using time-dependent transition intensities. Section 3.4 concludes this chapter and discusses potential future research related to our model.

3.2 A Simple Mixture of Markov Chains

In a population, there always exist healthier people and frailer people. To model this heterogeneity due to individuals' physical conditions, we consider a simple mixture of two three-state Markov chains. We assume that both Markov chains have the same state space, and the distinctions in the constant transition intensities of each Markov chain characterize the population; one is a healthier population, the other is a frailer population.

As shown in Figure 3.1, the random variable M determines the Markov chain. The mixing probabilities are $P(M = 1)$ and $P(M = 2)$. The transition intensities for the two Markov chains are $\{\mu_1^{ij} : i, j \in \{1, 2, 3\}, i \neq j\}$ and $\{\mu_2^{ij} : i, j \in \{1, 2, 3\}, i \neq j\}$, respectively.

This mixture of two Markov chains leads to a process that depends on an individual's sickness history, because intensities of observable transitions are functions of a history that contains the information about duration and spells of disability for an individual up to the

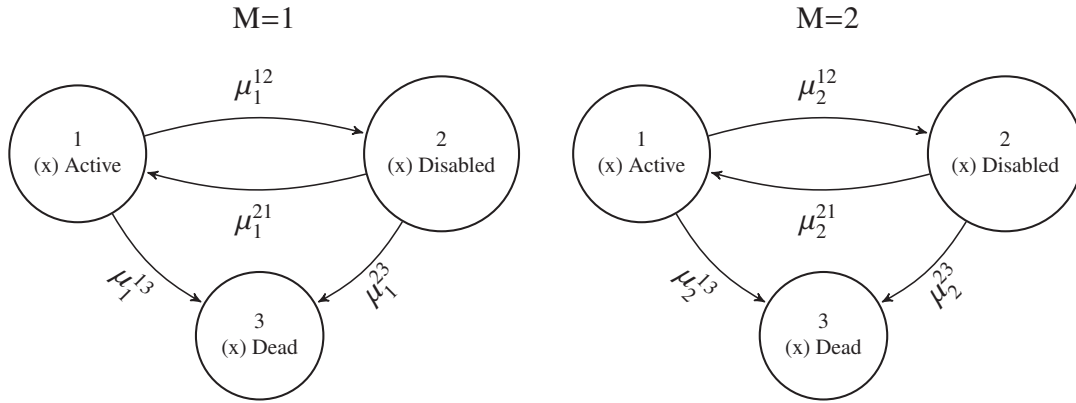


Figure 3.1: Simple mixture of two Markov chains

current time. The history up to time t is denoted by \mathcal{H}_t .

Using similar notation to Aalen (1988), in our mixture model, the transition intensity from observable state i to observable state j given the sickness history of an individual is denoted by $\lambda^{ij}(t | \mathcal{H}_t)$ where $i, j \in \{1, 2, 3\}$ and $i \neq j$. We have

$$\lambda^{ij}(t | \mathcal{H}_t) = \mu_1^{ij} P(M = 1 | \mathcal{H}_t) + \mu_2^{ij} P(M = 2 | \mathcal{H}_t). \quad (3.1)$$

Thus, the transition intensity of the mixture model is dependent on the history of the process through the conditional probabilities on the right-hand side of (3.1).

In general, for a mixture of n 3-state Markov chains, we have

$$\begin{aligned} \lambda^{ij}(t | \mathcal{H}_t) &= \frac{\sum_{m=1}^n \mu_m^{ij} P(\mathcal{H}_t | M = m) P(M = m)}{\sum_{m=1}^n P(\mathcal{H}_t | M = m) P(M = m)} \\ &= \frac{1}{P(\mathcal{H}_t)} \sum_{m=1}^n \mu_m^{ij} P(\mathcal{H}_t | M = m) P(M = m) \\ &= \sum_{m=1}^n \mu_m^{ij} P(M = m | \mathcal{H}_t). \end{aligned}$$

3.2.1 The Example Histories

We have established that the transition intensities for a mixture of Markov chains depend on the history of the process. In order to develop specific expressions for these intensities, additional notations are required.

In general, we let $X = \{X(t), t \geq 0\}$ be the multi-state process of interest, and let $N(t)$ be the total number of transitions by time t with $N = \{N(t), t \geq 0\}$. Also, let $N^{ij}(t)$ be the number of transitions from state i to state j during $[0, t]$. Let the time of the r^{th} transition be T_r for $r = 1, 2, \dots$, and $T_0 = 0$. Clearly, $T_r > T_{r-1}$. For simplicity, let $X_r = X(T_r)$. Finally, the history up to time t , $\mathcal{H}_t = \{x(s), 0 \leq s \leq t\}$ can be expressed as $\mathcal{H}_t = \{(T_r, X_r) : r = 0, 1, 2, \dots, N(t)\}$.

Example 3.2.1 Consider the following histories of four individuals,

$$\begin{aligned}\mathcal{H}_t^1 &= \{(0, 1)\} \\ \mathcal{H}_t^2 &= \{(0, 1), (\epsilon_1, 2), (t - \epsilon_2, 1)\} \\ \mathcal{H}_t^3 &= \{(0, 1), (t - \epsilon_2, 2)\} \\ \mathcal{H}_t^4 &= \{(0, 1), (\epsilon_1, 2)\}\end{aligned}$$

where $0 < \epsilon_1 < t - \epsilon_2 < t$, and ϵ_1 and ϵ_2 are small relative to t (see Figure 3.2).

Individuals 1 and 2 are in state 1 at time 0 and time t . However, individual 2 spends most of the time in state 2, suggesting that he/she is more likely to be frailer than individual 1. Also, individuals 3 and 4 are in state 1 at time 0 and state 2 at time t . Individual 3 spends most of his/her time in state 1, and individual 4 spends most of his/her time in state 2. This suggests that individual 4 is more likely to be frailer than individual 3.

These four individuals will be considered in our investigation in Section 3.3.1 of the impact of the history on the transition intensities.

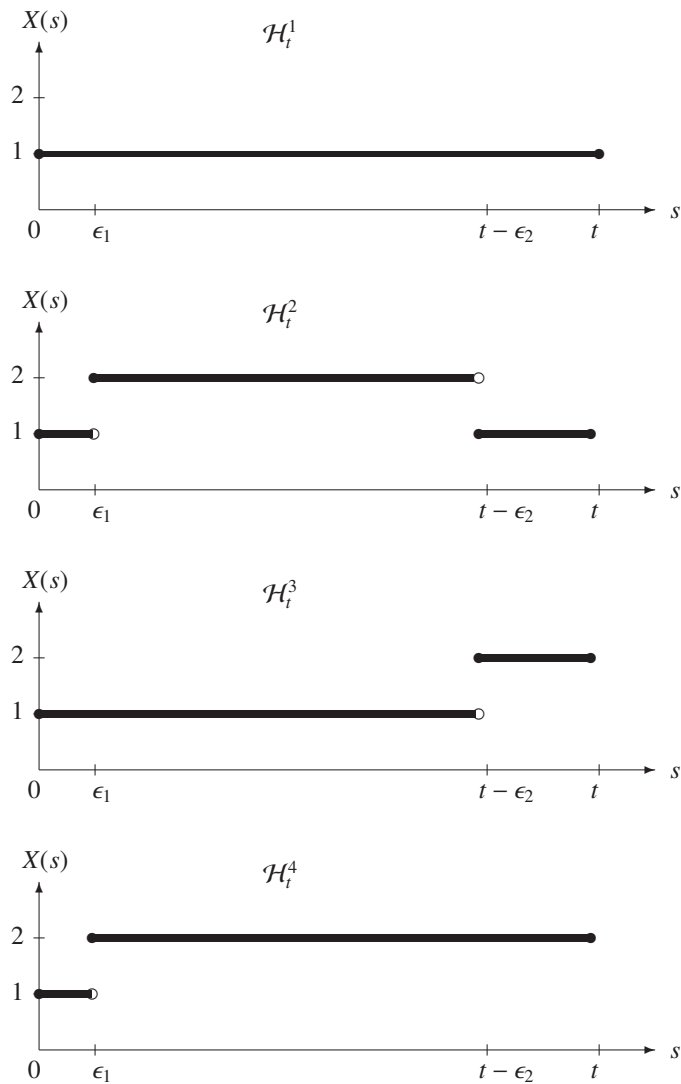


Figure 3.2: Example histories for four individuals

3.2.2 Calculation of Transition Intensities for Mixture Models

We established in Section 3.2 that, with a simple mixture of two 3-state Markov chains, the transition intensities depend on the history of the process. Since the third approach of constructing a mixture model in Section 2.4 is a general approach, we now derive transition intensities following that approach with mixing random variables Z^{ij} .

Suppose a mixture of Markov chains has state space $\{1, 2, \dots, q\}$. Let μ^{ij} be the constant transition intensity from state i to state j and the vector $\mathbf{Z} = \{Z^{ij}, i, j \in \{1, 2, \dots, q\}, i \neq j\}$ contains the mixing variables for the corresponding transition possibilities.

As shown by Aalen (1987):

$$\begin{aligned} \lambda^{ij}(t | \mathcal{H}_t) &= \lim_{h \rightarrow 0} \frac{P(X(t+h) = j | X(t) = i, \mathcal{H}_t)}{h} \\ &= \lim_{h \rightarrow 0} \frac{P(X(t+h) = j, \mathcal{H}_t | X(t) = i)}{h P(\mathcal{H}_t)} \end{aligned} \quad (3.2)$$

$$\begin{aligned} &= \frac{1}{E_{\mathbf{Z}}[P(\mathcal{H}_t | \mathbf{Z})]} \lim_{h \rightarrow 0} \frac{E_{\mathbf{Z}}[P(X(t+h) = j, \mathcal{H}_t | X(t) = i, \mathbf{Z})]}{h} \\ &= \frac{1}{E_{\mathbf{Z}}[P(\mathcal{H}_t | \mathbf{Z})]} E_{\mathbf{Z}} \left[\lim_{h \rightarrow 0} \frac{P(X(t+h) = j | X(t) = i, \mathbf{Z})}{h} P(\mathcal{H}_t | \mathbf{Z}) \right] \end{aligned} \quad (3.3)$$

$$= \frac{E_{\mathbf{Z}}[\mu^{ij} Z^{ij} P(\mathcal{H}_t | \mathbf{Z})]}{E_{\mathbf{Z}}[P(\mathcal{H}_t | \mathbf{Z})]}, \quad (3.4)$$

where expressions (3.3) and (3.4) follow from the fact that, conditional on \mathbf{Z} , $X(t+h)$ and \mathcal{H}_t are independent and X is a Markov chain with transition intensities $\mu^{ij} Z^{ij}$.

Now, $P(\mathcal{H}_t | \mathbf{Z})$ is the conditional probability density of the history up to time t given the mixing variables. In the other words, it is the probability density of the stays (in the several states) and the transitions that occurred during $[0, t]$ assuming the Markov process that results when we condition on the mixing variables.

To develop an expression for $P(\mathcal{H}_t | \mathbf{Z})$, we note that the probability that an individual stays in state i for the sojourns in state i that occur during $[0, t]$ is

$$\exp \left(- \sum_{j=1, j \neq i}^q \left(\mu^{ij} Z^{ij} \int_0^t I(x(s) = i) ds \right) \right),$$

where I is the indicator function, and $\{x(s), 0 \leq s \leq t\}$ is the observed history. Also, $\mu^{ij} Z^{ij}$ is the transition intensity from state i to state j , so that the intensity of the transition that occurred at time t_r is

$$\sum_{i,j} \mu^{ij} Z^{ij} I(x_{r-1} = i, x_r = j)$$

, where $x_k = x(t_k)$ and $\sum_{i,j}$ means $\sum_{i=1}^q \sum_{j=1, j \neq i}^q$. We also use $\prod_{i,j}$ to mean $\prod_{i=1}^q \prod_{j=1, j \neq i}^q$. Combining the probabilities of the stays in different states during $[0, t]$ and the intensities for the $n(t)$ transitions during $[0, t]$, we have

$$\begin{aligned} P(\mathcal{H}_t = \{x(s), 0 \leq s \leq t\} | \mathbf{Z}) &= \prod_{i=1}^q \exp\left(-\sum_{j=1, j \neq i}^q \mu^{ij} Z^{ij} \int_0^t I(x(s) = i) ds\right) \prod_{r=1}^{n(t)} \left(\sum_{i,j} \mu^{ij} Z^{ij} I(x_{r-1} = i, x_r = j)\right) \\ &= \exp\left(-\sum_{i,j} s^{ij} Z^{ij}\right) \prod_{r=1}^{n(t)} \left(\sum_{i,j} \mu^{ij} Z^{ij} I(x_{r-1} = i, x_r = j)\right), \end{aligned} \quad (3.5)$$

where $s^{ij} = \mu^{ij} \int_0^t I(x(s) = i) ds$.

Example 3.2.2 For individuals with the four histories in Example 3.2.1, we have

$$\begin{aligned} P(\mathcal{H}_t^1 | \mathbf{Z}) &= \exp(-(\mu^{13} Z^{13} + \mu^{12} Z^{12})t) \\ P(\mathcal{H}_t^2 | \mathbf{Z}) &= \exp(-(\mu^{13} Z^{13} + \mu^{12} Z^{12})(\epsilon_1 + \epsilon_2) - (\mu^{21} Z^{21} + \mu^{23} Z^{23})(t - \epsilon_2 - \epsilon_1)) \mu^{12} Z^{12} \mu^{21} Z^{21} \\ P(\mathcal{H}_t^3 | \mathbf{Z}) &= \exp(-(\mu^{13} Z^{13} + \mu^{12} Z^{12})(t - \epsilon_2) - (\mu^{21} Z^{21} + \mu^{23} Z^{23})\epsilon_2) \mu^{12} Z^{12} \\ P(\mathcal{H}_t^4 | \mathbf{Z}) &= \exp(-(\mu^{13} Z^{13} + \mu^{12} Z^{12})\epsilon_1 - (\mu^{21} Z^{21} + \mu^{23} Z^{23})(t - \epsilon_1)) \mu^{12} Z^{12} \end{aligned}$$

Equation (3.5) provides the probabilities needed in expression (3.4). However, in order to evaluate (3.4), we need a specific assumption about the distribution of \mathbf{Z} . Suppose $L(\mathbf{r})$ is the Laplace transform function for the mixing variables \mathbf{Z} . That is,

$$\begin{aligned} L_{\mathbf{Z}}(\mathbf{r}) &= E_{\mathbf{Z}} \left[\exp\left(-\left(\sum_{i,j} r^{ij} Z^{ij}\right)\right) \right] \\ &= \int \dots \int_{\mathbf{z}} \exp\left(-\left(\sum_{i,j} r^{ij} z^{ij}\right)\right) dF_{\mathbf{Z}}(\mathbf{z} | \mathcal{H}_t) \end{aligned}$$

From (3.4) and (3.5), we see that

$$\lambda^{ij}(t | \mathcal{H}_t) = -\frac{\mu^{ij}}{L_{\mathbf{Z}}^{(N(t))}(\mathbf{r})} \left(\frac{\partial}{\partial r^{ij}} L_{\mathbf{Z}}^{(N(t))}(\mathbf{r}) \right) \Big|_{\mathbf{r}=\mathbf{s}}, \quad (3.6)$$

where the form $L_{\mathbf{Z}}^{(\mathbf{a})}(\mathbf{r})$ is such that vector \mathbf{a} gives the number of derivatives of $L_{\mathbf{Z}}(\mathbf{r})$ taken with

respect to each component of \mathbf{r} , $N(t) = \{N^{12}(t), N^{13}(t), \dots, N^{q,q-1}(t)\}$, and $\mathbf{s} = \{s^{12}, s^{13}, \dots, s^{q,q-1}\}$. The expression on the right-hand side of equation (3.6) is an analogue to that given by Anlen (1988). Notice that expression (3.6) describes the dependence of the transition intensities on the different aspects of the history.

Suppose we express the simple mixture of two three-state Markov chains in terms of the third approach of constructing a general mixture model in Section 2.4. In this case, we let $\mu^{ij} = \mu_1^{ij}$ and a random variable B follow a Bernoulli distribution with $p = P(B = 1) = P(M = 2)$. Hence, the mixing variable is

$$Z^{ij} = \frac{\mu_1^{ij}(1 - B) + \mu_2^{ij}B}{\mu_1^{ij}},$$

and the Laplace transform has the form

$$\begin{aligned} L_{\mathbf{Z}}(\mathbf{r}) &= E \left[\exp \left(- \sum_{i,j} r^{ij} Z^{ij} \right) \right] \\ &= E \left[\exp \left(- \sum_{i,j} r^{ij} \frac{\mu_1^{ij}(1 - B) + \mu_2^{ij}B}{\mu_1^{ij}} \right) \right] \\ &= \exp \left(- \sum_{i,j} r^{ij} \right) E \left[\exp \left(- \sum_{i,j} r^{ij} B \frac{\mu_2^{ij} - \mu_1^{ij}}{\mu_1^{ij}} \right) \right] \\ &= \exp \left(- \sum_{i,j} r^{ij} \right) \left(\exp \left(- \sum_{i,j} r^{ij} \frac{\mu_2^{ij} - \mu_1^{ij}}{\mu_1^{ij}} \right) p + (1 - p) \right) \\ &= \exp \left(- \sum_{i,j} r^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}} \right) p + \exp \left(- \sum_{i,j} r^{ij} \right) (1 - p). \end{aligned} \quad (3.7)$$

After taking derivatives of result (3.7), by using equation (3.6), we get

$$\lambda^{k\ell}(t | \mathcal{H}_t) = \frac{\mu_2^{k\ell} \prod_{i,j} \left(\frac{\mu_2^{ij}}{\mu_1^{ij}} \right)^{N^{ij}(t)} \exp \left(- \sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}} \right) p + \mu_1^{k\ell} \exp \left(- \sum_{i,j} s^{ij} \right) (1 - p)}{\prod_{i,j} \left(\frac{\mu_2^{ij}}{\mu_1^{ij}} \right)^{N^{ij}(t)} \exp \left(- \sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}} \right) p + \exp \left(- \sum_{i,j} s^{ij} \right) (1 - p)}.$$

We can also use the third approach in Section 2.4 to construct a continuous mixture of three-

state Markov chains. Assume the mixing variables have independent gamma distributions such that $Z^{12} \sim \Gamma(\alpha_1, \beta_1)$, $Z^{21} \sim \Gamma(\alpha_2, \beta_2)$, $Z^{13} \sim \Gamma(\alpha_3, \beta_3)$, and $Z^{23} \sim \Gamma(\alpha_4, \beta_4)$, where the Laplace transform for $\Gamma(\alpha_i, \beta_i)$ is

$$\left(\frac{\beta_i}{\beta_i + r} \right)^{\alpha_i}.$$

Therefore, $\mathbf{r} = (r^{12}, r^{21}, r^{13}, r^{23})'$, and

$$L_{\mathbf{Z}}(\mathbf{r}) = \left(\frac{\beta_1}{\beta_1 + r^{12}} \right)^{\alpha_1} \left(\frac{\beta_2}{\beta_2 + r^{21}} \right)^{\alpha_2} \left(\frac{\beta_3}{\beta_3 + r^{13}} \right)^{\alpha_3} \left(\frac{\beta_4}{\beta_4 + r^{23}} \right)^{\alpha_4}.$$

After taking derivatives

$$\begin{aligned} L_{\mathbf{Z}}^{(N(t))}(\mathbf{r}) &= (-1)^{N^{12}(t)+N^{21}(t)+N^{13}(t)+N^{23}(t)} \frac{\beta_1^{\alpha_1} \beta_2^{\alpha_2} \beta_3^{\alpha_3} \beta_4^{\alpha_4}}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)\Gamma(\alpha_4)} \\ &\times \frac{\Gamma(\alpha_1 + N^{12}(t))\Gamma(\alpha_2 + N^{21}(t))\Gamma(\alpha_3 + N^{13}(t))\Gamma(\alpha_4 + N^{23}(t))}{(\beta_1 + r^{12})^{\alpha_1 + N^{12}(t)}(\beta_2 + r^{21})^{\alpha_2 + N^{21}(t)}(\beta_3 + r^{13})^{\alpha_3 + N^{13}(t)}(\beta_4 + r^{23})^{\alpha_4 + N^{23}(t)}}. \end{aligned}$$

Finally, the transition intensities given the histories are

$$\begin{aligned} \lambda^{12}(t | \mathcal{H}_t) &= \mu^{12} \frac{\alpha_1 + N^{12}(t)}{(\beta_1 + s^{12})}, \\ \lambda^{21}(t | \mathcal{H}_t) &= \mu^{21} \frac{\alpha_2 + N^{21}(t)}{(\beta_2 + s^{21})}, \\ \lambda^{13}(t | \mathcal{H}_t) &= \mu^{13} \frac{\alpha_3 + N^{13}(t)}{(\beta_3 + s^{13})}, \\ \lambda^{23}(t | \mathcal{H}_t) &= \mu^{23} \frac{\alpha_4 + N^{23}(t)}{(\beta_4 + s^{23})}. \end{aligned}$$

Therefore, assuming the mixing random variables follow gamma distributions yields nice properties for the transition intensities given the histories. That is, these transition intensities are functions of the number of transitions and the cumulated intensities in the states. By specifying the parameters for each gamma distribution, the intensity functions can be determined.

3.3 Empirical Investigations of Mixture Models

3.3.1 The Simple Model

Consider the simple case of a mixture of two three-state Markov chains as shown in Figure 3.1. To construct a mixture of a healthier Markov chain and a frailer Markov chain, we let

$$\mu_1^{12} = 0.05, \mu_1^{21} = 0.2, \mu_1^{13} = 0.01, \mu_1^{23} = 0.03,$$

and

$$\mu_2^{12} = 0.1, \mu_2^{21} = 0.1, \mu_2^{13} = 0.02, \mu_2^{23} = 0.06.$$

Therefore, $M = 1$ represents a healthier Markov chain because, compared to $M = 2$, the recovery rate is larger and mortality rate and morbidity rates are smaller. Also, we let $P(M = 1) = 0.6$.

As discussed in Section 3.2, given the histories for the four individuals in Example 3.2.1, we determine the intensities of observable transitions conditional on the applicable histories. In this section, we show $\lambda^{12}(t | \mathcal{H}_t^k)$ as an example, expressions for the other transition intensities are in Appendix A. We have

$$\lambda^{12}(t | \mathcal{H}_t^k) = 0.05P(M = 1 | \mathcal{H}_t^k) + 0.1P(M = 2 | \mathcal{H}_t^k)$$

for $k = 1, 2$.

