Modelling of Multi-State Panel Data:

The Importance of the Model Assumptions

by

Thandile John Mafu

Thesis presented in partial fulfilment of the requirements for the

degree of Master of Commerce at Stellenbosch University

department of Statistics and Actuarial Science

Supervisor: Dr Christoffel Joseph Brand Muller

December 2014

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2014

Summary

A multi-state model is a way of describing a process in which a subject moves through a series of states in continuous time. The series of states might be the measurement of a disease for example in state 1 we might have subjects that are free from disease, in state 2 we might have subjects that have a disease but the disease is mild, in state 3 we might have subjects having a severe disease and in last state 4 we have those that die because of the disease. So Markov models estimates the transition probabilities and transition intensity rates that describe the movement of subjects between these states. The transition might be for example a particular subject or patient might be slightly sick at age 30 but after 5 years he or she might be worse. So Markov model will estimate what probability will be for that patient for moving from state 2 to state 3.

Markov multi-state models were studied in this thesis with the view of assessing the Markov models assumptions such as homogeneity of the transition rates through time, homogeneity of the transition rates across the subject population and Markov property or assumption.

The assessments of these assumptions were based on simulated panel or longitudinal dataset which was simulated using the R package named ***msm package*** developed by Christopher Jackson (2014). The R code that was written using this package is attached as appendix. Longitudinal dataset consists of repeated measurements of the state of a subject and the time between observations. The period of time with observations in longitudinal dataset is being made on subject at regular or irregular time intervals until the subject dies then the study ends.

Opsomming

'n Meertoestandmodel is 'n manier om 'n proses te beskryf waarin 'n subjek in 'n ononderbroke tydperk deur verskeie toestande beweeg. Die verskillende toestande kan byvoorbeeld vir die meting van siekte gebruik word, waar toestand 1 uit gesonde subjekte bestaan, toestand 2 uit subjekte wat siek is, dog slegs matig, toestand 3 uit subjekte wat ernstig siek is, en toestand 4 uit subjekte wat aan die siekte sterf. 'n Markov-model raam die oorgangswaarskynlikhede en -intensiteit wat die subjekte se vordering deur hierdie toestande beskryf. Die oorgang is byvoorbeeld wanneer 'n bepaalde subjek of pasiënt op 30-jarige ouderdom net lig aangetas is, maar na vyf jaar veel ernstiger siek is. Die Markov-model raam dus die waarskynlikheid dat so 'n pasiënt van toestand 2 tot toestand 3 sal vorder.

Hierdie tesis het ondersoek ingestel na Markov-meertoestandmodelle ten einde die aannames van die modelle, soos die homogeniteit van oorgangstempo's oor tyd, die homogeniteit van oorgangstempo's oor die subjekpopulasie en tipiese Markov-eienskappe, te beoordeel.

Die beoordeling van hierdie aannames was gegrond op 'n gesimuleerde paneel of longitudinale datastel wat met behulp van Christopher Jackson (2014) se R-pakket genaamd msm gesimuleer is. Die R-kode wat met behulp van hierdie pakket geskryf is, word as bylae aangeheg. Die longitudinale datastel bestaan uit herhaalde metings van die toestand waarin 'n subjek verkeer en die tydsverloop tussen waarnemings. Waarnemings van die longitudinale datastel word met gereelde of ongereelde tussenposes onderneem totdat die subjek sterf, wanneer die studie dan ook ten einde loop.

Acknowledgments

I wish to acknowledge the following people for their contribution and support to the completion of this study:

But most importantly the Almighty above us, without Him the wisdom and courage to take this journey will be not be a reality.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, an overview of the research, the aim of the study as well as the structure of the thesis are presented.

1.1 Overview of the thesis

A multi-state model is a way of describing a process in which a subject moves through a series of states in continuous time. The series of states might be the measurement of a disease for example in state 1 we might have subjects that are free from disease, in state 2 we might have subjects that have a disease but the disease is mild, in state 3 we might have subjects having a severe disease and in last state 4 we have those that die because of the disease. So Markov models estimates the transition probabilities and transition intensity rates that describe the movement of subjects between these states. The transition might be for example a particular subject/ patient might be slightly sick at age 30 but after 5 years he/she might be worse. So Markov model will estimate what probability will be for that patient for moving from state 2 to state 3. For more information please refers to chapter 2 (Multi-state models).

Markov multistate models were studied with a view to assessing the assumptions of these models, such as homogeneity of the transition rates through time, homogeneity of the transition rates across the subject population and Markov property or assumption. The assumptions were studied in details. For more details about how to assess these assumptions please refers to chapter 4 (Model assessment).