Notice that, using Bayes' rule,

$$P(M = m | \mathcal{H}_t^k) = \frac{P(\mathcal{H}_t^k | M = m)P(M = m)}{P(\mathcal{H}_t^k | M = 1)P(M = 1) + P(\mathcal{H}_t^k | M = 2)P(M = 2)},$$

for $m = 1, 2$. For the applicable histories in Example 3.2.1,

$$\begin{aligned}
P(\mathcal{H}_t^1 | M = 1) &= \exp(-(\mu_1^{12} + \mu_1^{13})t) = e^{-(0.05+0.01)t}, \\
P(\mathcal{H}_t^1 | M = 2) &= \exp(-(\mu_2^{12} + \mu_2^{13})t) = e^{-(0.1+0.02)t}, \\
P(\mathcal{H}_t^2 | M = 1) &= \exp(-(\mu_1^{12} + \mu_1^{13})(\epsilon_1 + \epsilon_2)) \mu_1^{12} \exp(-(\mu_1^{21} + \mu_1^{23})(t - \epsilon_1 - \epsilon_2)) \mu_1^{21} \\
&= e^{-(0.2+0.03)(t-\epsilon_1-\epsilon_2)-(0.05+0.01)(\epsilon_1+\epsilon_2)}(0.05)(0.2), \\
P(\mathcal{H}_t^2 | M = 2) &= \exp(-(\mu_2^{12} + \mu_2^{13})(\epsilon_1 + \epsilon_2)) \mu_2^{12} \exp(-(\mu_2^{21} + \mu_2^{23})(t - \epsilon_1 - \epsilon_2)) \mu_2^{21} \\
&= e^{-(0.1+0.06)(t-\epsilon_1-\epsilon_2)-(0.1+0.02)(\epsilon_1+\epsilon_2)}(0.1)(0.1).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\lambda^{12}(t | \mathcal{H}_t^1) &= \frac{0.05e^{-0.06t}0.6 + 0.1e^{-0.12t}0.4}{e^{-0.06t}0.6 + e^{-0.12t}0.4}, \\
\lambda^{12}(t | \mathcal{H}_t^2) &= \frac{0.05e^{-0.23t+0.17(\epsilon_1+\epsilon_2)}(0.01)0.6 + 0.1e^{-0.16t+0.04(\epsilon_1+\epsilon_2)}(0.01)0.4}{e^{-0.23t+0.17(\epsilon_1+\epsilon_2)}(0.01)0.6 + e^{-0.16t+0.04(\epsilon_1+\epsilon_2)}(0.01)0.4}.
\end{aligned}$$

These intensities of transition from state 1 to state 2 are plotted in Figure 3.3. The solid curve represents the transition intensity $\lambda^{12}(t | \mathcal{H}_t^1)$, where $\mathcal{H}_t^1 = \{(0, 1)\}$; the dashed curve represents the transition intensity $\lambda^{12}(t | \mathcal{H}_t^2)$, where $\mathcal{H}_t^2 = \{(0, 1), (\epsilon_1, 2), (t - \epsilon_2, 1)\}$. The two thinner dashed lines represent the constant transition intensities of the two Markov chains, where the bottom one represents the healthier μ_1^{12} and the top one represents the frailer μ_2^{12} . The figure shows the limiting case as ϵ_1 and ϵ_2 go to 0. In the figure, at time 0, $\lambda^{12}(t | \mathcal{H}_t^1)$ and $\lambda^{12}(t | \mathcal{H}_t^2)$ have the same value, as \mathcal{H}_0^k provides no information. Notice that this value is simply $\mu_1^{12}P(M = 1) + \mu_2^{12}P(M = 2)$. As t increases from 0, we gain more and more information. As no transition to state 2 is observed in the history \mathcal{H}_t^1 , this history provides the information of a healthier individual. Thus, \mathcal{H}_t^1 gives us greater certainty that $M = 1$ as t increases. Therefore, $\lambda^{12}(t | \mathcal{H}_t^1) \rightarrow \mu_1^{12}$ as $t \rightarrow \infty$. Similarly, the history \mathcal{H}_t^2 gives us greater certainty that $M = 2$ as t increases, since the process spends almost all of its time in state 2. Therefore, $\lambda^{12}(t | \mathcal{H}_t^2) \rightarrow \mu_2^{12}$ as $t \rightarrow \infty$.

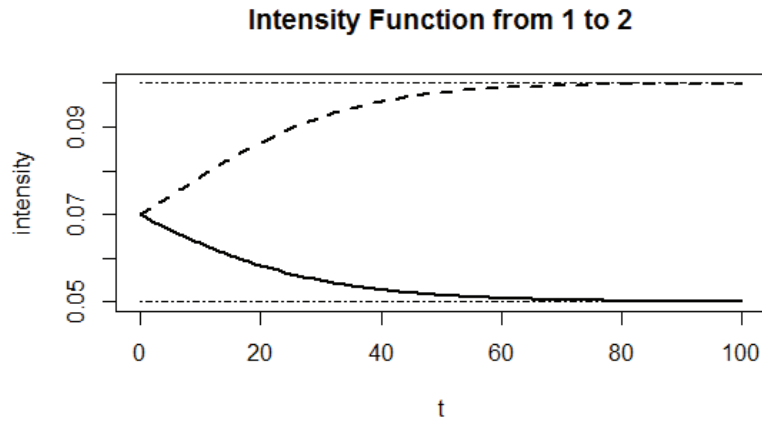


Figure 3.3: The morbidity rate given \mathcal{H}_t^1 (solid) and \mathcal{H}_t^2 (dashed)

Figures 3.4, 3.5 and 3.6 show the transition intensities, $\lambda^{21}(t | \mathcal{H}_t^k)$, $\lambda^{13}(t | \mathcal{H}_t^k)$ and $\lambda^{23}(t | \mathcal{H}_t^k)$ for applicable \mathcal{H}_t^k in Example 3.2.1. Note that, for $\lambda^{2j}(t | \mathcal{H}_t^k)$, the applicable histories are \mathcal{H}_t^3 and \mathcal{H}_t^4 . The appearance and interpretation of the graphs are similar to those for Figure 3.3. However, we see that the solid curve and dashed curve are reversed in Figure 3.4. This is because a healthier individual has a higher recovery rate.

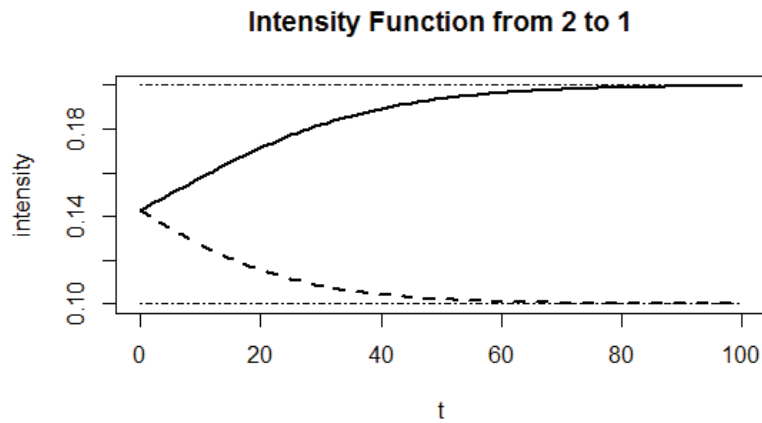


Figure 3.4: The recovery rate given \mathcal{H}_t^3 (solid) and \mathcal{H}_t^4 (dashed)

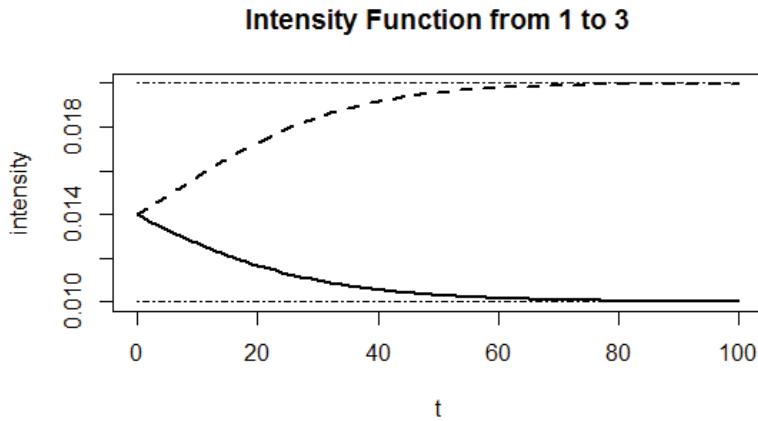


Figure 3.5: The mortality rate of an active individual given \mathcal{H}_t^1 (solid) and \mathcal{H}_t^2 (dashed)

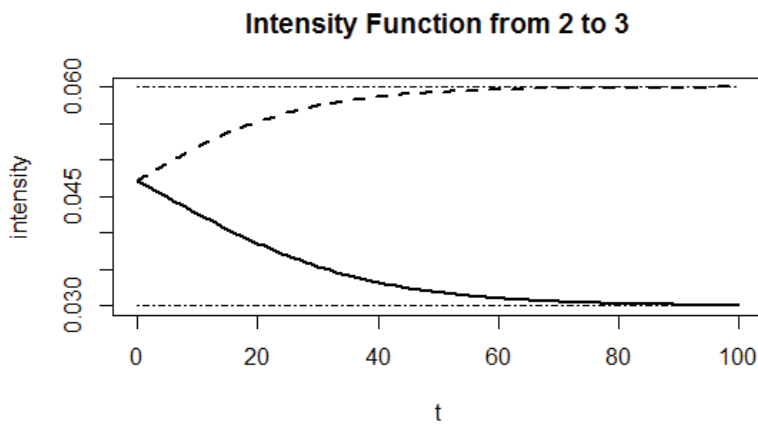


Figure 3.6: The mortality rate of a disabled individual given \mathcal{H}_t^3 (solid) and \mathcal{H}_t^4 (dashed)

3.3.2 A More Realistic Model

In reality, the force of mortality increases as an individual ages. Therefore, transition intensities are normally functions of time. In the three-state disability model, since the transition intensities from state 1 to states 2 and 3 represent the mortality rate of an individual, and the transition intensity from state 2 to state 3 represents the morbidity rate of an individual, they are increasing functions of time. However, the transition intensity from state 2 to state 1 represents the recovery rate; it is a decreasing function of time. We use the notation $\mu^{ij}(t)$ to represent the

time-dependent transition intensity from state i to state j for the three-state Markov chain. The resulting mixture model is depicted in Figure 3.7.

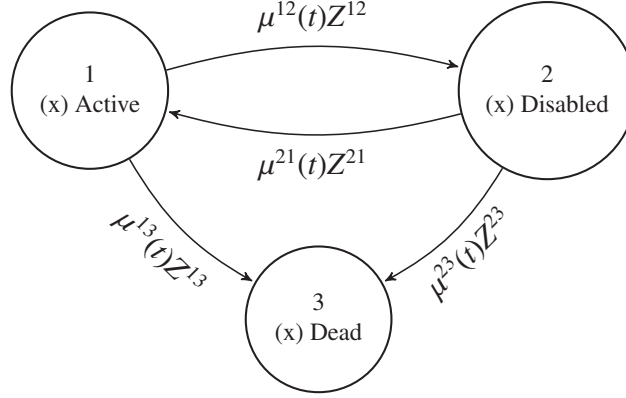


Figure 3.7: Mixture of three-state Markov chain with time-dependent intensities

For simplicity, consider the mixture of a healthier Markov chain and a frailer Markov chain. To implement the third approach of constructing a mixture model in Section 2.4 as discussed in Section 3.2.2, suppose

$$B = \begin{cases} 0 & \text{if } M = 1 \\ 1 & \text{if } M = 2 \end{cases}.$$

Thus, the mixture variables are

$$Z^{ij} = \frac{\mu_1^{ij}(1 - B) + \mu_2^{ij}B}{\mu_1^{ij}},$$

for $i, j = 1, 2, 3, i \neq j$.

As the mixture distributions are not changed from those in Section 3.2.2, $L_Z(\mathbf{r})$ is given by equation (3.7). However, the assumption of time-dependent transition intensities leads to

$$\lambda^{ij}(t | \mathcal{H}_t) = -\frac{\mu^{ij}(t)}{L_Z^{(N(t))}(\mathbf{r})} \left(\frac{\partial}{\partial r^{ij}} L_Z^{(N(t))}(\mathbf{r}) \right) \Big|_{r=s},$$

where $s = \{s^{12}, s^{21}, s^{13}, s^{23}\}$, and $s^{ij} = \int_0^t \mu^{ij}(s) I(x(s) = i) ds$.

Example 3.3.1 Using the mixture of two Markov chains discussed in Section 3.3.1 with the

constant mixing variables, we let $\mu^{12}(t) = \mu^{12}(0.001)(1.1)^t$, $\mu^{21}(t) = \mu^{21}(0.99)^t$, $\mu^{13}(t) = \mu^{13}(0.001)(1.1)^t$, and $\mu^{23}(t) = \mu^{23}(0.001)(1.1)^t$, where $\mu^{12} = \mu_1^{12}$, $\mu^{21} = \mu_1^{21}$, $\mu^{13} = \mu_1^{13}$, $\mu^{23} = \mu_1^{23}$.

Again, consider the histories for the four individuals in Example 3.2.1. We determine $\lambda^{12}(t | \mathcal{H}_t^k)$ in this section. The others can be found in Appendix A.

To explore the extreme case, in which the individual was active for the entire past lifetime, we have

$$\lambda^{12}(t | \mathcal{H}_t^1) = \frac{\mu_2^{12}}{\mu_1^{12}} \exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p)$$

$$\mu_1^{12}(0.001)(1.1)^t \frac{\exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p)}{\exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p)},$$

where

$$s^{12} = \int_0^t \mu^{12}(v) dv, \quad s^{13} = \int_0^t \mu^{13}(v) dv, \quad s^{21} = 0, \quad \text{and} \quad s^{23} = 0.$$

In the opposite extreme in which the individual started in active status, but became disabled immediately, remained disabled until recovery just before time t , we have

$$\lambda^{12}(t | \mathcal{H}_t^2) = \frac{\left(\frac{\mu_2^{12}}{\mu_1^{12}}\right)^2 \frac{\mu_2^{21}}{\mu_1^{21}} \exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p)}{\frac{\mu_2^{12}}{\mu_1^{12}} \frac{\mu_2^{21}}{\mu_1^{21}} \exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p)},$$

where

$$s^{12} \rightarrow 0, \quad s^{13} \rightarrow 0, \quad s^{21} \rightarrow \int_0^t \mu^{21}(v) dv, \quad s^{23} \rightarrow \int_0^t \mu^{23}(v) dv, \quad \text{as} \quad \epsilon_1, \epsilon_2 \rightarrow 0.$$

With the values of μ_1^{ij} and μ_2^{ij} from Section 3.3.1, we can calculate and plot $\lambda^{ij}(t | \mathcal{H}_t^k)$. In order to improve readability, we take natural logarithms of $\lambda^{12}(t | \mathcal{H}_t^k)$, $\lambda^{13}(t | \mathcal{H}_t^k)$ and $\lambda^{23}(t | \mathcal{H}_t^k)$. Figure 3.8, Figure 3.9, Figure 3.10 and Figure 3.11 show the plots of $\ln(\lambda^{12}(t | \mathcal{H}_t^k))$,

$\lambda^{21}(t | \mathcal{H}_t^k)$, $\ln(\lambda^{13}(t | \mathcal{H}_t^k))$ and $\ln(\lambda^{23}(t | \mathcal{H}_t^k))$, respectively.

Note that, each $\lambda^{ij}(t | \mathcal{H}_t^k)$ is determined using the applicable \mathcal{H}_t^k . That is, \mathcal{H}_t^1 and \mathcal{H}_t^2 for $i = 1$, and \mathcal{H}_t^3 and \mathcal{H}_t^4 for $i = 2$. Furthermore, as ϵ_1 and ϵ_2 go to 0, \mathcal{H}_t^1 and \mathcal{H}_t^3 represent healthier histories, while \mathcal{H}_t^2 and \mathcal{H}_t^4 represent frailer histories. For each plot, we set two “bench marks” represented by two dotted lines showing $\mu_1^{ij}(t)$ and $\mu_2^{ij}(t)$, while the solid curves represent the behaviour of (the logarithm of) transition intensities given the healthier histories and the dashed curves represent the behaviour of (the logarithm of) transition intensities given the frailer histories.

From Figure 3.8, we can see that the individual with the healthier history, namely \mathcal{H}_t^1 , tends to have a smaller intensity of disablement. Also, notice that $\lambda^{12}(t | \mathcal{H}_t^1)$ approaches $\mu_1^{12}(t)$ more slowly than $\lambda^{12}(t | \mathcal{H}_t^2)$ approaches $\mu_2^{12}(t)$. This is because long stays in state 2 provide more information about the value of M than long stays in state 1 due to the higher transition intensity from state 2. The same interesting results are shown in the Figure 3.10 and Figure 3.11 as well. Notice that Figure 3.9 shows the decreasing trend of $\lambda^{21}(t | \mathcal{H}_t^3)$ and $\lambda^{21}(t | \mathcal{H}_t^4)$, while the other three figures show the increasing trend. This is because we assume the recovery rate decreases as an individual ages, but the mortality rate and morbidity rate increase as an individual ages.

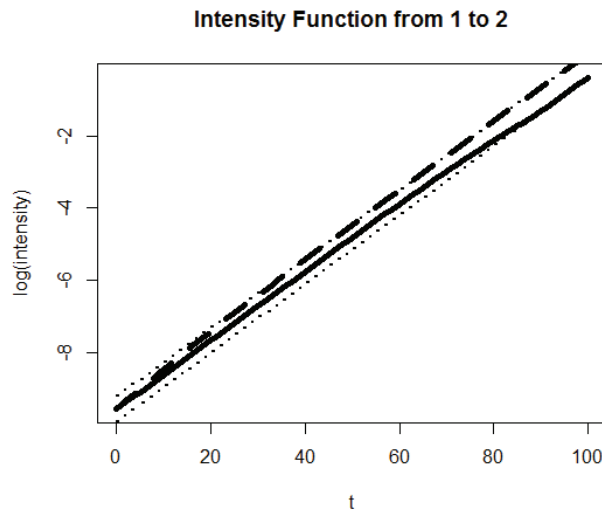


Figure 3.8: The morbidity rate of an individual given \mathcal{H}_t^1 (solid) and \mathcal{H}_t^2 (dashed), and the morbidity rate from each Markov chain (dotted)

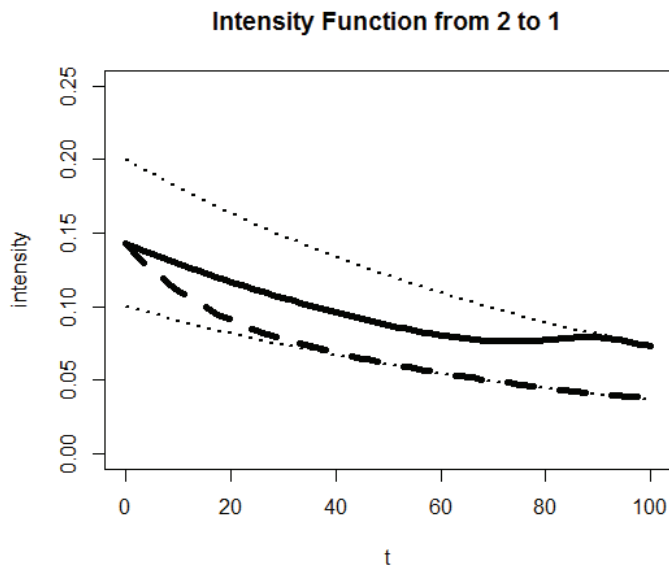


Figure 3.9: The recovery rate of a disabled individual given \mathcal{H}_t^3 (solid) and \mathcal{H}_t^4 (dashed), and the recovery rate from each Markov chain (dotted)

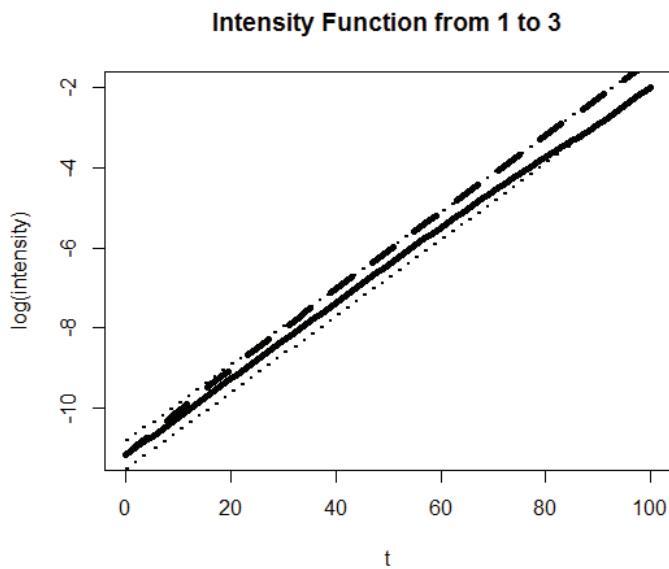


Figure 3.10: The mortality rate of an active individual given \mathcal{H}_t^1 (solid) and \mathcal{H}_t^2 (dashed), and the mortality rate from each Markov chain (dotted)

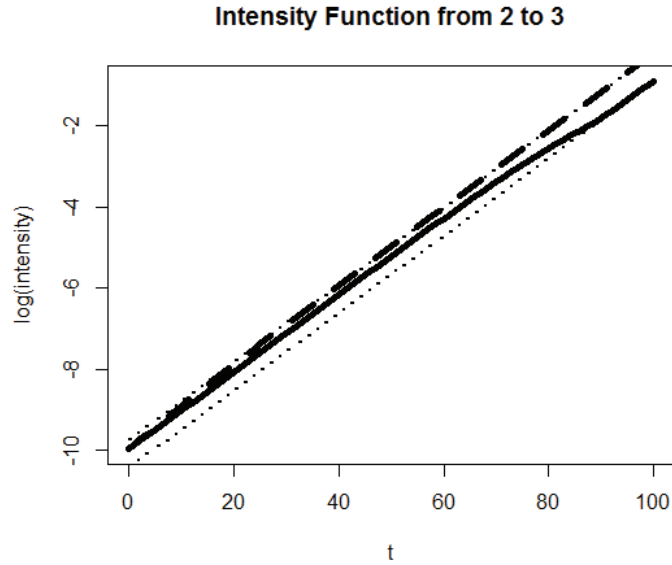


Figure 3.11: The mortality rate of a disabled individual given \mathcal{H}_t^3 (solid) and \mathcal{H}_t^4 (dashed), and the mortality rate from each Markov chain (dotted)

3.4 Conclusions and Discussion

This chapter investigates how the transition intensities of three-state disability processes, modelled using mixtures of Markov chains, depend on past transitions. By considering four assumed histories of being disabled in Example 3.2.1, we investigate a mixture of two simple Markov chains and two more realistic Markov chains in Section 3.3. As a result, the mixture model describes the disability process more practically as behaviours of the transition intensities conditional on their histories reasonably match what we expect.

Also, using Laplace transforms to determine the transition intensities at time t given the histories is quite mathematically convenient. Once the mixing distributions are given, intensity functions are determinable.

Further research is necessary to investigate how to estimate the parameters of the mixture models using an appropriate data set. We understand that this is challenging because of the

difficulties of the parameter identifiability. Also, we require a data set that contains information that goes beyond the observed transitions. That is, we require explanatory variables that reveal more about the health of an individual.

Chapter 4

Mixture of Markov Chains for the Joint-life Model

4.1 Introduction

Since many insurance and annuity products pay benefits that depend on the survival status of a married couple, there is value in studying the joint-life model. Norberg (1989) introduces a four-state continuous-time Markov chain with transition intensities depending on the survival status of two spouses. He shows that the remaining lifetime of an individual will be shorter after the death of his/her spouse.

However, the traditional Markov model for joint-life analysis is limited by the Markov property, for which transition intensities depend only on the current state and time. Hence, the time since the first death does not affect the future lifetime of the survivor.

In fact, the force of mortality for the survivor is dependent on when his/her spouse dies. Jagger and Sutton (1991) show an increased relative risk of mortality during marital bereavement. This phenomenon is known as “broken-heart syndrome”. To understand the nature of the dependence of the lifetimes of a couple, Hougaard (2000) describes three different types of dependence. According to the behaviour of the force of mortality for the survivor conditional

on the time of the spouse's death, he introduces three types of dependence:

1. *instantaneous dependence* is due to the common events that affect both lives at the same time;
2. *short-term dependence* involves an immediate increase to the mortality of the survivor upon the spouse's death with the extra force of mortality diminishing over time;
3. *long-term dependence* is generated by the common risk environment and also involves an immediate increase to the mortality of the survivor, but the extra force of mortality does not diminish over time.

Based on Hougaard's interpretation, Spreeuw (2006) explores the type of dependence produced by single parameter copula models. Furthermore, in order to investigate short-term dependence between the lifetimes of a couple, Spreeuw and Wang (2008) and Spreeuw and Owadally (2013) extend Norberg's work by allowing mortality to vary with the time elapsed since death of a spouse by introducing additional states. Also, Ji, et al (2011) extend Norberg's work by using a semi-Markov model.

In our study, we capture the short-term dependence of a couple by using mixtures of Markov chains. This approach allows the forces of mortality to depend on aspects of the past history of the process.

A summary of the data set we used is introduced in Section 4.2. A simple mixture of two four-state Markov chains and a generalized mixture model represented as a six-state Markov chain with unobservable states are discussed in Section 4.3. In comparison with the six-state Markov chain introduced by Spreeuw and Owadally (2013), the unobservable states in our model allow us to determine the rate of diminishment of the extra force of mortality of the survivor. The expressions for the forces of mortality conditional on the histories for both models are shown in this section as well. Afterwards, the maximum likelihood estimation (MLE) is discussed in Section 4.4. And then, with the data introduced in Section 4.2, Section

4.5 will specify our model assumptions and show the estimation results. Also, a graphical approach is used to examine the fit of the models. Section 4.6 concludes the chapter.

4.2 Data Used in Our Analysis

To explore the dependence of the lifetimes of a couple, we consider the expansion of the four-state Markov chain and use a data set containing mortality data for 805 married couples. The data set is taken from Allin (1988), which provides dates of birth, marriage and death over more than 200 years for individuals in a family tree. While we acknowledge that this kind of genealogical data may not be representative of the general population, we believe it is acceptable for our study of joint lifetimes of spouses. However, the results we obtain may not be applicable for general population.

Each couple includes a male and a female. The age at marriage is known for both. Also, the age at death or censoring is known for both. Censoring occurs if an individual is still alive at the time of data collection. It is assumed that all marriages continued until the earlier of the first death and the time of data collection.

There are 454 couples with both censored during the observation period, 43 males are still alive at the censoring time after the death of their wives and 74 females are still alive at the censoring time after the death of their husbands. For the remaining 234 couples, we have observed both deaths.

Tables 4.1 and 4.2 summarize the data. Each table contains three sub-tables, showing exposure, death counts and crude mortality rates. Table 4.1 summarizes the data for males, while Table 4.2 summarizes the data for females.

In Table 4.1, the first sub-table shows the number of years lived by males in different age ranges and for different timing of the wives' death. The second sub-table shows the total number of deaths in the corresponding groups. Taking the ratio of the number of deaths to the exposure for each group, results in the crude mortality rates given in the third sub-table.

Nine groups are considered here. To investigate the dependence of the lifetimes of a spouse, three status groupings for the female spouses are considered. Specifically, we consider whether the female is still alive, and if the female is no longer alive, has it been five years or not? For these groupings, the exposure and number of deaths of males are presented. For the purpose of investigating the age effects, three groupings are considered. The breakpoints are set at age 60 and age 75.

From the Table 4.1, we see that the rate of mortality for males increases upon their spouses' deaths. However, the rate of mortality more than five years after the spouses' death is lower than the rate of mortality during the first five years after their spouses' deaths. This phenomenon appears in all age ranges. We also observe that the additional mortality that occurs upon the death of the spouse diminishes more slowly at the older ages.

Table 4.1: Summary of male data

Exposure

Age	Spouse alive	Spouse dead < 5 years	Spouse dead \geq 5 years
Under 60	21394.4	161.5	231.9
Between 60 and 75	3762.1	141.1	384.3
Over 75	776.4	77.3	319.3

Deaths

Age	Spouse alive	Spouse dead < 5 years	Spouse dead \geq 5 years
Under 60	70	8	2
Between 60 and 75	101	10	14
Over 75	54	10	39

Crude Mortality Rates

Age	Spouse alive	Spouse dead < 5 years	Spouse dead \geq 5 years
Under 60	0.0033	0.0495	0.0086
Between 60 and 75	0.0268	0.0709	0.0364
Over 75	0.0696	0.1294	0.1221

Table 4.2 shows the corresponding information for females. Comparing Table 4.1 and Table 4.2, we observe that females generally have lower rates of mortality than males. When the survivor is younger than 60, the rates of mortality for males and females are not significantly different if the spouse is still alive. However, once the spouse dies, the rate of mortality for males increases more than for females. This additional rate of mortality for the males drops

more quickly than for females as the duration since the spouse dies increases. If the survivor is over 60 but younger than 75, similar changes in mortality rates are observed. When the survivor is over 75, females have a much lower force of mortality than males before the death of their spouses. However, females experience a larger additional rate of mortality when the spouse dies. For both males and females, the additional rate of mortality decreases slightly after five years since the spouse dies.

Table 4.2: Summary of female data

Exposure

Age	Spouse alive	Spouse dead < 5 years	Spouse dead \geq 5 years
Under 60	22440.4	327.6	470.3
Between 60 and 75	3008.0	311.1	1055.2
Over 75	484.6	143.3	880.9

Deaths

Age	Spouse alive	Spouse dead < 5 years	Spouse dead \geq 5 years
Under 60	61	7	9
Between 60 and 75	47	14	30
Over 75	11	15	83

Crude Mortality Rates

Age	Spouse alive	Spouse dead < 5 years	Spouse dead \geq 5 years
Under 60	0.0027	0.0214	0.0191
Between 60 and 75	0.0156	0.0450	0.0284
Over 75	0.0227	0.1047	0.0942

Our observations from this data set inspire us to explore how the duration of widow(er)hood impacts the force of mortality. Our data suggests that we have short-term dependence in the joint-life model. As we shall see, a mixture of Markov chains can be used to capture this behaviour.

4.3 Mixtures of Markov Chains Approach

An example of the first approach of constructing a mixture model in Section 2.4 is a simple mixture of two four-state Markov chains representing the joint-life model. In the simple mixture model, each Markov chain represents a population. One population may be healthier than

the other. This is reflected by lower forces of mortality. A consequence of this construction is that the force of mortality of an individual diminishes with time since the spouse's death. This is a behaviour that we want to capture with our model. However, this model does not capture any non-Markov behaviour that exists within each population. If we wish to capture the non-Markov behaviour within a single population, we can use a six-state Markov chain with unobservable states, which can be understood as the fourth approach of constructing a mixture model in Section 2.4. This is discussed in Subsection 4.3.2.

4.3.1 Simple Mixture of Two Four-state Markov Chains

According to Norberg's model, the four-state Markov chain for the joint-life model has state space $\{0,1,2,3\}$, where both (x) and (y) are alive when the process is in state 0, (x) is dead but (y) is alive when the process is in state 1, (x) is alive but (y) is dead when the process is in state 2, and both (x) and (y) are dead when the process is in the absorbing state 3. Therefore, the transition intensities $\mu^{01}(t)$ and $\mu^{23}(t)$ are essentially the forces of mortality for (x) , while the transition intensities $\mu^{02}(t)$ and $\mu^{13}(t)$ are essentially the forces of mortality for (y) . If a four-state Markov chain has dependent lifetimes, $\mu^{13}(t)$ and $\mu^{23}(t)$ are expected to be greater than $\mu^{02}(t)$ and $\mu^{01}(t)$, respectively. And this difference in the forces of mortality reflects the additional force of mortality for the survivor after the spouse dies.

One four-state Markov chain does not capture the diminishment of additional mortality that is normally observed. In other words, this model involves long-term dependence, whereas we also expect short-term dependence. We can achieve this behaviour by using a simple mixture of two continuous time four-state Markov chains. Consider such a mixture, as depicted in Figure 4.1. Let $\mu_m^{ij}(t)$ be the transition intensity from state i to state j at time t for Markov chain $M = m$, with $m = 1, 2$. Therefore, $\mu_m^{01}(t)$ and $\mu_m^{23}(t)$ are the forces of mortality for (x) at age $x + t$; $\mu_m^{02}(t)$ and $\mu_m^{13}(t)$ are the forces of mortality for (y) at age $y + t$.

The simple mixture of two Markov chains provides a model for a heterogeneous population. It allows us to represent each of the two sub-populations by a four-state Markov chain. For

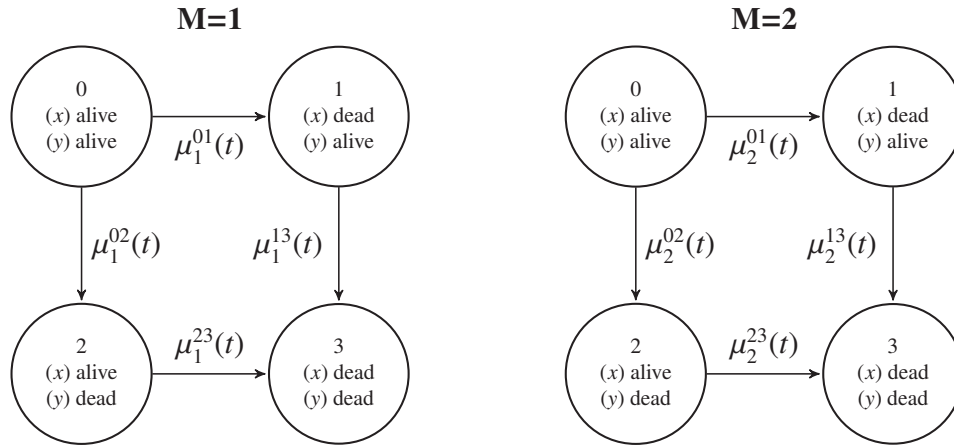


Figure 4.1: Four-state with mixture model

example, assume that $M = 1$ and $M = 2$ are the Markov chains with independent lifetimes and dependent lifetimes, respectively, where the forces of mortality before the first death are the same for $M = 1$ and $M = 2$. For $M = 1$, the force of mortality for each individual is not affected by the death of the spouse. On the contrary, the forces of mortality increase after the spouse dies if $M = 2$. Therefore, given the second death happens at an older age, there is a higher probability that $M = 1$.

However, the lifetimes for a couple are rarely independent in reality. In fact, the mixture of two Markov chains where both Markov chains involve dependent lifetimes is more practical and this will be assumed in the remainder of the chapter.

Let $M = 1$ and $M = 2$ be Markov chains with dependent lifetimes, where $\mu_1^{01}(t) = \mu_2^{01}(t)$ and $\mu_1^{02}(t) = \mu_2^{02}(t)$. Suppose that $M = 1$ represents a healthier population, which means there is less additional force of mortality for the survivor after the spouse dies.

4.3.2 A Six-state Markov Model with Unobservable States

Another approach to model the short-term dependence of joint lifetimes involves a six-state Markov chain as shown in Figure 4.2. The process is in state 0 when both (x) and (y) are alive. Upon death of (x) or (y) , the process enters state 1 or 3 respectively. In these states, the force of mortality of the survivor is higher than the force for that individual while in state 0. However,

the increased force of mortality will diminish over time, and this phenomenon is captured by allowing transitions from state 1 to state 2 and from state 3 to state 4, where states 2 and 4 have lower forces of mortality than states 1 and 3. Note that we cannot distinguish between state 1 and state 2 or between state 3 and state 4. The rate at which the force of mortality diminishes after the first death is determined by the constant transition rate from state 1 to state 2 or from state 3 to state 4.

In Figure 4.2, $\mu_x^{ij}(t)$ and $\mu_y^{ij}(t)$ represent the transition intensities from state i to state j for the corresponding individual. Transition intensities with a subscript x represent forces of mortality of (x) , and transition intensities with a subscript y represent forces of mortality of (y) .

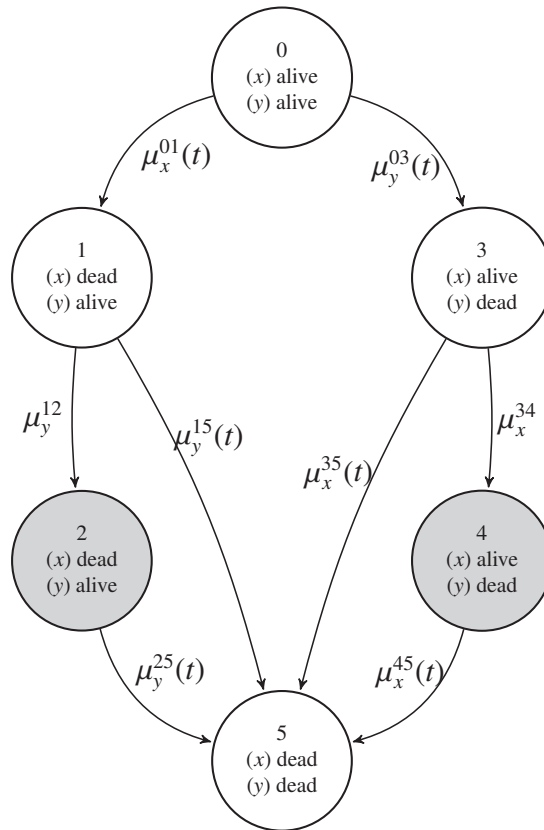


Figure 4.2: Six-state model

This six-state model can actually be represented as a more general mixture of two Markov chains. It follows the fourth approach in Section 2.4. The whole population is assumed to start from an initial Markov chain with a greater bereavement effect after the first death. At some

time after the first death happens, the process may move to a Markov chain with a lesser be-reavement effect. The possibility of this transition captures the diminishment of the additional force of mortality for the surviving spouse within the population as a whole. Therefore, there is a process $\{M(t), t \geq 0\}$ that indicates which Markov chain is applicable at each point in time. In addition, given an applicable history of survivorship, the probability of being in the initial Markov chain is decreasing with the time since the first death happens. Therefore, we observe the additional force of mortality for the survivor to diminish.

As shown in Figure 4.3, suppose Markov chain 1 has state space $\{0,1,3,5\}$ and Markov chain 2 has state space $\{0,2,4,6\}$. State 0 in both Markov chains is the state in which both (x) and (y) are alive, and state 5 and state 6 in the two Markov chains are the states in which both (x) and (y) are dead. States 1 and 2 are the states in which (x) is dead while (y) is still alive, and states 3 and 4 are the states in which (y) is dead while (x) is still alive. In this model, the initial Markov chain has more additional force of mortality after the spouse dies. The transition from Markov chain 1 to Markov chain 2 is only allowed after the first death happens. And the resulting history dependent mixing probability is associated with μ_y^{12} and μ_x^{34} , which are the rates at which the additional mortality diminishes.

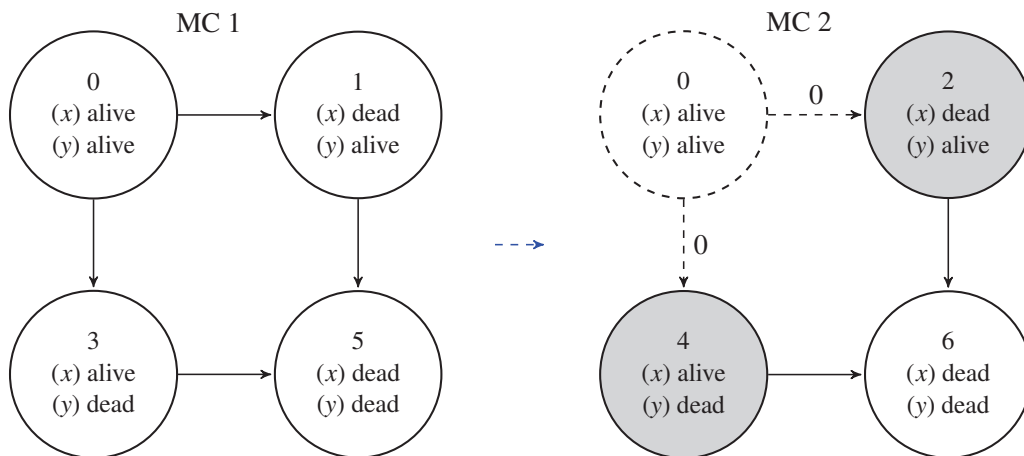


Figure 4.3: Mixture version of the six-state model

Although the six-state model introduces some computational complexity, the presence of the two additional states allows us to capture the short-term dependence between two spouses'

lifetimes.

Notice that our six-state model is similar to the model constructed by Spreeuw and Owadally (2013). However, our approach estimates the rate of transition between the two survivor states, whereas Spreeuw and Owadally (2013) estimates the time of this transition.

4.3.3 Transition Intensities

In many applications, the transition probabilities for a Markov process $X(t)$ are defined as ${}_t p_x^{ij} = Pr(X(x+t) = j | X(x) = i)$, and the transition intensities are given by $\mu_x^{ij} = \lim_{h \rightarrow 0^+} \frac{{}_t p_x^{ij}}{h}$ and $i \neq j = 0, 1, 2, 3$.

It is assumed that (1) the probability of two or more transitions in a small interval h is $o(h)$, and (2) ${}_t p_x^{ij}$ is a differentiable function of t . This leads to

$${}_t p_{x,y}^{00} = \exp\left(-\int_0^t \mu_x^{01}(s) + \mu_y^{02}(s) ds\right), \quad (4.1)$$

$${}_t p_{x,y}^{11} = \exp\left(-\int_0^t \mu_y^{13}(s) ds\right), \quad (4.2)$$

$${}_t p_{x,y}^{22} = \exp\left(-\int_0^t \mu_x^{23}(s) ds\right), \quad (4.3)$$

$$\text{and } {}_t p_{x,y}^{01} = \exp\left(-\int_0^t {}_s p_{x,y}^{00} \mu_x^{01}(s) {}_{t-s} p_x^{11} ds\right) \quad (4.4)$$

in the four-state Markov chain.

To illustrate how the time of first death will influence the force of mortality for the survivor, we consider the three histories of survivorship represented by $\mathcal{H}_t^k, k = 1, 2, 3$:

- \mathcal{H}_t^1 indicates that both (x) and (y) are alive at time t .
- \mathcal{H}_t^2 indicates that (y) dies before (x) at time t_y , (x) is still alive at t , $0 < t_y < t$.
- \mathcal{H}_t^3 indicates that (x) dies before (y) at time t_x , (y) is still alive at t , $0 < t_x < t$.

Notice that the history we refer to in the remainder of this chapter is the history of survivorship, whereas the history of the process involves unobservable information.

In order to glean information about the dependence of the two lifetimes, we can explore the behaviour of the conditional forces of mortality given applicable histories. Suppose the history dependent forces of mortality are denoted by $\lambda_x(t | \mathcal{H}_t^k)$ and $\lambda_y(t | \mathcal{H}_t^k)$ for $k = 1, 2, 3$. Obviously, conditional on \mathcal{H}_t^1 , we can determine the force of mortality for both (x) and (y). However, conditional on \mathcal{H}_t^2 , we can investigate only the force of mortality for (x), and conditional on \mathcal{H}_t^3 , we can investigate only the force of mortality for (y).

Simple Mixture of Two Four-state Markov chains

For a simple mixture of two Markov chains both with state space $\{0, 1, 2, 3\}$, we have:

$$\begin{aligned}
 \lambda_x(t | \mathcal{H}_t^1) &= \mu_1^{01}(x+t)P(M=1 | \mathcal{H}_t^1) + \mu_2^{01}(x+t)P(M=2 | \mathcal{H}_t^1), \\
 \lambda_y(t | \mathcal{H}_t^1) &= \mu_1^{02}(y+t)P(M=1 | \mathcal{H}_t^1) + \mu_2^{02}(y+t)P(M=2 | \mathcal{H}_t^1), \\
 \lambda_x(t | \mathcal{H}_t^2) &= \mu_1^{13}(x+t)P(M=1 | \mathcal{H}_t^2) + \mu_2^{13}(x+t)P(M=2 | \mathcal{H}_t^2), \\
 \lambda_y(t | \mathcal{H}_t^3) &= \mu_1^{23}(y+t)P(M=1 | \mathcal{H}_t^3) + \mu_2^{23}(y+t)P(M=2 | \mathcal{H}_t^3),
 \end{aligned} \tag{4.5}$$

and by the Bayes' rule

$$P(M = m | \mathcal{H}_t^k) = \frac{P(M = m)P(\mathcal{H}_t^k | M = m)}{P(\mathcal{H}_t^k)},$$

where

$$\begin{aligned}
P(\mathcal{H}_t^1 | M = m) &= {}^m_t P_{x,y}^{00}, \\
P(\mathcal{H}_t^2 | M = m) &= {}^m_{t_y} P_{x,y}^{00} \mu_m^{02}(y + t_y) {}^m_{t-t_y} P_{x+t_y}^{22}, \\
P(\mathcal{H}_t^3 | M = m) &= {}^m_{t_x} P_{x,y}^{00} \mu_m^{01}(x + t_x) {}^m_{t-t_x} P_{y+t_x}^{11},
\end{aligned}$$

and for a specified m , ${}^m_t P_{x,y}^{ii}$ has the same expression as ${}_t P_{x,y}^{ii}$ in Equations 4.1, 4.2, 4.3, and 4.4.

Note that all probabilities of histories shown in this thesis are actually probability densities even though probability notation is used.

Six-state Model

In the six-state model, given \mathcal{H}_t^1 , $X(t) \in \{0\}$; given \mathcal{H}_t^2 , $X(t) \in \{3, 4\}$; given \mathcal{H}_t^3 , $X(t) \in \{1, 2\}$.

Here, we determine $\lambda_x(t | \mathcal{H}_t^2)$ as an example. A similar technique can be used to determine the other λ s.

We have

$$\begin{aligned}
\lambda_x(t | \mathcal{H}_t^2) &= \mu_x^{35}(t) P(X(t) = 3 | \mathcal{H}_t^2) + \mu_x^{45}(t) P(X(t) = 4 | \mathcal{H}_t^2) \\
&= \mu_x^{35}(t) \frac{P(X(t) = 3, \mathcal{H}_t^2)}{P(\mathcal{H}_t^2)} + \mu_x^{45}(t) \frac{P(X(t) = 4, \mathcal{H}_t^2)}{P(\mathcal{H}_t^2)},
\end{aligned}$$

as

$$P(\mathcal{H}_t^2) = P(X(t) = 3, \mathcal{H}_t^2) + P(X(t) = 4, \mathcal{H}_t^2),$$

therefore,

$$\lambda_x(t | \mathcal{H}_t^2) = \frac{\mu_x^{35}(t) P(X(t) = 3, \mathcal{H}_t^2) + \mu_x^{45}(t) P(X(t) = 4, \mathcal{H}_t^2)}{P(X(t) = 3, \mathcal{H}_t^2) + P(X(t) = 4, \mathcal{H}_t^2)},$$

where

$$\begin{aligned} P(X(t) = 3, \mathcal{H}_t^2) &= {}_{t_y}P_{x,y}^{00} \mu_{y+t_y}^{03} {}_{t-t_y}P_{x+t_y}^{33} \\ &= \exp\left(-\int_0^{t_y} (\mu_x^{01}(s) + \mu_y^{03}(s))ds\right) \mu_{y+t_y}^{03} \exp\left(-\int_0^{t-t_y} (\mu_{x+t_y}^{35}(s) + \mu^{34})ds\right) \end{aligned}$$

and

$$P(X(t) = 4, \mathcal{H}_t^2) = {}_{t_y}P_{x,y}^{00} \mu_{y+t_y}^{03} \int_0^{t-t_y} {}_sP_{x+t_y}^{33} \mu_{x+t_y+s}^{34} {}_{t-t_y-s}P_{x+t_y+s}^{44} ds.$$

Therefore,

$$\lambda_x(t | \mathcal{H}_t^2) = \frac{\mu_x^{35}(t) {}_{t-t_y}P_{x+t_y}^{33} + \mu_x^{45}(t) \int_0^{t-t_y} {}_sP_{x+t_y}^{33} \mu_{x+t_y+s}^{34} {}_{t-t_y-s}P_{x+t_y+s}^{44} ds}{{}_{t-t_y}P_{x+t_y}^{33} + \int_0^{t-t_y} {}_sP_{x+t_y}^{33} \mu_{x+t_y+s}^{34} {}_{t-t_y-s}P_{x+t_y+s}^{44} ds}.$$

As mentioned in Section 4.3.2, the six-state Markov chain with unobservable states can also be represented as a generalized mixture of two four-state Markov chains, with a transition allowed from Markov chain 1 with state space $\{0,1,3,5\}$ to Markov chain 2 with state space $\{0,2,4,6\}$ after the first death happens. Clearly, if $X(t) \in \{1, 3\}$, then $M(t) = 1$, while if $X(t) \in \{2, 4\}$, then $M(t) = 2$. In this model representation, the probability of being in Markov chain 1 is 1 if both spouses are alive, and it decreases with time since the first death according to the diminish rate. As a result, the transition between Markov chain 1 and Markov chain 2 is represented by the diminish rate of additional force of mortality for the corresponding surviving spouse, given the applicable history of survivorship.