The assessments of these assumptions were based on simulated panel or longitudinal dataset which was simulated using the R package named ***msm package*** developed by Christopher Jackson (2005). The R code that was written using this package is attached as appendix. Longitudinal dataset consists of repeated measurements of the state of a subject and the time between observations. The period of time with observations in longitudinal dataset is being made on subject at regular or irregular time intervals until the subject dies then the study ends. For more information about longitudinal dataset please refers to chapter 5 (Data simulation and application).

Introduction

1.2 The aim of the thesis

Multi-state modelling has developed as the technique of choice when modelling panel or longitudinal data – data that include units that are observed across two or more points in time. A continuous time stochastic process is assumed to govern the multi-state process through its transition probabilities and transition rates. Estimating these transition probabilities or rates of the stochastic process lies at the heart of multi-state modelling. Three assumptions that are typically made regarding the transition rates before fitting a multi-state model are:

1) Homogeneity of the transition rates through time.

2) Homogeneity of the transition rates across the subject population.

3) The Markov assumption – the transition rates only depend on the history of the process through the current state.

Various authors have put forward methods to assess these assumptions before fitting a multi-state model. Unfortunately, as with many statistical techniques that have underlying assumptions, these methods are not always used to assess if these assumptions are valid before fitting a multi-state model. In this thesis, the results of a simulation study in which the importance of these three assumptions was assessed are presented. Simulated panel data sets are generated where these assumption are specifically violated. Standard multi-state model are then fitted to these data sets and the results obtained are discussed.

1.3 Structure of the thesis

Multi-state models has been discussed and explained in details in chapter 2 including its building blocks such stochastic process, transition probability and intensity matrix, Markov models, sojourn time, Model assumptions and Time homogeneous Markov model. In stochastic process the system enters a state, spends an amount of time called the sojourn time and then moves to another state where it spends another sojourn time, and so on. Transition probability and intensity matrix define probabilities and rates between the states for subject movements in the process. A Markov model is defined by a set of states as well as set of transitions with associated probabilities. In time homogeneous Markov models, all transition intensities are assumed to be constant as functions of time.

Introduction

In chapter 3, we introduced and explained the particulars of the multi-state models such as covariates as well model structures underlying in multi-state model. In this chapter we discuss in details the multi-state model features that can have significant influences in the model we fit. Model structure is defined by a set of states and a set of transitions with associated probabilities.

The Markov property and the homogeneity assumptions are strong assumptions that may lead to biased estimates if violated; therefore, it is very important to assess and further investigate a multistate model once it has been fitted to the model. The assessment of the model such as model assumptions validation, assessment of covariates effect in the model as well as model assessment using formal and informal tools has been investigated further in chapter 4.

The main purpose of this study was to assess the fit of model particular to assess or validate the Markov assumptions. In order to be able to assess those assumptions we firstly need a dataset that can be used to fit the model. Therefore with regard to this we will need to simulate a panel or longitudinal dataset that is suitable for Markov models. The last chapter 5 is concerned about simulation of dataset based on the Markov process, application of the simulated data to the model as well the representation of the results.

Chapter 2

Multi-state models

A multi-state model is a model for time to event data in which all subjects start in one state or possible more starting states and eventually may end up in one or more absorbing state(s). Alternatively it is a way of describing a process in which a subject moves through a series of states in continuous time. Some subjects are censored before they reach an absorbing state (dead state). For multi-state model a longitudinal or panel dataset is observed and investigated. A panel dataset is the one that follows a given sample of n subjects over time and provides multiple observations on each subject in the sample. Censored refers to the fact that some of subjects are dropped from the experiment which is highly expected since the subjects followed over time. Censored will cause problem in the study therefore it needs to be taken into account when modelling. When we considered a multi-state model, we want to investigate the effect of risk factors on the transitions through different states. In other words in multi-state modelling we study the relationships between the different predictors and the outcome (variable of interest). Variable of interest is the state each patient is in at each visit. Covariates must also be introduced in the model to assess their significant. In multi-state models the transition intensities (now explained in section 2.3) provide the hazards for movement from one state to another. These transitions intensities can be used to calculate the mean sojourn time in a given state. In this chapter, the stochastic process, the transition probability matrix, the transition intensity matrix, sojourn time, model assumptions, Markov models and time homogeneous Markov models are discussed in detail.

2.1 Stochastic process

A first order Markov process, $X(t)$, state that a stochastic process in which future knowledge about the process is provided only by the current state and is not altered with the additional knowledge of past states. This means that, the future state is independent of the past given the present state of the process (Ibe, 2009). That is,

$$P\left[X(t_n) \leq x_n \mid X(t_{n-1}) = x_{n-1}, X(t_{n-2}) = x_{n-2}, \dots, X(t_0) = x_0\right]$$
$$= P\left[X(t_n) \leq x_n \mid X(t_{n-1}) = x_{n-1}\right] \tag{2.1}$$

In stochastic process the system enters a state, spends an amount of time called the sojourn time (sojourn time has been discussed in section (2.5)) and then moves to another state where it spends another sojourn time, and so on. A stochastic process changes over time in an uncertain manner and its model (that is stochastic model) has five components such as time t,

Multi-state models

state *s,* activity (which depends on time), transition and stochastic process (a collection of random variables $X(t)$). The time can be either continuous or discrete parameter. The random variable in stochastic process is denoted by $X(t)$ and it represents the measurement that has been observed at the particular state at a given time for the particular subject. For an example if the study is concerned about measuring the patient's heart pulse during surgery then stochastic variable $X(t)$ will represents the occurrence of heartbeat at time *t* for that particular patient which is measured continuously. All the possible random variables $X(t)$ of stochastic process that are assumed are collected in a state space S where

$$ S \ = \ \left\{ \ s_1, s_2, s_3, \ldots, s_K \ \right\} \tag{2.2} $$

If S, is discrete, the then process is called a discrete-state stochastic process. Similarly if S is continuous, then the process is called a continuous-state stochastic process. The set of parameters of the stochastic process is denoted by T and it is usually a set of times. If T, is a countable set then the process is called a discrete-time stochastic process. If T, is an interval of real numbers then the process is called continuous-time stochastic process. If the Markov process is a discrete-time Markov process then the transitions occur at fixed points in time and we consider transition probabilities and if the Markov process is a continuous-time Markov process then the transitions can occur at any point in time and we consider transition rates.

To describe the Markov process let *S* defined above denote a set of states then

- The process moves successively from one state to another state having started in one of these states.
- If the process is currently in state *i*, then it moves to state *j* with a transition probability of p_{ij} (transition probability is discussed in the next section). The probability does not depend upon which states the process was in before the current state.
- The process can remain in the state it is in and this occurs with probability p_{ii}.
- The starting state defined in *S* is specified by the initial probability distribution, this is done by specifying a particular state as the starting state.

The absorbing state *i* of a Markov process is the state *i* in which the process will never leave that state for an example state 2 (Dead state) in 2-state Markov model in chapter (3) section

Multi-state models

(3.2) Figure (3.1). The Markov process is absorbing if it has at least one absorbing state and if from every state it is possible to go to an absorbing state. In a Markov process the state that is not absorbing is called transient. The first passage time of a certain state s_i in S is the time t at which $X(t) = s_i$ for the first time since the start of the process. The time of absorption of an absorbing state is the first passage time of that state. The recurrence time is the first time t at which the process has returned to its initial state. As the process progress over time $t,$ the history of the observation of the process over the interval $[0,t)$ will be generated, for an example the states previously visited, times of transitions etc.

2.2 Probability transition matrix

The transition probability matrix is the *K x K* matrix whose entry in *row i* and *column j* is the transition probability $P(t)$ and is denoted by

$$P(t) = \begin{bmatrix} P_{11}(t) & P_{12}(t) & \dots & P_{1K}(t) \\ P_{21}(t) & P_{22}(t) & \dots & P_{2K}(t) \\ \vdots & \vdots & \ddots & \vdots \\ P_{K1}(t) & P_{K2}(t) & \dots & P_{KK}(t \end{bmatrix} \qquad (2.3)$$

$P(t)$, denote transition probability matrix of a multi-state process at time t. The transition probability matrix (2.3) is a stochastic matrix because for any row *i,*

$$\sum_{j} p_{ij} = 1. \qquad (2.4)$$

The entries of probability transition matrix (2.3) are defined in (2.26) and these entries define transition/movement probabilities of subjects through states. The matrix $P(t)$ (2.3) is the transition probability matrix with its elements gives the probability of being in state *j* at time *t + s*, conditional on being in state *i* at time *s*. The transition is the movement from one state to another. The matrix P is time dependent and to emphasize that, the transition probability matrix should be denoted as $P(t)$ but in time homogeneous intensities the dependence of P on time will be omitted. In every transition probability matrix the probabilities must be greater than or equal to zero, and each row must sum to one that is

Multi-state models

$$P_{ij} \geq 0 \quad for \quad all \quad i, j \in \{1, \ldots, K\} \tag{2.5}$$

and

$$\sum_{j=1}^{K} P_{ij} = 1 \quad for \quad all \quad i, j \in \{1, \ldots, K\}, \tag{2.6}$$

To illustrate the transition probability matrix above let's use practical example where the transition probability matrix is assumed to be

$$P(t) = \begin{bmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0 & 0.5 \\ 0.25 & 0.25 & 0.5 \end{bmatrix} \tag{2.7}$$

This is a 3-state model (model structures has been discussed in chapter (3) with rainy, nice and snow states respectively. As indicated above, each row must sum to one for an example row 1 summation equals to 1 (0.5 + 0.25 + 0.25 = 1) and each probability must be greater than or equal to zero ($P_{12} = 0.25$).