We also determine $\lambda_x(t | \mathcal{H}_t^2)$ as an example. Analogous to Equation 4.5 for the simple mixture of two Markov chains, we have

$$\begin{aligned} \lambda_x(t | \mathcal{H}_t^2) &= \mu_x^{35}(t)P(M(t) = 1 | \mathcal{H}_t^2) + \mu_x^{45}(t)P(M(t) = 2 | \mathcal{H}_t^2) \\ &= \frac{\mu_x^{35}(t)P(M(t) = 1, \mathcal{H}_t^2) + \mu_x^{45}(t)P(M(t) = 2, \mathcal{H}_t^2)}{P(M(t) = 1, \mathcal{H}_t^2) + P(M(t) = 2, \mathcal{H}_t^2)}, \end{aligned}$$

where

$$\begin{aligned}
P(M(t) = 1, \mathcal{H}_t^2) &= {}_{t_y}P_{x,y}^{00} \mu_{y+t_y}^{03} {}_{t-t_y}P_{x+t_y}^{33} \exp(-\mu^{34}(t-t_y)) \\
&= \exp\left(-\int_0^{t_y} \mu_x^{01}(s) + \mu_y^{02}(s) ds\right) \mu_{y+t_y}^{03} \exp\left(-\int_0^{t-t_y} \mu_x^{35}(t_y+s) ds\right) \exp(-\mu_x^{34}(t-t_y))
\end{aligned}$$

and

$$P(M(t) = 2, \mathcal{H}_t^2) = {}_{t_y}P_{x,y}^{00} \mu_{y+t_y}^{03} \int_{t_y}^t {}_{s-t_y}P_{x+t_y}^{33} \mu_{x+t_y+s}^{34} {}_{t-s}P_{x+t_y+s}^{44} ds.$$

Notice that $P(M(t) = 1, \mathcal{H}_t^2)$ has the same expression as $P(X(t) = 3, \mathcal{H}_t^2)$ in the six-state model, and $P(M(t) = 2, \mathcal{H}_t^2)$ has the same expression as $P(X(t) = 4, \mathcal{H}_t^2)$ in the six-state model. Therefore, the six-state model can also be represented as a generalized mixture of two Markov chains, where a generalized mixture of Markov chains involves a stochastic process which indicates the Markov chain we are in and path-dependent mixing probabilities.

4.4 Estimation Method

We used the maximum likelihood estimation (MLE) method to estimate the parameters of interest. To construct the likelihood function, five cases are considered. Suppose the time of stopping observation (i.e. censoring time) is denoted by t_s . With c indicating the couple, δ_x^c and δ_y^c are the indicators of whether (x) and (y) are dead or alive at time t_s for the c^{th} couple. For example, $\delta_x^3 = 1$ and $\delta_y^3 = 0$ indicates the living status of the third couple: (x) dies before the censoring time t_s while (y) lives through t_s . Let t_x^c and t_y^c be the observed survival time since marriage for (x) and (y) of couple c . Also, let T_x^c and T_y^c be the random variables of the survival time. We will discuss the likelihood function for both the simple mixture of two Markov chains and the six-state model. Let the likelihood contribution for couple c be \mathcal{L}_c .

Therefore, the likelihood function is

$$\mathcal{L} = \prod_c \mathcal{L}_c.$$

Particularly, the likelihood contribution for couple c can be expressed as

$$\begin{aligned} \mathcal{L}_c &= Pr(T_x^c \geq t, T_y^c \geq t)^{(1-\delta_x^c)(1-\delta_y^c)} Pr(t_x^c < T_x^c < t_x^c + \Delta < t_s \leq T_y^c)^{\delta_x^c(1-\delta_y^c)} \times \\ &Pr(t_y^c < T_y^c < t_y^c + \Delta < t_s < T_x^c)^{(1-\delta_x^c)\delta_y^c} \times \\ &Pr(t_x^c < T_x^c < t_x^c + \Delta < t_y^c < T_y^c < t_y^c + \Delta)^{\delta_x^c \delta_y^c I(t_x^c < t_y^c)} \times \\ &Pr(t_y^c < T_y^c < t_y^c + \Delta < t_x^c < T_x^c < t_x^c + \Delta)^{\delta_x^c \delta_y^c I(t_y^c < t_x^c)}, \end{aligned}$$

where Δ is arbitrarily small.

For the purpose of simplifying the notation, we drop the superscription c for t_x^c , t_y^c , δ_x^c , and δ_y^c , and $t = t_s$ in the following expressions.

For the four-state Markov chain, the likelihood for couple c is then

$$\begin{aligned} \mathcal{L}_c &\propto \left({}_t P_{x,y}^{00}\right)^{(1-\delta_x)(1-\delta_y)} \times \left({}_{t_x} P_{x,y}^{00} \mu_{x+t_x}^{01} \quad {}_{t-t_x} P_{y+t_x}^{11}\right)^{\delta_x(1-\delta_y)} \times \left({}_{t_y} P_{x,y}^{00} \mu_{y+t_y}^{02} \quad {}_{t-t_y} P_{x+t_y}^{22}\right)^{(1-\delta_x)\delta_y} \times \\ &\left({}_{t_x} P_{x,y}^{00} \mu_{x+t_x}^{01} \quad {}_{t-t_x} P_{y+t_x}^{11} \mu_{y+t}^{13}\right)^{\delta_x \delta_y I(t_x < t_y)} \times \left({}_{t_y} P_{x,y}^{00} \mu_{y+t_y}^{02} \quad {}_{t-t_y} P_{x+t_y}^{22} \mu_{x+t}^{23}\right)^{\delta_x \delta_y I(t_y < t_x)} \end{aligned} \quad (4.6)$$

For the mixture models we are discussing in this chapter, by considering the histories, the possible of the likelihood contribution for couple c are as follows.

1. Both (x) and (y) are alive. (Notice this is the history \mathcal{H}_t^1 mentioned in Section 4.3.3.) In this case, we have $\delta_x = 0, \delta_y = 0, t_x = t_y = t$.

- With the notations introduced in Section 4.3.3, the probability density for the mixture of two four-state Markov chains is

$$Pr(T_x \geq t, T_y \geq t) = {}_t^1 P_{x,y}^{00} P(M = 1) + {}_t^2 P_{x,y}^{00} P(M = 2).$$

- The probability density for the six-state Model for this history is

$$Pr(T_x \geq t, T_y \geq t) = {}_t p_{x,y}^{00}.$$

2. (x) dies, (y) lives through t . (Notice this is the history \mathcal{H}_t^3 mentioned in Section 4.3.3).

In this case, we have $\delta_x = 1, \delta_y = 0, t_x < t_y = t$.

- The probability density for the mixture of two four-state Markov chains for this history is

$$\begin{aligned} \lim_{\Delta \rightarrow 0^+} \frac{1}{\Delta} Pr(t_x < T_x < t_x + \Delta < t \leq T_y) &= {}_{t_x}^1 p_{x,y}^{00} \mu_1^{01}(x + t_x) {}_{t-t_x}^1 p_{y+t_x}^{11} P(M = 1) + \\ & {}_{t_x}^2 p_{x,y}^{00} \mu_2^{01}(x + t_x) {}_{t-t_x}^2 p_{y+t_x}^{11} P(M = 2). \end{aligned}$$

- The probability density for the six-state Model with this history is

$$\begin{aligned} \lim_{\Delta \rightarrow 0^+} \frac{1}{\Delta} Pr(t_x < T_x < t_x + \Delta < t \leq T_y) &= \\ & {}_{t_x} p_{x,y}^{00} \mu_{x+t_x}^{01} \left({}_{t-t_x} p_{y+t_x}^{11} + \int_0^{t-t_x} {}_s p_{y+t_x}^{11} \mu_{y+t_x+s}^{12} {}_{t-t_x-s} p_{y+t_x+s}^{22} ds \right). \end{aligned}$$

3. (y) dies, (x) lives through t . (Notice this is the history \mathcal{H}_t^2 mentioned in Section 4.3.3).

In this case, we have $\delta_x = 0, \delta_y = 1, t_y < t_x = t$.

- The probability density for the mixture of two four-state Markov chains for this history is

$$\begin{aligned} \lim_{\Delta \rightarrow 0^+} \frac{1}{\Delta} Pr(t_y < T_y < t_y + \Delta < t \leq T_x) &= \\ & {}_{t_y}^1 p_{x,y}^{00} \mu_1^{02}(y + t_y) {}_{t-t_y}^1 p_{x+t_y}^{22} P(M = 1) + {}_{t_y}^2 p_{x,y}^{00} \mu_2^{02}(y + t_y) {}_{t-t_y}^2 p_{x+t_y}^{22} P(M = 2). \end{aligned}$$

- The probability density for the six-state Model with this history is

$$\lim_{\Delta \rightarrow 0^+} \frac{1}{\Delta} Pr(t_y < T_y < t_y + \Delta < t \leq T_x) =$$

$${}_{t_y}p_{x,y}^{00} \mu_{y+t_y}^{03} \left({}_{t-t_y}p_{x+t_y}^{33} + \int_0^{t-t_y} {}_s p_{x+t_y}^{33} \mu_{x+t_y+s}^{34} {}_{t-t_y-s} p_{x+t_y+s}^{44} ds \right).$$

4. Both are dead at t , (x) dies before (y). We refer to this as $\mathcal{H}_t^4 : \delta_x = 1, \delta_y = 1, t_x < t_y < t$

- The probability density for the mixture of two four-state Markov chains for this history is

$$\lim_{\Delta \rightarrow 0^+} \frac{1}{\Delta^2} Pr(t_x < T_x < t_x + \Delta < t_y \leq T_y < t_y + \Delta) =$$

$${}_{t_x}^1 p_{x,y}^{00} \mu_1^{01}(x + t_x) {}_{t_y-t_x}^1 p_{y+t_x}^{11} \mu_1^{13}(y + t_y) P(M = 1) +$$

$${}_{t_x}^2 p_{x,y}^{00} \mu_2^{01}(x + t_x) {}_{t_y-t_x}^2 p_{y+t_x}^{11} \mu_2^{13}(y + t_y) P(M = 2).$$

- The probability density for the six-state Model with this history is

$$\lim_{\Delta \rightarrow 0^+} \frac{1}{\Delta^2} Pr(t_x < T_x < t_x + \Delta < t_y \leq T_y < t_y + \Delta) =$$

$${}_{t_x} p_{x,y}^{00} \mu_{x+t_x}^{01} \left({}_{t_y-t_x} p_{y+t_x}^{11} \mu_{y+t_y}^{15} + \int_0^{t_y-t_x} {}_s p_{y+t_x}^{11} \mu_{y+t_x+s}^{12} {}_{t-t_x-s} p_{y+t_x+s}^{22} \mu_{y+t_y}^{25} ds \right).$$

5. Both are dead at t , (y) dies before (x). We refer to this as $\mathcal{H}_t^5 : \delta_x = 1, \delta_y = 1, t_y < t_x < t$.

- The probability density for the mixture of two four-state Markov chains for this history is

$$\lim_{\Delta \rightarrow 0^+} \frac{1}{\Delta^2} Pr(t_y < T_y < t_y + \Delta < t_x \leq T_x < t_x + \Delta) =$$

$${}_{t_y}^1 p_{x,y}^{00} \mu_1^{02}(y + t_y) {}_{t_x-t_y}^1 p_{x+t_y}^{22} \mu_1^{23}(x + t_x) P(M = 1) +$$

$${}_{t_y}^2 p_{x,y}^{00} \mu_2^{02}(y + t_y) {}_{t_x-t_y}^2 p_{x+t_y}^{22} \mu_2^{23}(x + t_x) P(M = 2).$$

- The probability density for the six-state Model with this history is

$$\lim_{\Delta \rightarrow 0^+} \frac{1}{\Delta^2} Pr(t_y < T_y < t_y + \Delta < t_x \leq T_x < t_x + \Delta) =$$

$${}_{t_y}P_{x,y}^{00} \mu_{y+t_y}^{03} \left({}_{t_x-t_y}P_{x+t_y}^{33} \mu_{x+t_x}^{35} + \int_0^{t_x-t_y} {}_sP_{x+t_y}^{33} \mu_{x+t_y+s}^{34} {}_{t-t_y-s}P_{x+t_y+s}^{44} \mu_{x+t_x}^{45} ds \right).$$

Therefore, the likelihood contribution for couple c can be determined by taking the product of the probability densities corresponding to the five cases above with each raised to the appropriate indicator power as in expression (4.6).

4.5 Fitting and Checking the Model

4.5.1 Specific Model Assumptions

We start this section by introducing some specific assumptions about the forms of the transition intensities. In the mixture of two four-state Markov chains and the six-state Markov chain with unobservable states, we assume the force of mortality for each individual follows Gompertz' law with a parameter for mortality improvement. This is appropriate since Gompertz distribution reasonably captures human mortality over a broad range of ages. Let (x) refer to the male and (y) refer to the female. For i and j representing mortality transitions of (x) , let the intensities be $\mu_m^{ij}(t) = \alpha_m^{ij} \exp(\beta^H(x+t) - \gamma v_x)$, and for i and j representing mortality transitions of (y) , let the intensities be $\mu_m^{ij}(t) = \alpha_m^{ij} \exp(\beta^W(y+t) - \gamma v_y)$. In these expressions, α_m^{ij} , β^H , β^W , and γ are parameters, and v_x and v_y are the times of birth of (x) and (y) respectively. (Note that the H and W in β^H and β^W indicate husband and wife respectively.)

Notice that in the mixture of two Markov chains, only α_m^{ij} varies between Markov chains, but β^H , β^W and γ are assumed to be the same for both Markov chains. In other words, the forces of mortality in the two Markov chains are proportional.

For the simple mixture of two four-state Markov chains, consider $M = 1$ to be a healthier Markov chain, which is assumed to have dependent lifetimes with lower forces of mortality compared to the frailer Markov chain $M = 2$. In addition, we assume $\mu_1^{01}(t) = \mu_2^{01}(t)$ and

$$\mu_1^{02}(t) = \mu_2^{02}(t).$$

For the six-state Markov chain with unobservable states, we assume that the additional force of mortality for the male will diminish completely, in the other words, $\mu_x^{45}(t) = \mu_x^{01}(t)$. We check this assumption later and find it to be reasonable.

Also, we let the diminish rates μ_y^{12} and μ_x^{34} be constant for simplicity.

4.5.2 Estimation Results

Our parameter estimates are shown in Tables 4.3 and 4.4.

Table 4.3 shows the estimates of parameters in the simple mixture of two four-state Markov chains, both involving dependent lifetimes. We observe that $\hat{\beta}^H$ is greater than $\hat{\beta}^W$, which indicates that the force of mortality for males is increasing faster than that for females. For males, $\hat{\alpha}_1^{01}$ ($= \hat{\alpha}_2^{01}$) is greater than $\hat{\alpha}_1^{23}$ but smaller than $\hat{\alpha}_2^{23}$. The fact is that, for a mixture of Markov chains, the force of mortality for the male after his wife dies depends on how long both of them are alive. The longer they live, the higher is the conditional probability that the couple is in the healthier population. For females, $\hat{\alpha}_1^{02}$ ($= \hat{\alpha}_2^{02}$) is smaller than $\hat{\alpha}_1^{13}$ and $\hat{\alpha}_2^{13}$. The latter two estimates are almost identical, and this is consistent with Table 4.2, which shows that the crude mortality rates for females do not decrease very much after the increase that occurs when their spouses die.

Table 4.3: Estimates from mixture of two four-state Markov chains

	$\hat{\alpha}_m^{01}$	$\hat{\alpha}_m^{23}$	$\hat{\alpha}_m^{02}$	$\hat{\alpha}_m^{13}$	$\hat{\beta}^H$	$\hat{\beta}^W$	$\hat{\gamma}$	$\hat{P}(M = m)$
m=1	0.000479	0.000333	0.001328	0.002964	0.083678	0.061216	0.010266	0.452159
m=2	0.000479	0.001283	0.001328	0.002964	0.083678	0.061216	0.010266	0.547841

The shaded values are estimates for the parameters that are assumed to be the same for both Markov chains.

Table 4.4 shows the estimates of parameters in the six-state Markov chain with unobservable states. With the assumption that the additional force of mortality for males will diminish

completely, we get results consistent with Table 4.3.

As in the mixture of two four-state Markov chains, $\hat{\beta}^H$ is greater than $\hat{\beta}^W$. In fact, these parameter estimates are very close to those in Table 4.3. The estimate $\hat{\alpha}^{15}$ represents the constant factor for the higher force of mortality for the female upon the death of the male. Since this estimate is greater than $\hat{\alpha}^{25}$, we achieve a diminishment of the additional force of mortality with time since the male's death. We observe the same relationship between $\hat{\alpha}^{35}$ and $\hat{\alpha}^{45}$. Note that $\hat{\alpha}^{45}$ is constrained to equal $\hat{\alpha}^{01}$, since we observe no significant difference between these parameters. This means that the additional mortality experienced by males when their spouses die diminishes completely.

Notice that $\hat{\gamma}$ is approximately 0.01 in both Tables 4.3 and 4.4, indicating that mortality improved by about 1% per year.

Table 4.4: Estimates from six-state Markov chain with unobservable states

$\hat{\alpha}^{01}$	$\hat{\alpha}^{45}$	$\hat{\alpha}^{35}$	$\hat{\alpha}^{03}$	$\hat{\alpha}^{25}$	$\hat{\alpha}^{15}$
0.000484	0.000484	0.001469	0.001175	0.002494	0.007630
$\hat{\mu}_y^{12}$	$\hat{\mu}_x^{34}$	$\hat{\beta}^H$	$\hat{\beta}^W$	$\hat{\gamma}$	
2.496666	0.341883	0.082385	0.062021	0.009812	

$$\alpha^{01} = \alpha^{45}.$$

We find that there is significant variability associated with some parameter estimates, this is due both to the size of the data set and the rather flat likelihood that results when we have unobservable states in our model.

4.5.3 Model Testing

Since Neyman and Pearson (1933) proposed the likelihood ratio test, it has become one of the most popular methods for testing restrictions on the parameters of a model. Before concluding the model assumptions in Section 4.5.1, we implement several likelihood ratio tests to make decision about parameter restrictions. In this section, we will discuss three of these likelihood

ratio tests.

To check the reasonableness of using Gompertz' law for the forces of mortality, we first adjust for the fact that we have different force of mortality for each year of birth. Specifically, our model forces of mortality are proportional Gompertz forces of mortality for different years of birth, and

$$\text{Force of mortality for year of birth } v = (\text{Force of mortality for year of birth } 0) e^{-\gamma v}.$$

This is a proportional hazards model with a Gompertz baseline hazard function. Therefore, we can check the Gompertz assumption by comparing this fitted baseline hazard function with estimates of the baseline hazard function obtained using a Cox proportional hazards model. According to this model, the hazard function (force of mortality) $\lambda(t)$ for the failure time random variable T of an individual with a covariate v is $\lambda(t | v) = \lambda_0(t) \exp(\phi v)$, where $\lambda_0(t)$ is an unspecified baseline hazard function and ϕ is a regression coefficient.

After imposing the parameter restrictions that result from our likelihood ratio tests and checking the Gompertz assumption, we compare the simple mixture of two four-state Markov chains to the six-state model by adopting one couple as an example.

Likelihood Ratio Tests

For both the simple mixture of two Markov chains and the six-state model, the first concern is whether β^H should be the same as β^W , Therefore we perform a likelihood ratio test based on the null hypothesis:

$$H_0: \beta^H = \beta^W.$$

The asymptotic distribution for likelihood ratio statistic $\Lambda = -2(\ell(\theta | H_0) - \ell(\theta | H_1))$ is well-known as a limiting central chi-square distribution under the H_0 .

For the six-state model, the log-likelihood function under the null hypothesis model equals -2642.396, while the log-likelihood function under the alternative hypothesis is -2637.173,

therefore, the p-value is 0.0012, assuming 1 degree of freedom.

For the simple mixture of two Markov chains, the log-likelihood function under the null hypothesis model equals -2645.936, while the log-likelihood function under the alternative hypothesis is -2640.041, therefore, the p-value is 0.0006.

Thus, both tests show that there is a significant difference between β^H and β^W for both models.

Then, with $\beta^H \neq \beta^W$, the second concern is whether x and y should share the same mortality improvement factor, in the other words,

$$H_0: \gamma^H = \gamma^W.$$

For the six-state model, the likelihood ratio test has p-value 0.0679. For the simple mixture of two Markov chains, the likelihood ratio test has p-value 0.0041. Thus, under a 5% significance level, we cannot reject that $\gamma^H = \gamma^W$ for the six-state model. However, there is significant evidence to suggest that x and y should have different mortality improvement factors, γ , for the simple mixture of two Markov chains. In order to implement the comparison of the six-state model and the simple mixture of two Markov chains, the assumptions need to be consistent. For simplicity, we have assumed $\gamma^H = \gamma^W$ for both models.

Therefore, with assumptions of $\beta^H \neq \beta^W$ and $\gamma^H = \gamma^W = \gamma$, the third concern is unique for the six-state model—whether the additional force of mortality for males diminishes completely or not. The null hypothesis is

$$H_0: \alpha^{01} = \alpha^{45}.$$

The likelihood ratio test has p-value equals to 0.9189. This shows no significant difference between α^{01} and α^{45} . Therefore, we assume that the additional force of mortality for males diminishes completely in the six-state model.

Behaviour of Integrated Force of Mortality Checking

Since the forces of mortality in our models are proportional, we can check their behaviours using the Cox proportional hazards model with a covariate to reflect the mortality improvement. Let $\widehat{H}_0^{ij}(t) \exp(\hat{\phi}v)$ represent the fitted Cox proportional hazard model for the integrated forces of mortality. We can compare $\widehat{H}_0^{ij}(t)$ with $\int_0^t \lambda^{ij}(s | \mathcal{H}_s) ds$ from our fitted Markov models. This is shown in Figure 4.4. The black curves and the grey solid curves represent the integrated forces of mortality. The grey solid curves represent the $\widehat{H}_0^{ij}(t)$, the black solid curves represent the integrated forces of mortality from the six-state model, and the black dashed curves represent the integrated forces of mortality from the simple mixture of two Markov chains. The two grey dashed curves in each panel show a 95% confidence interval for $H_0^{ij}(t)$.

The figure shows that the fit is better after the spouse dies in both models, and the best fit is for females after the spouse dies. The six-state model provides a better fit for the males with spouse dead. In the other cases, the fits of the two models are quite similar.

Comparison of The Two Models

In order to compare how the two models capture the impact of the time since first death, one couple is taken as an example. With different assumptions about the first death, the behaviour of the force of mortality of the surviving spouse is displayed in Figure 4.5. The three panels on the top show the forces of mortality for the male as a function of the time since the female died. Similarly, the three panels at the bottom show the forces of mortality for the female as a function of the time since the male died. The black curves show forces of mortality based on the fitted six-state model and the grey curves show forces of mortality based on the fitted mixture of two Markov chains. The dashed curves, which represent the forces of mortality while the spouse is alive, are included for comparison.

From Figure 4.5, we observe that the force of mortality for the survivor is highest immediately after the spouse dies and decreases with the time since this death happened. As the solid curves are approaching the dashed curves when duration increases, we see how much

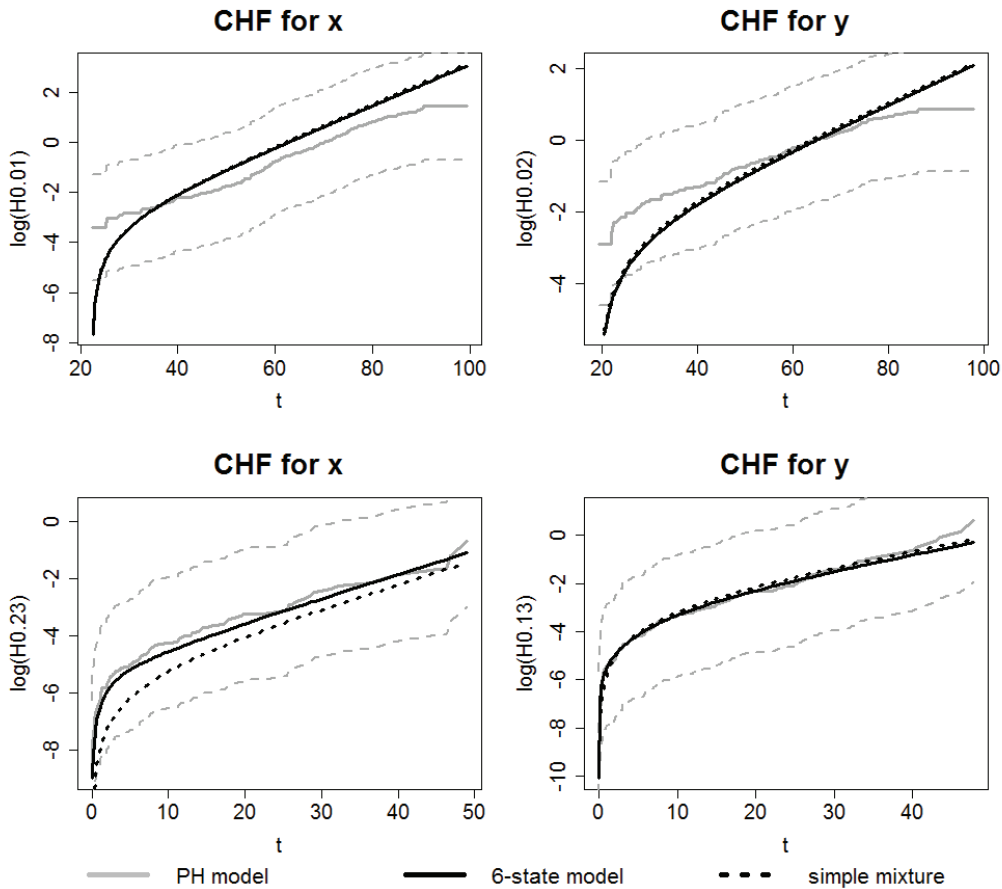


Figure 4.4: Baseline Cumulative Hazard Function. The top left panel shows the cumulative function of $\lambda^{01}(t | \mathcal{H}_t^1)$. The bottom left panel shows the cumulative function of $\lambda^{23}(t | \mathcal{H}_t^2)$. The top right panel shows the cumulative function of $\lambda^{02}(t | \mathcal{H}_t^1)$. The bottom right panel shows the cumulative function of $\lambda^{13}(t | \mathcal{H}_t^3)$.

additional force of mortality eventually exists. On the other hand, by vertically comparing the three panels for each gender, we can see that the longer the duration, the slower the force of mortality decreases.

Comparing the black curves with the grey curves, we see that although both models show the additional force of mortality for the survivor fades away, the six-state model displays a greater additional force of mortality upon the first death.

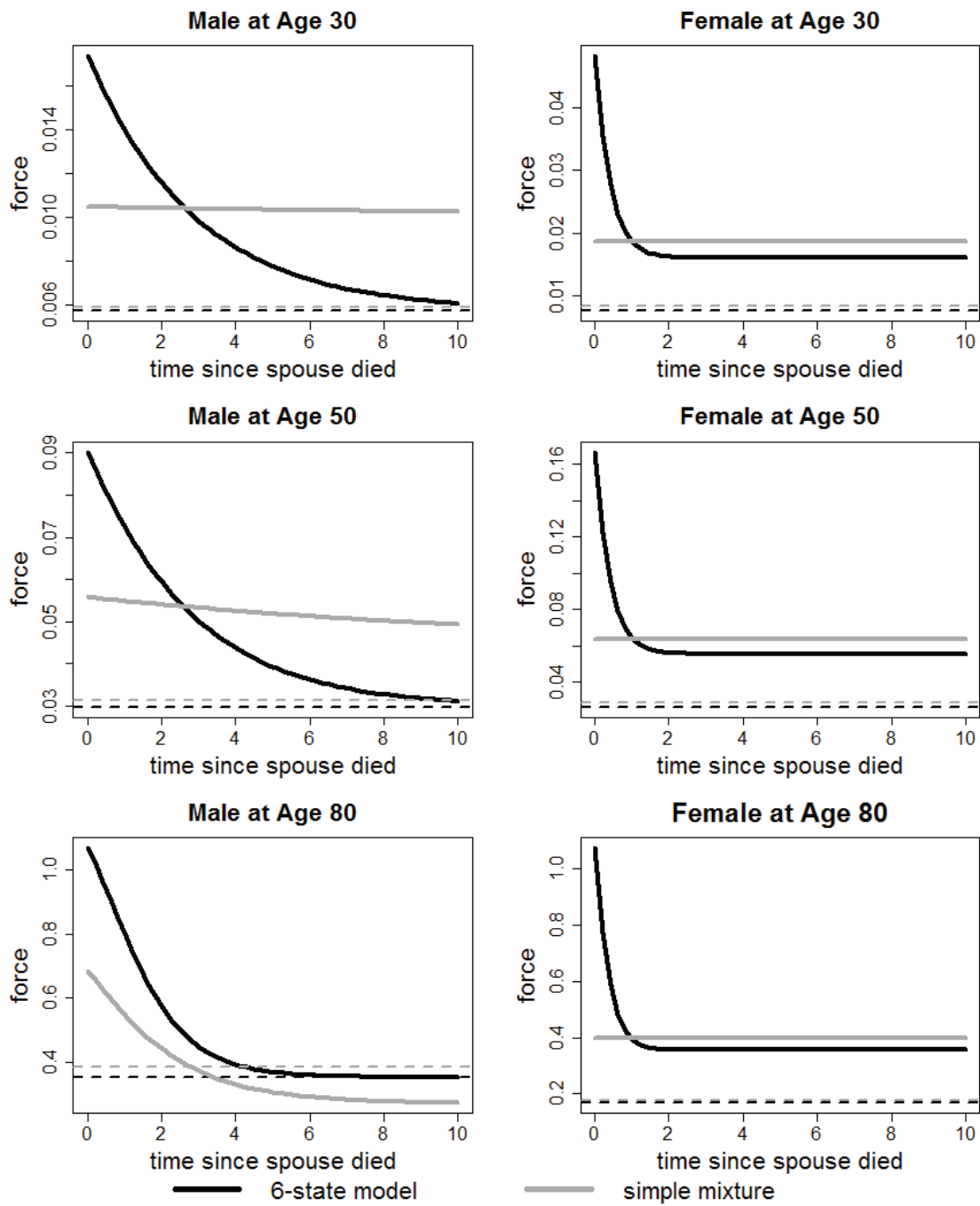


Figure 4.5: The force of mortality for a surviving spouse of a fixed age at various times since the first death.

4.6 Conclusion and Discussion

This chapter investigates a new method to model the dependence of joint lifetimes. Using a simple mixture of two four-state Markov chains, we can capture dependence that arises due to heterogeneity of the population. Furthermore, this approach allows the history of survivorship to provide information about the forces of mortality. Applying a six-state Markov chain with unobservable states, we can capture the short-term dependence of spouses by introducing the diminishment of the additional force of mortality of the survivor. Under both approaches, by viewing the forces of mortality conditional on different histories, we observe the immediate increase of the force of mortality for the survivor when the spouse dies and the speed of diminishment of the additional forces of mortality afterwards.

With the data set from Allin (1998) and the MLE method, both fitted models behave roughly as we expected. Specifically, when one spouse dies, the force of mortality for the surviving spouse at a fixed age decreases with time since the first death. As our motivation is to capture this behaviour, which is known as short-term dependence of joint lifetimes, both of the two models accomplish this task. The six-state model is an expression of a generalized mixture of two Markov chains. Compared to the simple mixture of two four-state Markov chains, the short-term dependence is more apparent in the six-state model.

Further study is necessary if one desires a well-fitting model. Our models do not perfectly capture the forces of mortality for the couples, but this is not the main purpose of this chapter. The forces of mortality for males after their spouses' death are overestimated. This could be due to our assumptions on the forces of mortality. After all, Gompertz law is a fairly simple model for the forces of mortality. Also, we have modelled the mortality improvement in a simple way.

Chapter 5

Modelling a Physiological Aging Process by Mixture of Markov Chains

5.1 Introduction

For the pricing and valuation of life insurance and annuity products, many actuarial models are used to estimate the mortality rates. The usual life table assumes that the population under study is homogeneous. However, in reality, there are many factors that affect the health of humans. Thus, many models have been developed for the heterogeneity of a population. Vaupel, Manton and Stallard (1979) introduced a frailty model, which contains a random variable indicating the individual frailty. Given a frailty model, many assumptions on the frailty distribution are studied. Vaupel, Manton and Stallard (1979), Lancaster (1979), Lancaster and Nickell (1980) used the gamma distribution to model frailties. Vaupel and Yashin (1983) also proposed to use a two point distribution, the uniform distribution, the Weibull distribution and the lognormal distribution to model frailties. Hougaard (1984) and Hougaard (1986) illustrated the inverse gaussian distribution to model frailties. Meyrike and Sherris (2013) developed a method to adjust annuity prices to allow for frailty.

Traditionally, many mortality models and projections are based on people's calendar ages.

However, for one individual, he/she may be healthier or frailer than the others at the same calendar age. Therefore, the calendar age cannot perfectly reflect the health information of one individual. In fact, the differences in the health status of people at the same calendar age lead to different physiological ages, where “healthier” can be translated as “younger”, and “frailer” can be translated as “older”.

Lin and Liu (2007) introduced a physiological ageing model that can link the ages of individuals to their physiological ages by assuming the time of death follows a phase-type distribution. Su and Sherris (2012) compared a frailty model with a gamma frailty distribution to the physiological ageing model using Australian cohort data (1940 and 1945). They conclude that the frailty model with an assumption of Gompertz mortality results in log mortality rates that increase linearly with age, while the physiological ageing model exhibits a concave behaviour of log mortality rates at the older ages.

Extensions of Lin and Liu (2007) with consideration of health cost information were discussed by Govorun, et al (2016). In this chapter, we also develop several extensions of Lin and Liu (2007) to investigate how fast people are ageing given their health status. Since we focus on heterogeneity of the ageing speeds in a population, we mix over transition intensities between phases instead of forces of mortality for the phases as in the frailty model.

In this chapter, Section 5.2 reviews the physiological ageing model introduced by Lin and Liu (2007). Section 5.3 explores several discrete mixture models by extending the physiological ageing model. Starting with a simple mixture of two Markov processes, two more complicated mixture models involving embedded Markov processes are discussed afterwards. Section 5.4 investigates a continuous mixture of Markov processes by assuming transition intensities follow a continuous distribution. Also, a mixture model with an embedded autoregressive process is discussed. Finally, Section 5.5 concludes this chapter and discusses some potential future work.

5.2 The Physiological Aging Model

Traditionally, survival models specified deterministic age-specified mortality rates for a population. However, changes in the force of mortality should be mainly linked to the health status of an individual rather than his/her calendar age. In order to capture this fact, Lin and Liu (2007) proposed a model that “can link its parameters to the biological/physiological mechanism of ageing to a certain extent”. They consider a Markov process with state space that consists of a single absorbing state representing death and physiological states $E = \{1, 2, 3, \dots, n\}$, in which each state is referred to as a phase that represents the physiological age based on the health status.

As shown in Figure 5.1, the model assumes that newborns start from the very first phase. And starting from birth, each individual proceeds through the physiological states until death. Therefore, two parameters are used for each physiological age $i, i = 1, 2, \dots, n$. The parameter λ_i is used to describe the development of the ageing process, and q_i is used to describe the force of mortality in that physiological state. The time spent in phase i follows an exponential distribution with mean $\frac{1}{\lambda_i}$. Since there is one single absorbing state, the time of death follows a phase-type distribution.

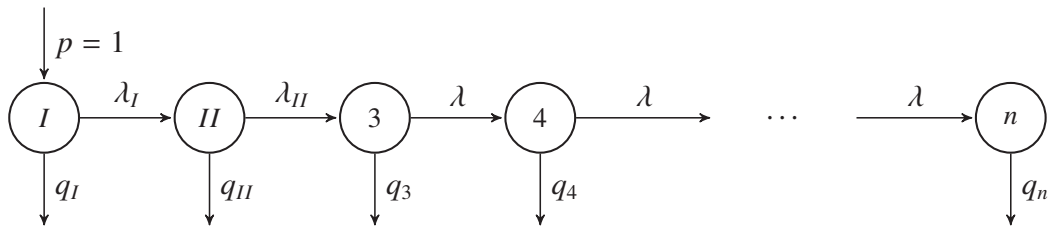


Figure 5.1: The physiological ageing model

A general theory of the phase-type distribution was given by Neuts (1981) and Asmussen (1987). For the model of Lin and Liu (2007), with the n transient physiological states and one single state representing death, let $\{Y_t, t \geq 0\}$ be a time-homogeneous continuous-time Markov ageing process. The initial distribution of the process is given by the n -component

vector $\alpha = (1, 0, 0, \dots, 0)'$. The infinitesimal generator of the Markov process is expressed as a $(n + 1) \times (n + 1)$ matrix

$$Q = \begin{pmatrix} \Lambda & \mathbf{q} \\ \mathbf{0}' & 0 \end{pmatrix},$$

where $\mathbf{0}$ is a column vector of zeroes and the $n \times n$ sub-generator matrix Λ is given by

$$\Lambda = \begin{pmatrix} -(\lambda_1 + q_1) & \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & -(\lambda_2 + q_2) & \lambda_2 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & -(\lambda_{n-2} + q_{n-2}) & \lambda_{n-2} & 0 \\ 0 & 0 & \cdots & 0 & -(\lambda_{n-1} + q_{n-1}) & \lambda_{n-1} \\ 0 & 0 & \cdots & 0 & 0 & -q_n \end{pmatrix}. \quad (5.1)$$

Therefore, the survival time T is said to follow a phase-type distribution with representation (α, Λ) . By computing the matrix exponential, the transition probability matrix is

$$P(t) = \exp(\Lambda t).$$

For an individual aged x in state i , the probability of being in state j at age $x + t$ is then $\Pr(Y_{x+t} = j \mid Y_x = i) = P_{ij}(t)$. And the survival function of a newborn is

$$\begin{aligned} S(x) &= \Pr(Y_x \neq n + 1) \\ &= \alpha' \exp(\Lambda x) \mathbf{1}, \end{aligned}$$

where $\mathbf{1}$ is column vector of ones. Also, let $f_x(i) = \Pr(Y_x = i \mid Y_x \neq n + 1)$ and $\mathbf{f}_x =$

$(f_x(1), f_x(2), \dots, f_x(n))$. Then

$$f_x = \frac{\alpha' \exp(\Lambda x)}{\alpha' \exp(\Lambda x) \mathbf{1}}.$$

The age-specific force of mortality is then given by

$$\mu_x = \sum_{i=1}^n f_x(i) q_i. \quad (5.2)$$

Lin and Liu (2007) also consider the first k states labeled as Roman-numbered phases to represent a developmental period in which newborns adapt to the environment and develop to reach their maximal physiological performance. During this period, transition intensities from one phase to the next phase are different. However, beyond this developmental period, transition intensities from one phase to the next phase are the same.

To model the mortality, they add a phase dependent accident parameter to reflect changes in the forces of mortality for a period in early adulthood due to the behaviour-related accident rate. Together, the force of mortality for phase i is given by

$$q_i = \begin{cases} b + a + i^p \cdot q & \text{for } i_1 < i \leq i_2 \\ b + i^p \cdot q & \text{for } i \leq i_1, \ i_2 < i < n, \\ \lambda_n + b + n^p \cdot q & \text{for } i = n \end{cases}$$

where a is the behaviour-related accident rate which appears only between phases i_1 and i_2 , b is a background rate, q and p reflect the impact of ageing. Notice that the expression for q_n is used by Lin and Liu (2007) for the purpose of truncating the state space so that it is finite.

To investigate this model, the Swedish population cohort data (1911) is fitted in Lin and Liu (2007). With the assumptions that $n = 200$ and $k = 4$, the parameter estimates are shown in the Table 5.1. Since our models in this chapter are developed from their model, the plots and analysis in the rest of this chapter are also based on these estimates.

Table 5.1: Parameter Estimates for Swedish Cohort (1911)

λ	b	a	$[i_1, i_2]$	q	p		
2.3707	9.0987e-04	2.8939e-03	[33,70]	1.8872e-15	6		
Development Period				$k = 4$			
q_I	λ_I	q_{II}	λ_{II}	q_{III}	λ_{III}	q_{IV}	λ_{IV}
0.1671	1.7958	0.0097	0.5543	0.0003	3.5061	0.0149	0.6535

Figure 5.2 shows the age-specific force of mortality that results from equation (5.2) using the parameter estimates in Table 5.1.

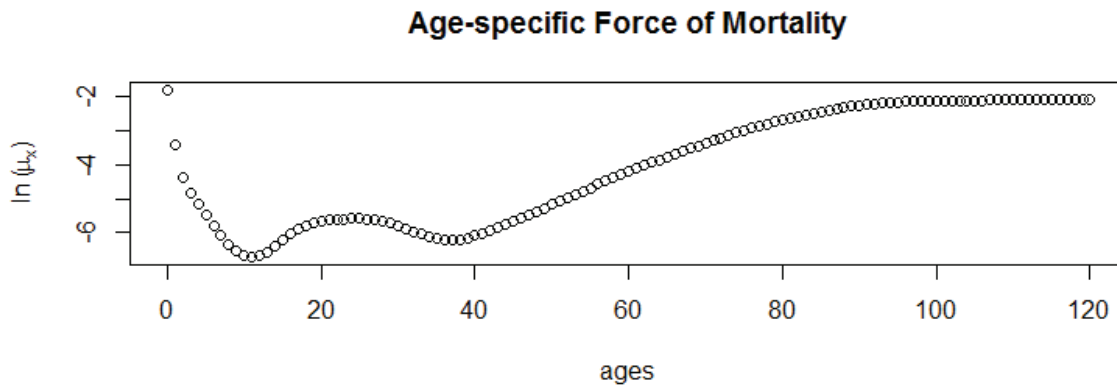


Figure 5.2: Age-specific force of mortality for Swedish Cohort (1911)

5.3 A Discrete Mixture of Markov Processes

Although Lin and Liu (2007) have estimated the variety of forces of mortality given an individual’s physiological age, the assumption of fixed transition intensities between phases (the λ_i ’s) creates a difficulty to model a heterogeneous population. However, in reality, many factors affect the speed of an individual’s ageing. In order to capture this fact, this section develops several discrete mixtures of Markov processes. In Section 5.3.1, we start with a simple attempt by using a mixture of two Markov processes, where each Markov process has its unique fixed

λ_i 's. Afterwards, in section 5.3.2, we develop a mixture of three Markov processes, and transitions among the three processes are allowed. Finally, in section 5.3.3, we develop a mixture by applying the mover-stayer model idea.

5.3.1 Simple Mixture of Two Markov Processes

In reality, there are millions of reasons for people to age at different speeds. Some objective factors may influence the health of people for their entire lifetimes, such as genetic differences, living environment, long-term lifestyle habits, and so on. These factors create heterogeneity of the population. To reflect this, we can group people into a healthier population and a frailer population, where intuitively, the healthier population is ageing more slowly than the frailer population. A mixture of two Markov processes can be used to investigate a population that combines two such sub-populations.

Figure 5.3 shows a mixture of two Markov processes; each Markov process is a physiological ageing process. We use similar notations as Lin and Liu (2007): λ_i is the transition intensity to the next phase for the phase i in the usual physiological ageing (non-mixture) model, and q_i is the force of mortality for phase i . Now, suppose that $\lambda_i^H = \lambda_i(1 - \rho)$ is the transition intensity for phase i in the healthier Markov process, where ρ is a positive constant smaller than 1, and that $\lambda_i^F = \lambda_i(1 + \rho)$ is the transition intensity for phase i in the frailer Markov process. Notice that the initial probability of entering each process is p and $1 - p$, which is determined by the composition of the population.

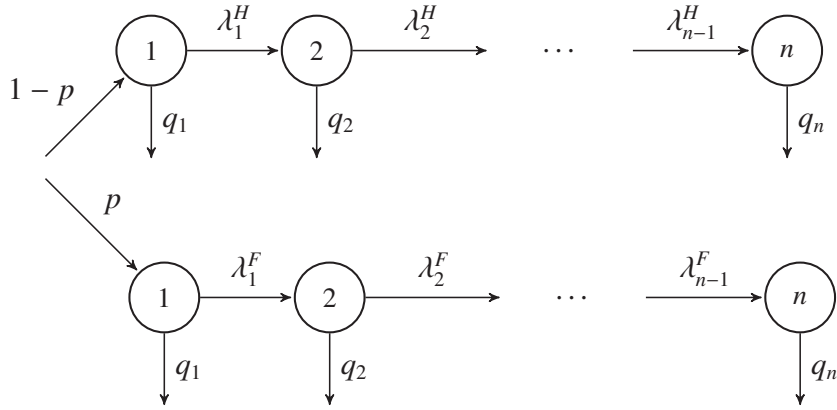


Figure 5.3: Mixture of two Markov processes

Since the mixture of phase-type distributions is still a phase-type distribution, in order to understand this mixture model, we can treat the mixture of two processes as one big phase-type distribution. Suppose X_t is the model state at age t with the state space $\{1, 2, \dots, 2n, 2n + 1\}$. When $X_t \in \{1, 2, \dots, n\}$, the individual is in the healthier process and when $X_t \in \{n + 1, n + 2, \dots, 2n\}$, the individual is in the frailer process. $X_t = 2n + 1$ means the individual's death. Compared to the physiological ageing process $\{Y_t, t > 0\}$, notice that $Y_t = i$ corresponds to $X_t \in \{i, n + i\}$ for $i = 1, 2, \dots, n$ and $Y_t = n + 1$ corresponds to $X_t = 2n + 1$.

Therefore, for the mixture of a healthier Markov process and a frailer Markov process, the time until death follows a phase-type distribution with a $2n \times 2n$ sub-generator matrix

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}^H & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}^F \end{pmatrix},$$

and initial distribution

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_{2n})'.$$

The $\mathbf{\Lambda}^H$ and $\mathbf{\Lambda}^F$ matrices are in the same form as expression (5.1). Also, $\alpha_1 = 1 - p$, $\alpha_{n+1} = p$, and $\alpha_i = 0$ for $i = 2, 3, \dots, n, n + 2, n + 3, \dots, 2n$.

Using the Swedish population cohort data (1911), the phase distributions for several ages

are shown in Figure 5.4 for the choices $\rho = 0.07$ and $\rho = 0.2$. Both plots compare the phase distributions of the original model from Lin and Liu (2007) (non-mixture model) and the mixture of two Markov processes at age 30, 45, 60, 75 and 90, where the solid lines represent the non-mixture model and the circles represent the mixture of two Markov processes. Each age is shown in a unique colour. Notice that the first plot assumes ρ is 0.07. This means, in the mixture model, ageing speed in the healthier process is 7% slower than in the non-mixture process, and ageing speed in the frailer process is 7% faster than in the non-mixture process. Notice that all the later models will adopt the assumption of $\rho = 0.07$ for convenience. The second plot assumes that ρ equals 0.2, which creates a greater difference between the healthier and frailer processes.

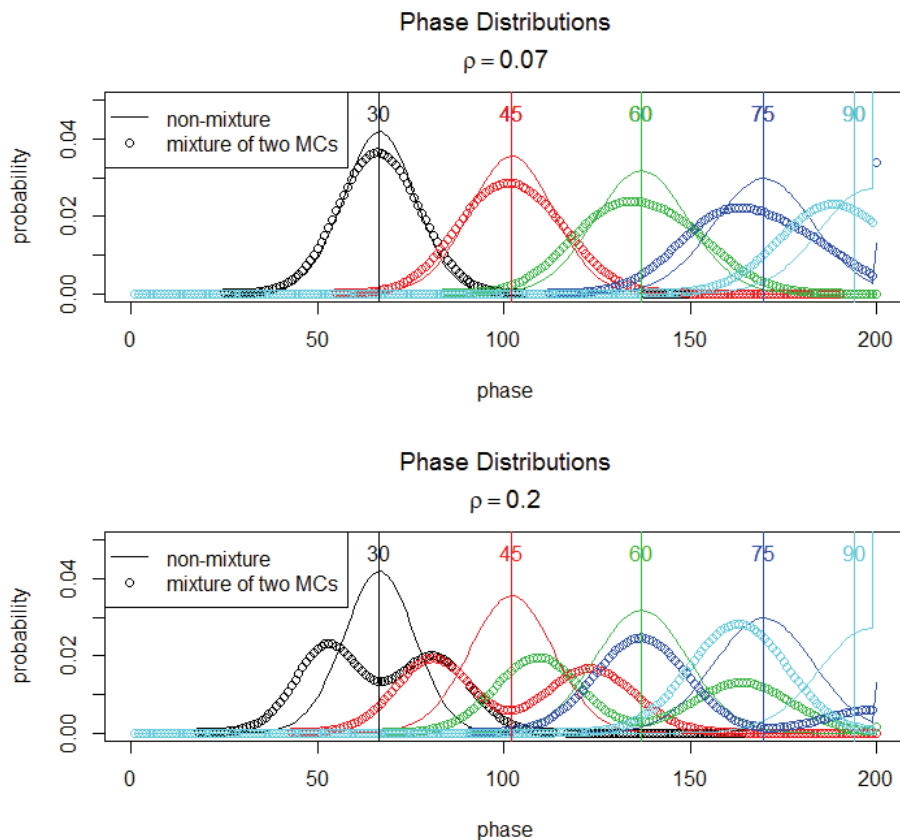


Figure 5.4: Phase distributions of mixture of two Markov processes

From the plot, we observe a significant increase in the variance when we use a mixture of two Markov processes. As expected, increasing the difference between the speeds of ageing in the two processes will increase the variance. When the difference is large enough, such as $\rho = 0.2$, each phase distribution has two modes. However, the appearance of bimodal distribution is not desirable because we expect that the continuum of ageing speeds in the population leads to a physiological ageing distribution with a single mode. In order to capture a more realistic behaviour, we may consider a continuous mixture of Markov processes, which will be discussed in a later section of this chapter.

5.3.2 Mixture of Three Markov Processes

The mixture of two Markov processes captures the fact that the ageing of persons in a population can be affected by some objective factors. However, there are always possibilities that the ageing speed of one individual can suddenly change due to a change of living environment, lifestyle or some other elements that play crucial roles in affecting his/her health. To reflect this in our model, transitions between processes should be allowed.

Some sociological factors may affect the health status of the residents in a particular area. For a simple example, residents in a heavily polluted industrial city are ageing faster than the average ageing speed of the overall population of the country. But residents in a rural area with no factory at all are ageing more slowly than the average ageing speed of the overall population of the country. If government realizes the pollution problem in that industrial city, then they may move several factories from the industrial city to the rural area. Theoretically, this implementation of pollution control strategies will improve the speed of ageing for the residents in the industrial city but speed up residents' ageing in the rural area. Thus, in this example, the ageing speeds of residents in the two areas are moving towards to the average speed of the whole population of the country.

To model the reality in this example, we consider a mixture of three Markov processes: a healthier process, a frailer process, and a process for someone of average health. As shown in

Figure 5.5, the initial probabilities of entering the average, healthier, and frailer process are p_A , p_H , and p_F respectively. Notice that $p_A + p_H + p_F = 1$. Similar to the mixture of two Markov processes, λ_i^H and λ_i^F are constructed by λ_i and ρ , representing the transition intensities for phase i in the healthier process and the frailer process. λ_i represents the transition intensity for phase i in the average process.

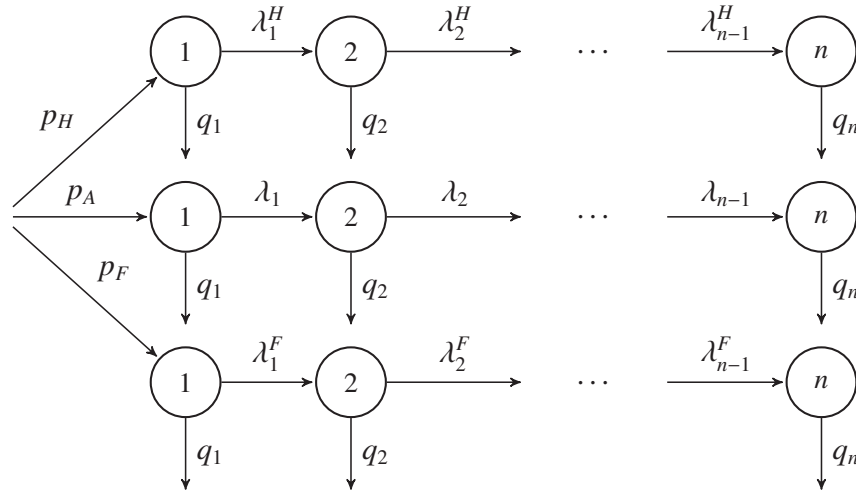


Figure 5.5: Mixture of three Markov processes

To model changes of ageing speed, an embedded discrete-time Markov process is involved, as shown in Figure 5.6. We assume that transitions between the healthier, the average, and the frailer processes can occur only at the times of transition from one phase to the next and that the phase of the individual is preserved when transitions between processes occur. The probabilities of moving between processes can be expressed as a 3×3 probability matrix:

$$\mathbf{P} = \begin{matrix} & \begin{matrix} H & A & F \end{matrix} \\ \begin{matrix} H \\ A \\ F \end{matrix} & \begin{pmatrix} p^{HH} & p^{HA} & 0 \\ p^{AH} & p^{AA} & p^{AF} \\ 0 & p^{FA} & p^{FF} \end{pmatrix} \end{matrix} .$$

For simplicity, we assume there is no direct transition between the healthier process and the

frailer process.

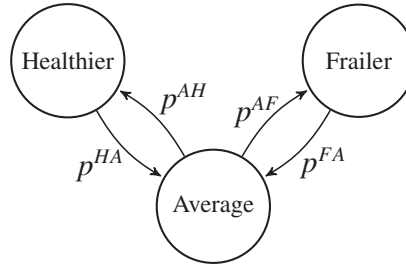


Figure 5.6: Embedded discrete-time process

To understand this mixture model, suppose that the process indicating the model state at age t , $\{X_t, t > 0\}$, has state space $\{1, 2, \dots, 3n, 3n + 1\}$, and suppose $\{J_t, t > 0\}$ is the Markov process for which J_t indicates the physiological ageing process the individual is in at age t . Thus, $J_t = H$ corresponds to $X_t \in \{1, 2, \dots, n\}$, $J_t = A$ corresponds to $X_t \in \{n + 1, n + 2, \dots, 2n\}$, and $J_t = F$ corresponds to $X_t \in \{2n + 1, 2n + 2, \dots, 3n\}$. Notice that J_t is not defined after the individual dies, in the other words when $X_t = 3n + 1$. Compared to the physiological ageing process $\{Y_t, t > 0\}$, notice that $Y_t = i$ corresponds to $X_t \in \{i, n + i, 2n + i\}$ for $i = 1, 2, \dots, n$ and $Y_t = n + 1$ corresponds to $X_t = 3n + 1$.

Therefore, for this model, the time until death follows a phase-type distribution with a $3n \times 3n$ sub-generator matrix Λ given by

$$\Lambda = \begin{pmatrix} \Lambda^H p^{HH} & (\Lambda^H - \text{diag}(\Lambda^H)) p^{HA} & \mathbf{0} \\ (\Lambda^A - \text{diag}(\Lambda^A)) p^{AH} & \Lambda^A p^{AA} & (\Lambda^A - \text{diag}(\Lambda^A)) p^{AF} \\ \mathbf{0} & (\Lambda^F - \text{diag}(\Lambda^F)) p^{FA} & \Lambda^F p^{FF} \end{pmatrix},$$

and initial distribution

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_{3n})'.$$

The matrices Λ^H , Λ^A , and Λ^F are in the same form as expression (5.1). Also, $\alpha_1 = p_H$, $\alpha_{n+1} = p_A$, $\alpha_{2n+1} = p_F$, and $\alpha_i = 0$ for $i = 2, 3, \dots, n, n + 2, n + 3, \dots, 2n, 2n + 2, 2n + 3, \dots, 3n$.

Notice that this mixture of three Markov processes is a more complicated mixture model.

The mixing probabilities depend on the individual's age. Specifically, given an individual is alive at time t , the mixing probabilities are

$$\begin{aligned} P(J_t = H | Y_t \neq n + 1) &= \frac{\sum_{i=1}^n (\boldsymbol{\alpha}' \exp(\boldsymbol{\Lambda}t))_i}{\boldsymbol{\alpha}' \exp(\boldsymbol{\Lambda}t) \mathbf{1}}, \\ P(J_t = A | Y_t \neq n + 1) &= \frac{\sum_{i=n+1}^{2n} (\boldsymbol{\alpha}' \exp(\boldsymbol{\Lambda}t))_i}{\boldsymbol{\alpha}' \exp(\boldsymbol{\Lambda}t) \mathbf{1}}, \\ P(J_t = F | Y_t \neq n + 1) &= \frac{\sum_{i=2n+1}^{3n} (\boldsymbol{\alpha}' \exp(\boldsymbol{\Lambda}t))_i}{\boldsymbol{\alpha}' \exp(\boldsymbol{\Lambda}t) \mathbf{1}}. \end{aligned}$$

Using the Swedish population cohort data (1911), the phase distributions for several ages are shown in Figure 5.7. The solid curves represent phase distributions of the three non-mixture Markov processes, where green curves represent the healthier process, red curves represent the frailer process, black curves represent the average process. The circles represent phase distributions of the mixture of three Markov processes. From the top to the bottom, the three panels show the phase distributions of a mixture of three Markov processes at age 30, 45, and 60. In these plots, we assume an equal chance to initially enter each process, and we assume the embedded probability matrix to be

$$\mathbf{P} = \begin{pmatrix} 0.9 & 0.1 & 0 \\ 0.4 & 0.2 & 0.4 \\ 0 & 0.1 & 0.9 \end{pmatrix}.$$

This implies that there is a much higher probability of staying in the healthier or the frailer process once an individual enters one of those two processes. Compared to the non-mixture process, additional variance is observed due to this construction.

In this model, we notice that the increase in the ages causes an increase in the variance. Numerical evidences are given: for the non-mixture model, variances at ages 30, 45 and 60 are 94.28, 128.59 and 160.03, respectively; for the mixture of three Markov processes with the probability matrix described above, variances at ages 30, 45 and 60 are 99.48, 136.59 and

170.46, respectively. Another expected phenomenon worthy of mention is that more probability of moving towards the average process results in less additional variance. For instance, assuming the embedded probability matrix is

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 0.5 & 0.5 \end{pmatrix},$$

the variances at ages 30, 45 and 60 are 94.90, 129.47 and 161.17, which are much closer to the variances of the non-mixture model.

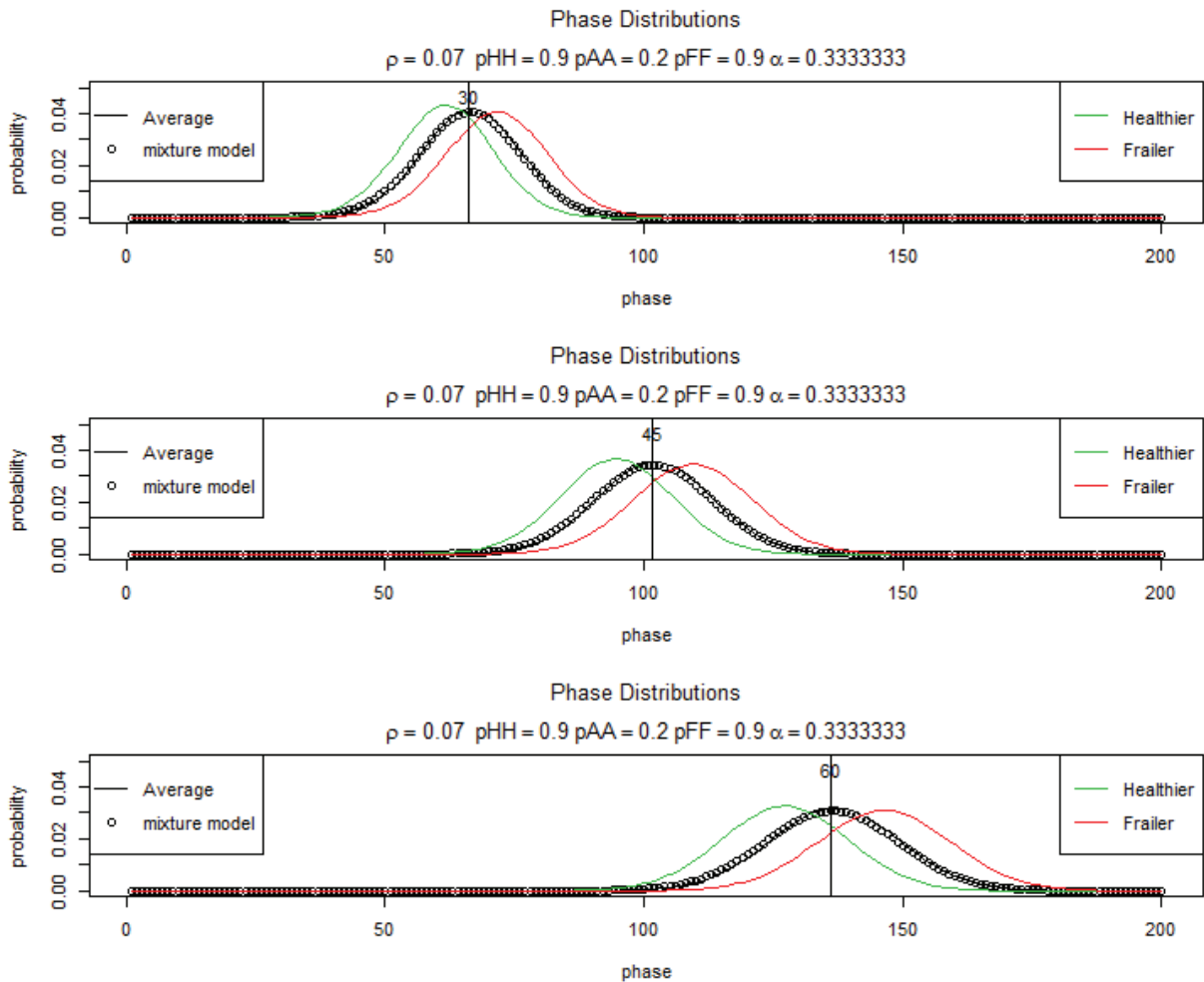


Figure 5.7: Phase distributions of the mixture of three Markov processes

5.3.3 A Mixture Involving The Mover-stayer Model

From the mixture of three Markov processes, we observe that there is more additional variance if there is a higher probability of being in the healthier or the frailer process. To investigate the probability of being in these two extreme processes, we next investigate the chance of staying in a healthier process or moving to the frailer process by applying the idea as a mover-stayer model. The mover-stayer model is a model introduced by Blumen, Kogan, and McCarthy (1955). They consider the workers in industries as movers or stayers. We can also take this idea to model a mixture of two Markov processes.

Consider two Markov processes. Similar to the mixture of two Markov processes in Section 5.3.1, there is an initial probability of being in each process. For an individual who enters one of the processes, his/her ageing speed can stay the same until death or can change over time due to life changes. To reflect this reality, we allow the individual to be a stayer or a mover. Being a stayer means once one starts in a process, he/she will stay in that process; being a mover means that an individual can make transitions between the two processes according to probabilities captured by an embedded two-state discrete time process. Figure 5.8 shows an embedded process for a mover in a mixture of one healthier process and one frailer process.

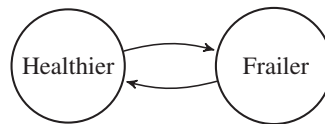


Figure 5.8: Transitions of a mover

To understand this mixture model, suppose that the process indicating the model state at age t , $\{X_t, t > 0\}$, has state space $\{1, 2, \dots, 4n, 4n + 1\}$, and suppose $\{J_t, t > 0\}$ is the Markov process for which J_t indicates the physiological ageing process the individual is in at age t . Thus, $J_t = H$ corresponds to $X_t \in \{1, 2, \dots, n\} \cup \{2n + 1, 2n + 2, \dots, 3n\}$, $J_t = F$ corresponds to $X_t \in \{n + 1, n + 2, \dots, 2n\} \cup \{3n + 1, 3n + 2, \dots, 4n\}$, and J_t is not defined after the individual dies, that is, when $X_t = 4n + 1$. Suppose that the individual is a stayer when $X_t \in \{1, 2, \dots, 2n\}$,

and the individual is a mover when $X_t \in \{2n + 1, 2n + 2, \dots, 4n\}$. In terms of the physiological ageing process $\{Y_t, t > 0\}$, notice that $Y_t = i$ corresponds to $X_t \in \{i, n + i, 2n + i, 3n + i\}$ for $i = 1, 2, \dots, n$ and $Y_t = n + 1$ corresponds to $X_t = 4n + 1$.

In this mixture model, some probabilities need to be specified by additional information about an individual, such as

1. initial distribution:
 - (a) probability of being a stayer in the healthier process $p_{s,H}$,
 - (b) probability of being a stayer in the frailer process $p_{s,F}$,
 - (c) probability of being a mover in the healthier process $p_{m,H}$,
 - (d) probability of being a mover in the frailer process $p_{m,F}$;
2. given being a mover, the probability matrix

$$\mathbf{P} = \begin{array}{c} H \quad F \\ \begin{array}{c} H \\ F \end{array} \begin{pmatrix} p^{HH} & p^{HF} \\ p^{FH} & p^{FF} \end{pmatrix} \end{array}.$$

Notice that the probability of being a mover is denoted by $p_m = p_{m,H} + p_{m,F}$. Therefore, for this model, the time until death follows a phase-type distribution with a $4n \times 4n$ sub-generator matrix $\mathbf{\Lambda}$ given by

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}^H & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}^F & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Lambda}^H p^{HH} & (\mathbf{\Lambda}^H - \text{diag}(\mathbf{\Lambda}^H)) p^{HF} \\ \mathbf{0} & \mathbf{0} & (\mathbf{\Lambda}^F - \text{diag}(\mathbf{\Lambda}^F)) p^{FH} & \mathbf{\Lambda}^F p^{FF} \end{pmatrix},$$

and initial distribution

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_{4n})'.$$

The matrices Λ^H , and Λ^F are in the same form as expression (5.1). Also, $\alpha_1 = p_{s,H}$, $\alpha_{n+1} = p_{s,F}$, $\alpha_{2n+1} = p_{m,H}$, $\alpha_{3n+1} = p_{m,F}$, and $\alpha_i = 0$ for $i \neq 1, n + 1, 2n + 1, 3n + 1$.

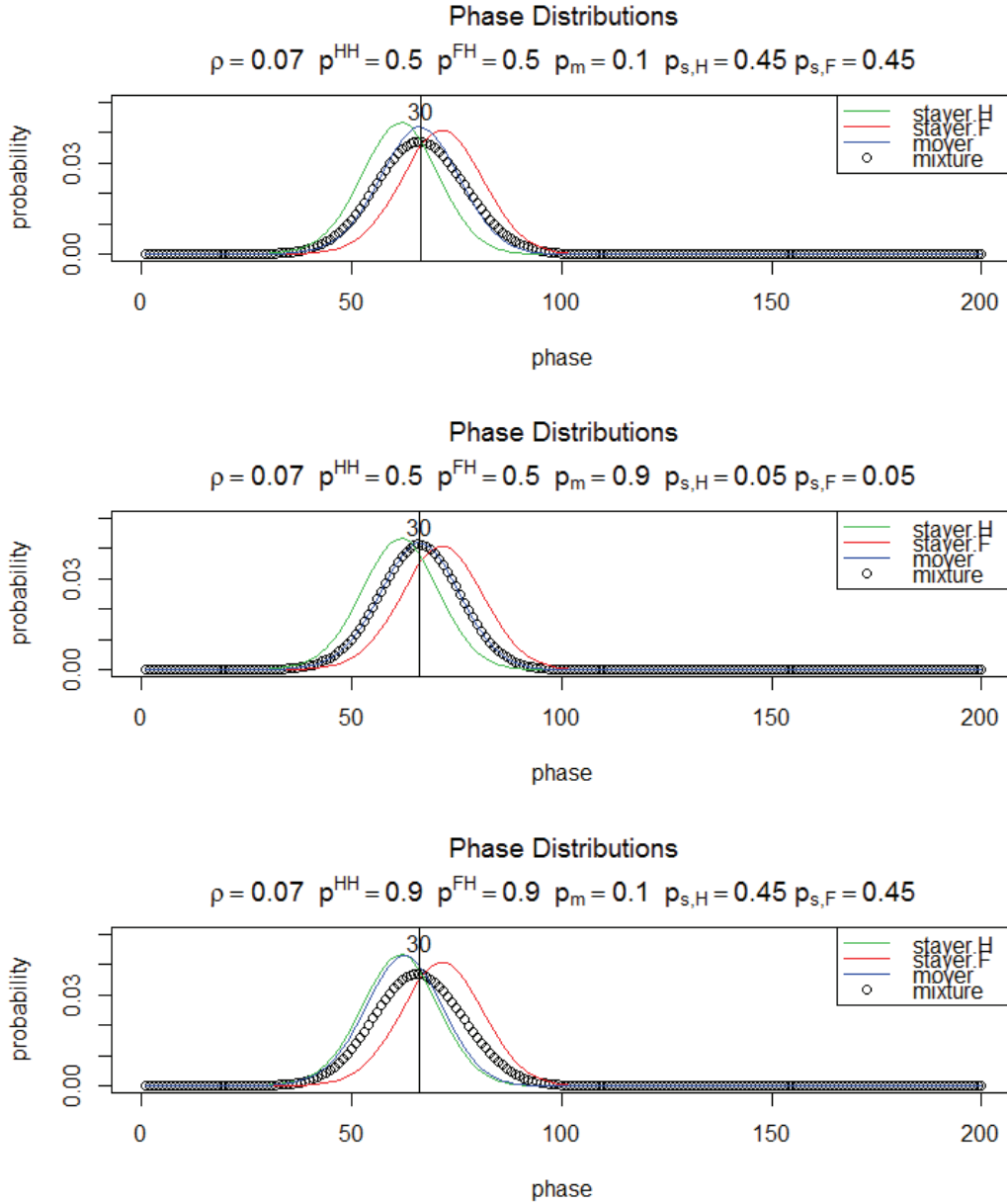


Figure 5.9: Phase distributions of the mixture of mover-stayer processes

Using the Swedish population cohort data (1911), the phase distributions for age 30 with different probability assumptions are shown in Figure 5.9. In the three plots, we assume $\alpha =$

$(0.25, 0, \dots, 0, 0.25, 0, \dots, 0, 0.25, 0, \dots, 0, 0.25, 0, \dots, 0)$. The top two panels assume that, for a mover,

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix},$$

They compare the change due to changes in the probability of being a mover. As we observe, the larger the probability of being a mover, the less the additional variance. The bottom panel shows that given a higher probability of being a mover and higher probability that a mover is in the healthier process, we observe that the mixture distribution puts more weight on the healthier process.

5.4 A Continuous Mixture of Markov Processes

As we observe in the simple mixture of two Markov processes, if there is a large difference between the ageing speeds of the two processes, an undesirable bimodal distribution appears. In general, a discrete mixture of Markov processes can lead to a multi-modal distribution. However, in reality, the ageing speeds of persons can be very highly variable and it is always possible that the physiological age of one healthy individual is far younger than the physiological age of one frail individual at the same calendar age. In order to model various ageing speeds while eliminating the undesirable bimodal distribution, we consider a continuous mixture of Markov processes. Section 5.4.1 investigates a simple continuous mixture of Markov processes by introducing a continuous distributed random variable to the transition intensities between phases. Section 5.4.2 develops this continuous mixture as an autoregressive process.

5.4.1 Simple Continuous Mixture of Markov processes

In reality, the heterogeneous ageing speeds of people are determined by the characteristic of each individual. In the physiological ageing process, the ageing speeds of one individual are

presented by the transition intensities between phases. In order to capture this reality involving variability of ageing speeds, we construct a mixture of Markov processes by considering the randomness in the transition intensities between phases.

Suppose Z is a standard normal random variable, σ is a parameter that reflects the variability of ageing speeds for a population, and λ_i is the transition intensity from phase i to phase $i + 1$ in a non-mixture model. The transition intensity from phase i to phase $i + 1$ in this mixture model is assumed to be $\lambda_i e^{Z\sigma}$. Therefore, the mixing density for this continuous mixture is the standard normal density.

By introducing randomness to the transition intensities between phases, we introduce additional variance to the phase distribution. For the physiological ageing process $\{Y_t, t > 0\}$, we have $\text{Var}[Y_t] = \text{Var}[E[Y_t | Z]] + E[\text{Var}[Y_t | Z]]$, where $E[\text{Var}[Y_t | Z]]$ is the variance due to non-mixture physiological ageing process and $\text{Var}[E[Y_t | Z]]$ is the additional variance due to the randomness Z .

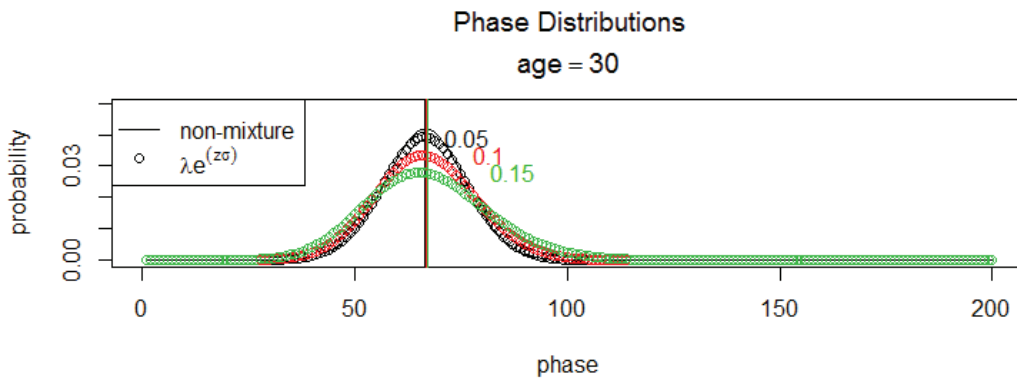


Figure 5.10: Phase distributions of the continuous mixture of Markov processes

Using the Swedish population cohort data (1911), the phase distributions for age 30 with assumptions $\sigma = 0.05$, $\sigma = 0.1$, and $\sigma = 0.15$ are shown in Figure 5.10. In the plot, the black solid line represents the phase distribution for the non-mixture model, the black circles represent the phase distribution for the continuous mixture model with $\sigma = 0.05$, the red circles represent the phase distribution for the continuous mixture model with $\sigma = 0.1$, and the green

circles represent the phase distribution for the continuous mixture model with $\sigma = 0.15$. As we expected, the larger the value of σ , the more additional variance appears.

5.4.2 Embedded Autoregressive Process

Practically, it is not reasonable to assume that Z is a constant throughout an individual's lifetime. Similar to the mixture of three Markov processes with an embedded discrete-time process, there are changes in an individual's ageing speed due to a variety of factors, for example, lifestyle changes. To achieve these changes, the transition intensity from phase i to phase $i + 1$ is assumed to be $\lambda_i e^{Z_i}$ for $i = 1, 2, \dots, n - 1$. Therefore, this mixture model can be understood as a continuous mixture of Markov processes with an embedded process such that transitions between processes can occur only at the times of transition from one phase to the next, and the phase of the individual is preserved when transitions between processes occur. However, Z_1, Z_2, \dots, Z_{n-1} are not independently distributed. In fact, Z_i should be correlated to Z_{i-1} . To achieve this, the embedded process $\{Z_i, i = 1, 2, \dots, n - 1\}$ is assumed to be an autoregressive process. That is, with constant ϕ and σ , $Z_i = \phi Z_{i-1} + W_i$, where $W_i \sim N(0, \sigma)$ and $Z_0 = 0$.

The assumption on the embedded autoregressive process brings complexities to fit this mixture model. Therefore, we use a simulation method to investigate this mixture model. Using the Swedish population cohort data (1911), the phase distributions for ages 30, 45, and 60 with fixed $\sigma = 0.05$ are shown in Figure 5.11. Again, the solid lines represent the phase distributions for the non-mixture model, and the circles represent the phase distribution for the mixture model. The top panel assumes $\phi = 0$, means that the transition intensities between phases are independent of each other. Compared to the non-mixture model, we observe no significant change in the variance of this model. This makes sense because of the law of large numbers. If the phases are independent, then, on average, the transition intensity for each phase will be the same as the non-mixture process. The second and third panels increase ϕ to be 0.05 and 0.8, respectively. We expect that stronger dependence of Z_i on Z_{i-1} will result in greater variability of the phase distribution at a given age. However, in these plots, we do not observe a

significant change in the variances of these two models either. The reason for this phenomenon is that the power functions of these small ϕ 's drops to 0 quickly. If we have assumed $\phi = 0.95$, as shown in figure 5.12, we observe significant additional variances for each phase distribution.

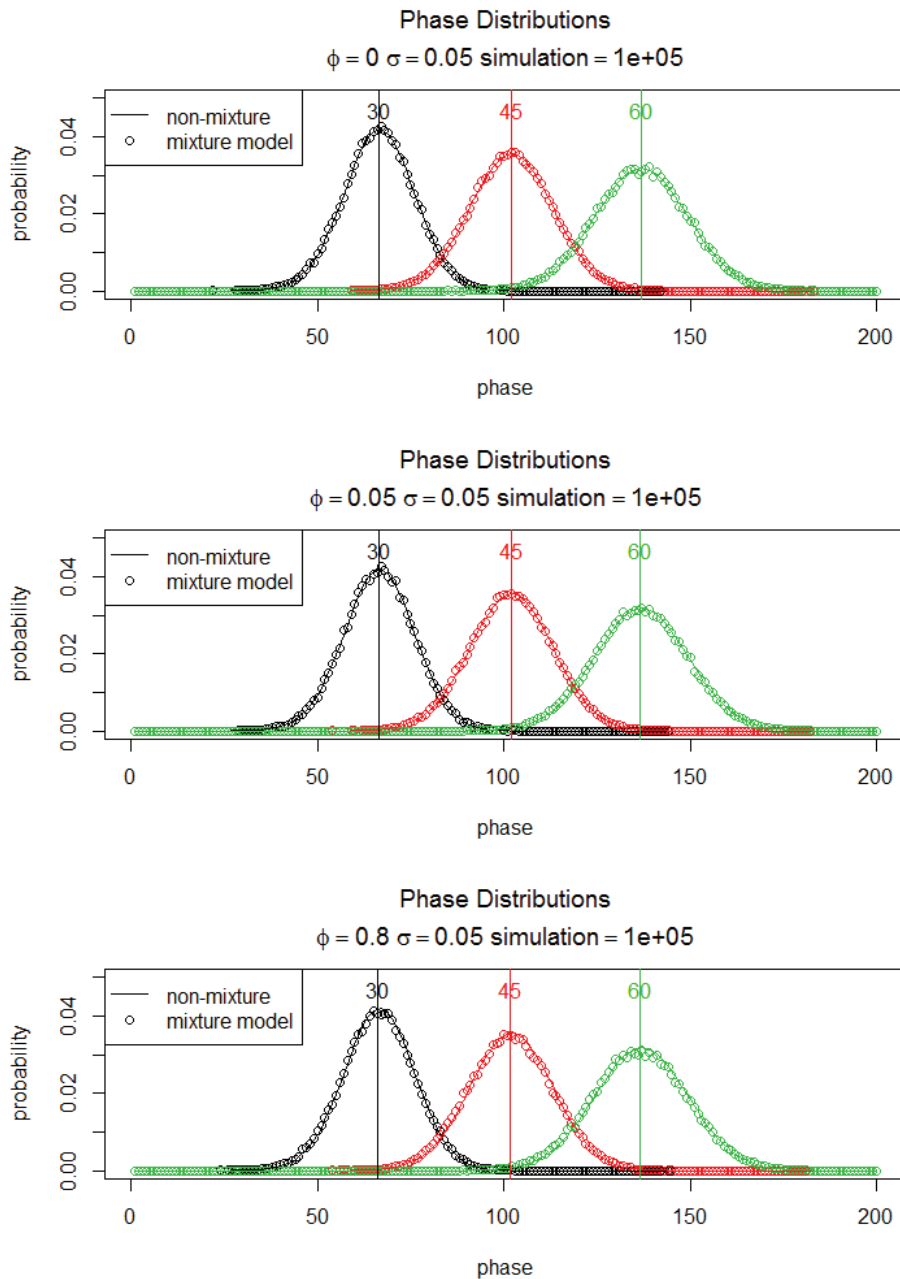


Figure 5.11: Phase distributions of autoregressive process mixture model with assumptions on ϕ .

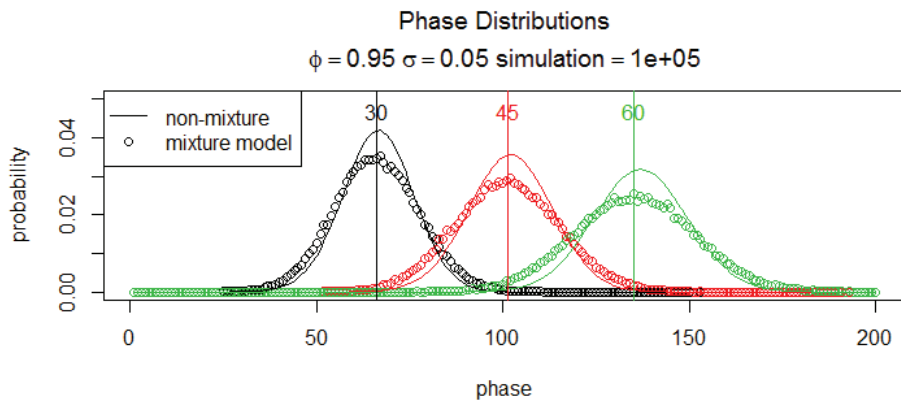


Figure 5.12: Phase distributions of autoregressive process mixture model with $\phi = 0.95$.

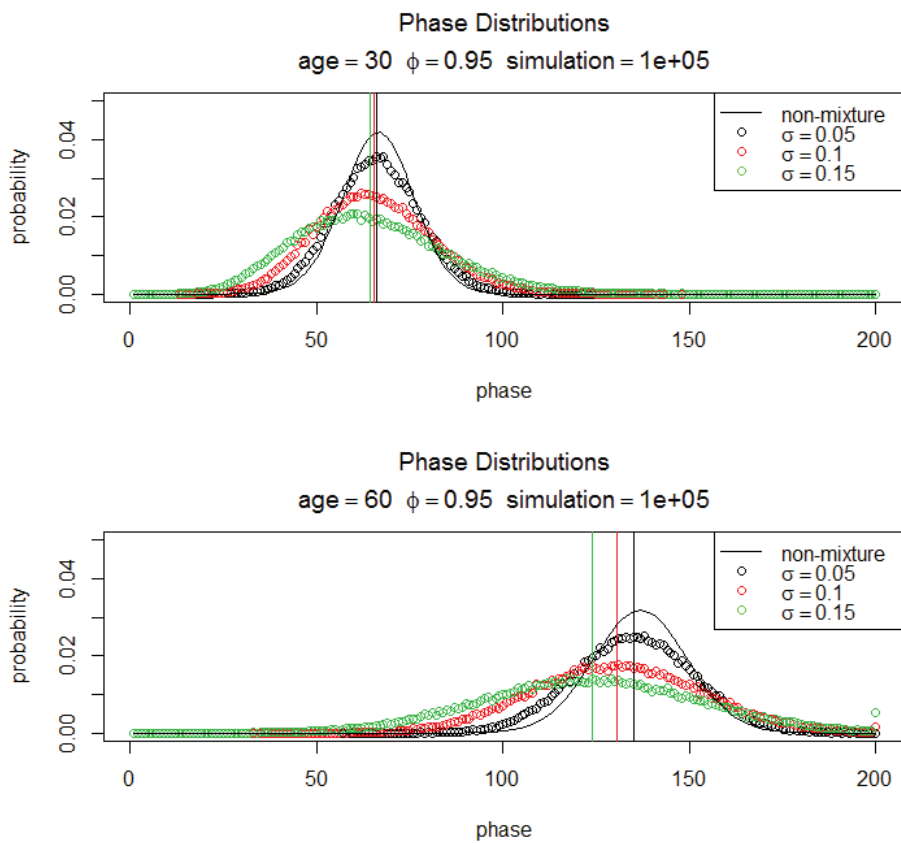


Figure 5.13: Phase distributions of autoregressive process mixture model with assumptions on σ .

We are also interested in how σ affects the phase distributions. In order to observe significant changes in variances, we assume $\phi = 0.95$ and compare the phase distribution with $\sigma = 0.05$, $\sigma = 0.1$, and $\sigma = 0.15$ in Figure 5.13. The top panel shows phase distributions at age 30, and the bottom panel shows phase distributions at age 60. As we expect, the larger the σ , the more additional variance. Also, the larger the age, the more change we observe.

5.5 Conclusion and Future Research

For the purpose of capturing the variability of ageing speed in reality, this chapter investigates several approaches to extending the physiological ageing process. Using a simple discrete mixture of Markov processes can achieve differences in the ageing speeds of a finite number of population groups. With an embedded Markov process, the discrete mixture model can achieve changes in the ageing speeds of a given individual. Grouping the population into movers and stayers allows one to reflect the fact that some factors can change during an individual's lifetime and other factors are more stable. Using a continuous mixture of Markov processes can enable one to model the heterogeneity by capturing large and small differences in the ageing speeds of individuals. With an embedded autoregressive process, the continuous mixture model can achieve changes in the ageing speed of each individual.

We go from discussing our simplest mixture model to more complicated mixture models. These models capture the reality better, but we are aware that these mixture models require more information from the data. That may include socioeconomic information, geographic information, demographic information, lifestyle information, and so on. Therefore, potential future research on this project may involve fitting the model with an appropriate data set.

Chapter 6

Conclusion

6.1 Concluding Remarks

Although Markov chains are commonly used to model the changes of individuals' statuses, most of these processes in reality show a non-Markov behaviour. This thesis investigates another approach to capture the heterogeneity in individuals' health and mortality by using mixtures of Markov chains. We investigate the rates of transitions in several multi-state models conditional on additional information about the process or the individual. Specifically, we develop our mixture models by considering three multi-state models: the three-state disability process, the four-state joint-life model, the phase-type physiological ageing model. Each model attracts us with different beauties. The interest in how the history of being disabled affects future probabilities of being disabled drives us to mix over the disability processes; the interest in how the death of an individual affects the future force of mortality for the spouse drives us to mix over the joint-life models; the interest in how to capture the heterogeneity of individuals' ageing speeds drives us to mix over the physiological ageing processes.

The simple mixtures of two disability processes discussed in Chapter 3 is the first model that inspires us to explore the impact of past histories on individuals' future mortality, morbidity and recovery rates. When we use a mixture of two Markov chains with same state space, we

can only observe whether the individual is active, disabled, or dead; we cannot observe whether the process is in Markov chain 1 or Markov chain 2. Therefore, the transition intensity is determined by the average of intensities of all possible transitions from the observed current state. This average intensity is associated with the whole history of the process. Also, we consider a continuous mixture of Markov chains, where the mixing variables follows a gamma distribution. In general, for any mixture of Markov chains with a known distribution for the mixing variables, we can use Laplace transforms to implement calculations of the history-dependent transition intensities. Our numerical examples of a mixture of two homogeneous three-state Markov chains and a mixture of two Markov chains with time-dependent intensities show appropriate results of how the transition intensities given the history of the process behave.

The mixture of two joint-life models discussed in Chapter 4 provides a new approach to investigate the dependence of joint lifetimes. The results show us that using a mixture model can successfully reflect the information provided by the survivorship of a couple. The simple mixture of two four-state Markov chains captures the dependence that arises due to heterogeneity in the health of couples. The more generalized mixture, also represented as a six-state model, captures the short-term dependence of spouses. Two additional states in the six-state model allow an unobservable diminishment of the surviving spouse's additional force of mortality to occur. To investigate both mixtures, a genealogical data set is used, and the MLE method is adopted. The results of both fitted models show that the force of mortality for the widow(er) at a fixed age decreases with time since the spouse's death. As a result, the six-state model captures more non-Markov behaviour than the simple mixture of two four-state models.

The discrete and continuous mixtures of physiological ageing processes considered in Chapter 5 can capture the heterogeneity in ageing speeds of a population in a wide variety of circumstances. The simple discrete mixture of Markov processes achieves differences in the ageing speeds for a finite number of population groups. Increasing the number of Markov processes to be mixed captures ageing speeds that are affected by more risk factors. When an embedded Markov chain is involved, changes in the ageing speeds of a given individual can

be achieved. Using the mover-stayer model for changes in the ageing process allows one to reflect the fact that risk factors affecting people's health can change over time or be relatively stable. Using continuous mixtures of Markov processes allows the ageing speeds of individuals to vary continuously. When an embedded autoregressive process are involved, changes in each individual's ageing speeds is specifically reflected.

6.2 Discussion and Further Research

In Chapter 3, for simplicity, we used Bernoulli distributions and gamma distributions as example distributions for the discrete and continuous mixing variables. However, there exists a wide range of distributions to be chosen from as the mixing distribution of a mixture model. The decision on the mixing distribution should be based on the specific circumstances.

For simplicity, we used Gompertz' law for the forces of mortality in Chapter 4. Actually, a variety of assumptions could be made for the forces of mortality. Also, notice that we expressed the six-state model as a generalized mixture model. However, instead of mimicing the six-state model, we can construct a generalized mixture model directly. That is, if $M(t)$ represents the mixing variable at time t , then by specifying the probability behaviour of $\{M(t), t \leq 0\}$, we can create a generalized mixture model.

For the mixtures of physiological ageing processes in Chapter 5, we only considered the mixtures of transition intensities between physiological states. This is because our interest is in investigating differences in people's ageing speeds. However, the assumption of fixed forces of mortality at each physiological age is not necessary. We can mix over the state-specific forces of mortality. In doing so, however, we lose the physiological age interpretation of the states.

Mixtures of Markov chains capture some realism, but involve additional complexity. Also, to use mixture models require additional information from the data. For example, in the mixture of disability processes, information about sickness history is required, such as the frequency of being disabled (sick) and the duration of disability (sickness) every time. To use the mixture

of joint-life models requires information about the survivorship of a couple, such as how long it has been since the spouse died. To use the mixture of physiological ageing process requires more information about the person, such as some personal information, socioeconomic information. Thus, using mixture models does complicate the implementation. Therefore, there is a need for future research on the application of mixture models to data.

Also, mixture models introduce some complexity to estimation of parameters, because of the larger numbers of parameters and the unobservable information of the process. This may lead to parameter identifiability problems, which can be addressed by imposing structural assumptions, such as parametric forms for the transition intensities. Normally, MLE is a useful method to estimate parameters. MLE for the simplest mixture of Markov chains, the mover-stayer model, is discussed by Frydman (1984). When unobservable quantities are involved, the well-known (expectation-maximization) EM algorithm may be a better option. The EM algorithm offers efficiency for the MLE. For a likelihood function, it might be difficult to maximize it. The EM algorithm first finds the expected value of the log-likelihood with respect to the unknown data given the observed data (E-step), and then maximizes the expectation computed in the first step (M-step). The EM algorithm have been used to estimate parameter in the mover-stayer model by Fuchs and Greenhous (1988), and in mixtures of Markov chains by Frydman (2005). Future studies on the estimation can be based on the work of these authors.

Bibliography

- [1] AALEN, O. (1987). Mixing distributions on a Markov Chain. *Scand J Statist* **14**, 281-289.
- [2] AALEN, O. (1988). Dynamic description of a Markov Chain with random time scale. *Mathematical Scientist* **13**, 90-103.
- [3] ADLER, N.E., BOYCE, T., CHESNEY, M.A., COHEN, S., FOLKMAN, S., AND SYME, S.L. (1994). Socioeconomic status and health: the challenge of the gradient. *American Psychologist* **49**, 15-24.
- [4] ALLIN, G.R. (1988). *Allin Families of North Devonshire*. Gateway Press, Inc., Baltimore.
- [5] ALLISON, D.B., FONTAINE, K.R., MANSON, J.E., STEVENS, J., AND VANITALLIE, T.B. (1999). Annual deaths attributable to obesity in the United States. *Journal of the American Medical Association*, 282(16).
- [6] ANDERSON, K.M., CASTELLI, W.P., LEVY, D. (1987). Cholesterol and mortality: 30 years of follow-up from the Framingham study. *Journal of the American Medical Association* **257**, 2176-2180.
- [7] ASMUSSEN, S. (1987). *Applied Probability and Queues*. New York: John Wiley.
- [8] ATTANASIO, O.P., AND EMMERSON, C. (2001): Differential mortality in the UK. *Institute for Fiscal Studies*, Working Paper no. 01/16.
- [9] BAGSHAW, S.M., EGI, M., GEORGE, C., BELLOMO, R. (2009). Early blood glucose control and mortality in critically ill patients in Australia. *Critical Care Medicine* **37**, 463-470.

- [10] BERNOULLI, D. (1766). Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir. *Hist. Acad. Roy. Sci.*, Année MDCCLX, Mémoires, 1-45. (English translation entitled 'An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it' In *Smallpox Inoculation: An Eighteenth Century Mathematical Controversy*, Bradley L. Adult Education Department: Nottingham, 1971, 21)
- [11] BLUMEN, I., KOGAN, M., AND MCCARTHY, P.J. (1955). *The industrial mobility of labor as a probability process*. Ithaca, NY: Cornell University Press.
- [12] BROWN, R., AND MCDAID, J. (2003). Factors affecting retirement mortality. *North American Actuarial Journal* **7**, 24-43.
- [13] DAMICO, G., GUILLEN, M., MANCA, R. (2009). Full backward non homogeneous semi-Markov processes for disability insurance models: a Catalunya real data application. *Insurance: Mathematical and Economics* **45**, 173-179.
- [14] DAMICO, G., GUILLEN, M., MANCA, R. (2013). Semi-Markov Disability Insurance Models. *Communications in Statistics Theory and Methods* **42**, 2872-2888.
- [15] DE MOIVRE, A. (1725). *Annuities on Lives: With Several Tables, Exhibiting at One View, the Values of Lives, for Different Rates of Interest*. MDCCLII.
- [16] DEATON, A., AND PAXSON, C. (1999). Mortality, education, income, and inequality among American cohorts. *National Bureau of Economic Research*, Working Paper no. 7140.
- [17] DEATON, A., AND LUBOTSKY, D. (2003). Mortality, inequality and race in American cities and states. *Social Science and Medicine* **56**(6), 1139-1153.
- [18] DEMPSTER, A.P., LAIRD, N.M., AND RUBIN, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Ser. B*, **39**, 1-38.

- [19] DICKSON, D.C.M, HARDY, M.R., AND WATERS, H.R. (2009). *Actuarial Mathematics for Life Contingent Risks*. Cambridge University Press, New York.
- [20] ELO, I.T., AND PRESTON, S.H. (1997). Racial and ethnic differences in mortality at older ages. *Racial and Ethnic Differences in The Health of Older Americans*, 10-42.
- [21] FRYDMAN, H. (1984). Maximum likelihood estimation in the mover-stayer model. *Journal of the American Statistical Association* **79**, 632-638.
- [22] FRYDMAN, H. (2005). Estimation in the mixture of Markov Chains moving with different speeds. *Journal of the American Statistical Association* **100**, 1046-1053.
- [23] FUCHS, C., AND GREENHOUSE, J.B. (1988). The EM algorithm for maximum likelihood estimation in the mover-stayer model. *Biometrics* **44**, 605-613.
- [24] GODTFREDSEN, N.S., HOLSE, C., PRESCOTT, E., VESTBO, J., OSLER, M. (2002). Smoking reduction, smoking cessation, and mortality: A 16-year follow-up of 19,732 men and women from the Copenhagen centre for prospective population studies. *American Journal of Epidemiology* **156**, 994-1001.
- [25] GOMPERTZ, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of Royal Society of London* **115**, 513-583.
- [26] GOVORUN, M, JONES, L. B., LIU, X, AND STANDFORD, D.A. (2016). Physiological age, health costs and their interrelation. submitted.
- [27] HABERMAN, S., AND PITACCO, E. (1999). *Actuarial Models for Disability Insurance*. CRC Press LLC, Boca Raton.
- [28] HASSAN-ZADEH, A., JONES, B.L., STANFORD, D.A. (2014). The use of phase-type models for disability insurance calculations. *Scandinavian Actuarial Journal* **2014** (8), 714-728.

- [29] HOEM, J.M. (1969). Markov Chain models in life insurance. *Blätter der DGVMF* **9**(2), 91-107.
- [30] HOEM, J.M. (1972). Inhomogeneous semi-Markov processes, select actuarial tables, and duration-dependence in demography. *Population Dynamics* (T.N.E. Greville, editor). Academic Press, New York, 251-296.
- [31] HORIUCHI, S., AND WILMOTH, J.R. (1998). Deceleration in the age pattern of mortality at older ages. *Demography* **35**(4), 391-412.
- [32] HOUGAARD, P. (1984). Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika* **71**, 75-83.
- [33] HOUGAARD, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73**, 387-396.
- [34] HOUGAARD, P. (2000). *Analysis of Multivariate Survival Data*. Springer, New York.
- [35] HUMMER, R.A. (1996). Black-White differences in health and mortality: a review and conceptual model. *The Sociological Quarterly* **37**(1), 105-125.
- [36] JAGGER, C., AND SUTTON, C.J. (1991). Death after marital bereavement—is the risk increased? *Statistics in Medicine* **10**, 395-404.
- [37] JI, M. (2011). Markovian approaches to joint life mortality with applications in risk management. PhD Thesis, University of Waterloo, Canada.
- [38] JI, M., HARDY, M.R., AND LI, J.S.-H. (2011). Markovian approaches to joint life mortality. *North American Actuarial Journal* **15**, 357-376.
- [39] JI, M., HARDY, M.R., AND LI, J.S.-H. (2012). A Semi-Markov Multiple State Model for Reverse Mortgage Terminations. *Annals of Actuarial Science* **6** (2), 235-257.

- [40] JONES, B.L. (1994). Actuarial calculation using a Markov model. *Transactions of the Society of Actuaries* **46**, 69-77.
- [41] JONES, B.L. (1995). A stochastic population model for high demand CCRC's. *Insurance: Mathematical and Economics* **16**, 69-77.
- [42] JONES, B.L. (1996). Transient results for a high demand CCRC model. *Scandinavian Actuarial Journal*, 165-182.
- [43] JONES, B.L. (1997a). Stochastic models for continuing care retirement communities. *North American Actuarial Journal* **1**, 50-73.
- [44] JONES, B.L. (1997b). Methods for the analysis of CCRC data. *North American Actuarial Journal* **1**, 40-54.
- [45] JONES, B.L., AND WU, W. (2014). Transition modeling. In *Predictive Modeling Application in Actuarial Science: I Predictive Modeling Techniques*, Frees, E.W., Derrig, R.A. and Meyers, G. (Editors). Cambridge University Press, New York.
- [46] JUDGE, K. (1995). Income distribution and life expectancy: a critical appraisal. *British Medical Journal* **311**, 1282-1285.
- [47] KALLAN, J. (1997). Effects of sociodemographic variables on adult mortality in the United States: comparisons by sex, age, and cause of death. *Social Biology* **44**(1-2),136-147.
- [48] KESTENBAUM, B (2000). Mortality Statistics for the Oldest-Old: An Evaluation of Canadian Data. *Demographic Research* **2**(2).
- [49] KIRKWOOD, T.B.L. (2015). Deciphering death: a commentary on Gompertz (1825) On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Phil. Trans. R. Soc. B* 370: 20140379.

- [50] KNOX, D., AND TOMLIN, A (1997). An analysis of Pensioner Mortality by preretirement income. Working paper series No. 44. Centre for Actuarial Studies, University of Melbourne. Australia.
- [51] KWON, H., AND JONES, B.L. (2006). The impact of the determinants of mortality on life insurance and annuities. *Insurance: Mathematical and Economics* **38** (2), 271-288.
- [52] LANCASTER, T. (1979). Econometric methods for the duration of unemployment. *Econometrica* **47**, 939-56.
- [53] LANCASTER, T. AND NICKELL, S.J. (1980). The analysis of re-employment probabilities for the unemployed (with discussion). *J. R. Statist. Soc. A* **143**, 141-65.
- [54] LANDI, F.L., ZUCCALA, G., GAMBASSI, G., INCALZI, R.A., MANIGRASSO, L., PAGANO, F., CARBONIN, P., BERNABEI, R. (1999). Body mass index and mortality among older people living in the community. *Journal of American Geriatric Society* **47**, 1072-1076.
- [55] LANTZ, P.M., HOUSE, J.S., LEPKOWSKI, J.M., WILLIAMS, D.R., MERO, R.P., AND CHEN, J. (1998). Socioeconomic factors, health behaviors and mortality. *Journal of the American Medical Association* **279**, 1703-1708.
- [56] LIOCHEV, S.I.(2015). Which Is the Most Significant Cause of Aging? *Antioxidants* **4**, 793-810.
- [57] LIN, X.S., AND LIU, X. (2007). Markov aging process and phase-type law of mortality. *North American Actuarial Journal* **11**, 92-109.
- [58] LIN, X.S., AND LIU, X. (2012). A subordinated Markov model for stochastic mortality. *European Actuarial Journal* **2**(1), 105-127.
- [59] LOBMAYER, P., AND WILKINSON, R.G. (2000). Income inequality and mortality in 14 developed countries. Trafford Center for Medical Research, University of Sussex.

- [60] MACDONALD, A.S., WATERS, H.R., AND WEKWETE, C.T. (2005a). A model for coronary heart disease and stroke, with applications to critical illness insurance underwriting I: The model. *North American Actuarial Journal* **9** (1), 13-40.
- [61] MACDONALD, A.S., WATERS, H.R., AND WEKWETE, C.T. (2005b). A model for coronary heart disease and stroke, with applications to critical illness insurance underwriting II: Applications.. *North American Actuarial Journal* **9** (1), 41-56.
- [62] MAEGEBIER, A. (2013). Valuation and risk assessment of disability insurance using a discrete time trivariate Markov renewal reward process. *Insurance: Mathematics and Economics* **53** (3), 802-811.
- [63] MAKEHAM, W.M. (1860). On the law of mortality and the construction of annuity tables. *Journal of the Institute of Actuaries* **8**, 301-310.
- [64] MEYRICKE, R., SHERRIS, M. (2013). The determinants of mortality heterogeneity and implications for pricing annuities. *Insurance: Mathematics and Economics* **53** (2), 379-387.
- [65] MITNITSKI, A , ROCKWOOD. K (2016). The rate of aging: the rate of deficit accumulation does not change over the adult life span. *Biogerontology* **17**, 199C204.
- [66] MOLER, C., VAN LOAN, C. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review* **45** (1), 3-49.
- [67] NEUTS, M.F. (1981). *Matrix-Geometric Solutions in Stochastic Models*. Baltimore: Johns Hopkins University Press.
- [68] NEYMAN, J., PEARSON, E.S. (1933). On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Philosophical Transactions of the Royal Society of London* **231**, 289-337.
- [69] NORBERG, R. (1989). Actuarial analysis of dependent lives. *Bulletin de l'Association Suisse des Actuaries*, 243-254.

- [70] NORBERG, R. (1995). Differential equations for moments of present values in life insurance. *Insurance: Mathematics and Economics* **17**(2), 171-180.
- [71] OLSHANSKY, S.J., CARNES, B.A. (1997). Ever since Gompertz. *Demography* **34**(1), 1-15.
- [72] PAPPAS, G., QUEEN, S., HADDEN, W., AND FISHER, G. (1993). The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986. *New England Journal of Medicine* **329**(2), 103-109.
- [73] PITACCO, E. (1995a). Actuarial models for pricing disability benefits: towards a unifying approach. *Insurance: Mathematics and Economics* **16**, 39-62.
- [74] PITACCO, E. (2003). Survival models in actuarial mathematics: from Halley to longevity risk. Invited lecture at 7th International Congress INSURANCE: MATHEMATICS and ECONOMICS, ISFA, Lyon, France.
- [75] PRESCOTT, E., OSLER, M., ANDERSEN, P.K., HEIN, H.O., BORCH-JOHNSEN, K., LANGE, P., SCHNOHR, P., VESTBO, J. (1998). Mortality in women and men in relation to smoking. *International Epidemiological Association* **27**, 27C32.
- [76] PRESTON, S., AND ELO, I. (1995). Are educational differentials in adult mortality increasing in the United States? *Journal of Aging and Health* **7**, 476-496.
- [77] RAMLAU-HANSEN, H. (1988). The emergence of profit in life insurance. *Insurance: Mathematics and Economics* **7**(4), 225-236.
- [78] RAMLAU-HANSEN, H. (1991). Distribution of surplus in life insurance. *Astin Bulletin* **21** (1), 57-71.
- [79] RASTAS, S., PIIRTTILA, T., VIRAMO, P., VERKKONIEMI, A., HALONEN, P., JUVA, K., NIINISTO, L., MATTILA, K., LANSIMIES, E., AND SULKAVA, R. (2006). Association Between Blood Pressure and Survival over 9 Years in a General Population Aged 85 and Older. *Journal of The American Geriatrics Society* **54**, 912-918.

- [80] RICKAYZEN, B.D., WALSH, D.E.P. (2002). A multi-state model of disability for the United Kingdom: implications for future need for long-term care for the elderly. *British Actuarial Journal*, **8**, 341-393.
- [81] ROGERS, R.G. (1995). Marriage, sex, and mortality. *Journal of Marriage and Family* **57**, 515-526.
- [82] ROTHSCHILD, M., AND STIGLITZ, J. (1976). Equilibrium in competitive insurance markets-an essay on the economics of imperfect information. *Quarterly Journal of Economics*. 629-650.
- [83] SEAL, H.L. (1977). Studies in the history of probability and statistics. XXXV. Multiple decrements or competing risks. *Biometrika* **64**, 429-439.
- [84] SORLIE, P.D., BACKLAND, E., KELLER, J.B. (1995). U.S. mortality by economic, demographic, and social characteristics: the National Longitudinal Mortality Study. *American Journal of Public Health* **85**(7), 949-956.
- [85] SPREEUW, J. (2006). Types of dependence and time-dependent association between two lifetimes in single parameter copula models. *Scandinavian Actuarial Journal* **5**, 286-309.
- [86] SPREEUW, J., AND WANG, X. (2008). Modelling the short-term dependence between two remaining lifetimes. *Cass Business School Discussion Paper*.
- [87] SPREEUW, J., AND OWADALLY, I. (2013). Investigating the broken-heart effect: a model for short-term dependence between the remaining lifetimes of joint-lives. *Annals of Actuarial Science* **7**(2), 236-257.
- [88] SU, S, AND SHERRIS, M. (2012). Heterogeneity of Australian population mortality and implications for a viable life annuity market. *Insurance: Mathematics and Economics* **51**, 322-332.

- [89] TESSIER, J.F., NEJARI, C., LETENNEUR, L., BARBERGER-GATEAU, P., DARTIGUES, J.F., AND SALAMON, R. (2000). Smoking and eight-year mortality in an elderly cohort. *International Journal of Tuberculosis and Lung Disease* **4**, 698C704.
- [90] TOLLEY, H.D., MANTON, K.G. (1991). Intervention effects among a collection of risks. *Transactions of the Society of Actuaries* **43**.
- [91] TRAVATO, F., AND LALU, N.K. (1998). Contribution of cause-specific mortality to changing sex differences in life expectancy: seven nations case study. *Social Biology* **45**,1-20.
- [92] VAILLANT, G.E., AND MUKAMAL, K. (2001). Successful Aging. *Journal of Psychiatry* **158**(6), 839-847.
- [93] VALLIN, J. (1995). Can sex differentials in mortality be explained by socioeconomic mortality differentials? *Adult Mortality in Developed Nations*, edited by A. Lopes, G. Caselli and T. Valkonen. Pp. 179-200. Oxford: Clarendon Press.
- [94] VAUPEL, J.W., MANTON, K.G., STALLARD, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439-54.
- [95] VAUPEL, J.W., YASHIN, A.I. (1983). The deviant dynamics of death in heterogeneous populations, RR-83-1, Laxenburg, Austria: International Institute for Applied Systems Analysis.
- [96] WATERS, H.R. (1984). An approach to the study of multiple state models. *Journal of the Institute of Actuaries* **111**(2), 363-374.
- [97] WATERS, H.R. (1990). The recursive calculation of the moments of the profit on a sickness insurance policy. *Insurance: Mathematics and Economics* **9**, 101-113.
- [98] WATERS, H.R., AND WILKIE, A.D. (1988). A short note on the construction of life tables and multiple decrement models. *Journal of the Institute of Actuaries* **114**, 569-580.
- [99] WEI, M.J., KAMPERT, B., BARLOW, C.E., NICHAMAN, M.Z., GIBBONS, L.W., PAFFENBARGER JR., R.S., AND BLAIR, S.N. (1999). Relationship between low cardiorespiratory fitness and

- mortality in normal-weight, overweight and obese men. *Journal of the American Medical Association* **282**(16), 1547.
- [100] WETTERSTRAND, W.H. (1981) Parametric models for life insurance mortality data: Gompertz's law over time. *Transactions of Society of Actuaries* **33**, 159-179.
- [101] WILLIAMS, D.R., AND COLLINS, C. (1996). U.S. socioeconomic and racial differences in health: patterns and explanations. *Annual Review of Sociology* **21**, 349-386.
- [102] WOLFSON, M., ROWE, G., GENTLEMAN, J.F., AND TOMIAK, M. (1990). Earnings and Death-Effects Over a Quarter Century. *Statistics Canada*.
- [103] WOLTHUIS, H. (2003). *Life Insurance Mathematics - The Markovian Model*. IAE, Universiteit van Amsterdam, Amsterdam, second edition.

Appendix A

Appendix

Expressions for the remaining transition intensities in Section 3.3.1

$$\lambda^{21}(t | \mathcal{H}_t^3) = \frac{0.2e^{-0.06(t-\epsilon_2)-0.23\epsilon_2}(0.05)0.6 + 0.1e^{-0.12(t-\epsilon_2)-0.16\epsilon_2}(0.1)0.4}{e^{-0.06(t-\epsilon_2)-0.23\epsilon_2}(0.05)0.6 + e^{-0.12(t-\epsilon_2)-0.16\epsilon_2}(0.1)0.4}$$

$$\lambda^{21}(t | \mathcal{H}_t^4) = \frac{0.2e^{-0.23(t-\epsilon_1)-0.06\epsilon_1}(0.05)0.6 + 0.1e^{-0.16(t-\epsilon_1)-0.12\epsilon_1}(0.1)0.4}{e^{-0.23(t-\epsilon_1)-0.06\epsilon_1}(0.05)0.6 + e^{-0.16(t-\epsilon_1)-0.12\epsilon_1}(0.1)0.4}$$

$$\lambda^{13}(t | \mathcal{H}_t^1) = \frac{0.01e^{-0.06t}0.6 + 0.02e^{-0.12t}0.4}{e^{-0.06t}0.6 + e^{-0.12t}0.4}$$

$$\lambda^{13}(t | \mathcal{H}_t^2) = \frac{0.01e^{-0.23t+0.17(\epsilon_1+\epsilon_2)}(0.01)0.6 + 0.02e^{-0.16t+0.04(\epsilon_1+\epsilon_2)}(0.01)0.4}{e^{-0.23t+0.17(\epsilon_1+\epsilon_2)}(0.01)0.6 + e^{-0.16t+0.04(\epsilon_1+\epsilon_2)}(0.01)0.4}$$

$$\lambda^{23}(t | \mathcal{H}_t^3) = \frac{0.03e^{-0.06(t-\epsilon_2)-0.23\epsilon_2}(0.05)0.6 + 0.06e^{-0.12(t-\epsilon_2)-(0.1+0.06)\epsilon_2}(0.1)0.4}{e^{-0.06(t-\epsilon_2)-0.23\epsilon_2}(0.05)0.6 + e^{-0.12(t-\epsilon_2)-0.16\epsilon_2}(0.1)0.4}$$

$$\lambda^{23}(t | \mathcal{H}_t^4) = \frac{0.03e^{-0.23(t-\epsilon_1)-0.06\epsilon_1}(0.05)0.6 + 0.06e^{-0.16(t-\epsilon_1)-0.12\epsilon_1}(0.1)0.4}{e^{-0.23(t-\epsilon_1)-0.06\epsilon_1}(0.05)0.6 + e^{-0.16(t-\epsilon_1)-0.12\epsilon_1}(0.1)0.4}$$

Expressions for the rest transition intensities in Section 3.3.2

$$\lambda^{21}(t | \mathcal{H}_t^3) = \frac{\mu_2^{21} \mu_2^{12}}{\mu_1^{21} \mu_1^{12}} \exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p)$$

$$\mu_1^{21}(0.99^t) \frac{\mu_2^{12}}{\mu_1^{12}} \exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p),$$

where

$$\text{as } \epsilon_1, \epsilon_2 \rightarrow 0, s^{12} \rightarrow \int_0^t \mu^{12}(v) dv, s^{13} \rightarrow \int_0^t \mu^{13}(v) dv, s^{21} \rightarrow 0, s^{23} \rightarrow 0.$$

$$\lambda^{21}(t | \mathcal{H}_t^4) = \frac{\mu_2^{21} \mu_2^{12}}{\mu_1^{21} \mu_1^{12}} \exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p)$$

$$\mu_1^{21}(0.99^t) \frac{\mu_2^{12}}{\mu_1^{12}} \exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p),$$

where

$$\text{as } \epsilon_1, \epsilon_2 \rightarrow 0, s^{12} \rightarrow 0, s^{13} \rightarrow 0, s^{21} \rightarrow \int_0^t \mu^{21}(v) dv, s^{23} \rightarrow \int_0^t \mu^{23}(v) dv.$$

$$\lambda^{13}(t | \mathcal{H}_t^1) = \frac{\mu_2^{13}}{\mu_1^{13}} \exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p)$$

$$\mu_1^{13}(0.001)(1.1^t) \frac{\mu_2^{13}}{\mu_1^{13}} \exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p),$$

where

$$\text{as } \epsilon_1, \epsilon_2 \rightarrow 0, s^{12} \rightarrow \int_0^t \mu^{12}(v) dv, s^{13} \rightarrow \int_0^t \mu^{13}(v) dv, s^{21} \rightarrow 0, s^{23} \rightarrow 0.$$

$$\lambda^{13}(t | \mathcal{H}_t^2) = \frac{\mu_2^{13} \mu_2^{12} \mu_2^{21}}{\mu_1^{13} \mu_2^{12} \mu_1^{21}} \exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p)$$

$$\mu_1^{13}(0.001)(1.1^t) \frac{\mu_2^{13} \mu_2^{12} \mu_2^{21}}{\mu_1^{13} \mu_2^{12} \mu_1^{21}} \exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p),$$

$$\frac{\mu_2^{12} \mu_2^{21}}{\mu_1^{12} \mu_1^{21}} \exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p)$$

where

$$\text{as } \epsilon_1, \epsilon_2 \rightarrow 0, s^{12} \rightarrow 0, s^{13} \rightarrow 0, s^{21} \rightarrow \int_0^t \mu^{21}(v) dv, s^{23} \rightarrow \int_0^t \mu^{23}(v) dv.$$

$$\lambda^{23}(t | \mathcal{H}_t^3) = \frac{\mu_2^{23} \mu_2^{12}}{\mu_1^{23} \mu_1^{12}} \exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p)$$

$$\mu_1^{23}(0.001)(1.1^t) \frac{\mu_2^{23} \mu_2^{12}}{\mu_1^{23} \mu_1^{12}} \exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p),$$

$$\frac{\mu_2^{12}}{\mu_1^{12}} \exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p)$$

where

$$\text{as } \epsilon_1, \epsilon_2 \rightarrow 0, s^{12} \rightarrow \int_0^t \mu^{12}(v) dv, s^{13} \rightarrow \int_0^t \mu^{13}(v) dv, s^{21} \rightarrow 0, s^{23} \rightarrow 0.$$

$$\lambda^{23}(t | \mathcal{H}_t^4) = \frac{\mu_2^{23} \mu_2^{12}}{\mu_1^{23} \mu_1^{12}} \exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p)$$

$$\mu_1^{23}(0.001)(1.1^t) \frac{\mu_2^{23} \mu_2^{12}}{\mu_1^{23} \mu_1^{12}} \exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p),$$

$$\frac{\mu_2^{12}}{\mu_1^{12}} \exp\left(-\sum_{i,j} s^{ij} \frac{\mu_2^{ij}}{\mu_1^{ij}}\right) p + \exp\left(-\sum_{i,j} s^{ij}\right) (1-p)$$

where

$$\text{as } \epsilon_1, \epsilon_2 \rightarrow 0, s^{12} \rightarrow 0, s^{13} \rightarrow 0, s^{21} \rightarrow \int_0^t \mu^{21}(v) dv, s^{23} \rightarrow \int_0^t \mu^{23}(v) dv.$$

Curriculum Vitae

Name: Yuzhou Zhang

Post-Secondary Education and Degrees: University of Western Ontario
London, ON
2009 - 2012 HBSc

University of Western Ontario
London, ON
2012 - 2016 Ph.D.

Honours and Awards: Graduate Student Teaching Award
2013-2014

Related Work Experience: Teaching Assistant
The University of Western Ontario
2012 - 2016