In case of an n-step state transition probability matrix, let $p_{ij}(n)$ denote the conditional probability that the process will be in state j after exactly n transitions, given that it is presently in state i (Ibe, 2009). That is,

$$p_{ij}(n) = P[X_{m+n} = j \mid X_m = i]$$
$$p_{ij}(0) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{2.8}$$
$$p_{ij}(1) = p_{ij},$$

To illustrate this lets consider two-step transition probability $p_{ij}(2)$, which is defined by

$$p_{ij}(2) = P[X_{m+2} = j \mid X_m = i] \tag{2.9}$$

if $m = 0$, then

Multi-state models

$$p_{ij}(2) = \sum_{k} p_{kj} p_{ik}$$
$$= \sum_{k} p_{ik} p_{kj}.$$

(2.10)

The summation is taken over all possible intermediate states *k*. This means that the probability of starting in state *i* and being in state *j* at the end of the second transition is the probability that we first go immediately from state *i* to an intermediate state *k* and then immediately from state *k* to state *j*. The $p_{ij}(n)$, is the *ijth* entry (that is *ith row, jth column*) in the matrix P^n. That is,

$$P^n = \begin{bmatrix} p_{11}(n) & p_{12}(n) & \dots & p_{1N}(n) \\ p_{21}(n) & p_{22}(n) & \dots & p_{2N}(n) \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1}(n) & p_{N2}(n) & \dots & p_{NN}(n) \end{bmatrix}$$

(2.11)

where *N*, represent the number of state. If *n* is equal to 1 then the above matrix is called *one-step* probability matrix.

The *n*-step transition probabilities can be obtained by multiplying the transition probability matrix by itself *n* times. To illustrate this, let

$$P = \begin{bmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \end{bmatrix}$$

(2.12)

Then

$$P^2 = P \times P = \begin{bmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \end{bmatrix} \times \begin{bmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \end{bmatrix}$$
$$= \begin{bmatrix} 0.4 & 0.6 \\ 0.36 & 0.64 \end{bmatrix}$$

(2.13)

The P^2, is the *2*-step transition probability matrix obtained using the definition in (2.2.2) above. From the *2*-step transition probability matrix P^2, we obtain $p_{11} = 0.4$, $p_{12} = 0.6$, $p_{21} = 0.36$

Multi-state models

and $p_{22} = 0.64$ entries. The *n*-step transition probability matrix $p_{ij}(n)$ does not depend on i as $n \to \infty$. This means that $P[X(n) = j]$ approaches a constant as $n \to \infty$. If the limit exists in the Markov chains the limiting-state probabilities is defined as

$$\lim_{n \to \infty} P[X(n) = j] = \pi_j \quad j = 1,2,\ldots,N . \tag{2.14}$$

If the limiting-state probabilities exist and do not depend on the initial state, then we have

$$\lim_{n \to \infty} p_{ij}(n) = \pi_j = \lim_{n \to \infty} \sum_k p_{ik}(n-1) p_{kj}$$
$$= \sum_k \pi_k p_{kj} \tag{2.15}$$

Letting the limiting-state probability vector $\pi = [\pi_1, \pi_2, \ldots, \pi_N]$, results in

$$\pi_j = \sum \pi_k p_{kj}$$
$$\pi = \pi P \tag{2.16}$$
$$\sum_j \pi_j = 1.$$

If each column of transition probability matrix sum to 1 then the transition probability matrix is defined to be a doubly stochastic matrix, That is,

$$\sum_i p_{ij} = 1. \tag{2.17}$$

This means that apart from each row sum to 1 also each column must sum to 1. If the transition probability matrix is a doubly stochastic matrix with the transition probabilities of a Markov chain with *N* states, then the limiting-state probabilities are defined by

$$\pi_i = \frac{1}{N}, \quad i = 1,2,\ldots,N. \tag{2.18}$$

To illustrate the doubly stochastic matrix, let *P* be defined as

$$P = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} \tag{2.19}$$

Multi-state models

From this transition probability matrix *P,* it can be seen that each column sum to 1 and also each row sum to 1. The limiting-state probabilities exist and are given by

$$\pi_1 = \pi_2 = \frac{1}{2},$$ (2.20)

since *N*= 2.

2.3 Transition intensity matrix

The intensity between two states *i* and *j,* is the rate of change of the probability P_{ij} in a very small time interval Δt. For the formal definition of intensity from state *i* to state *j* at time *t* please refer to definition (2.28) and also the entries in transition intensity matrix are defined by (2.28). All possible intensities between the various states are collected in the transition intensity matrix which is denoted by Q with dimension of *(K x K)*. For example, for the *K* states the transition intensity matrix would be

$$Q(\lambda) = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1K} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{K1} & \lambda_{K2} & \dots & \lambda_{KK} \end{bmatrix}$$ (2.21)

The parameter λ in (2.21) represents independent parameters and it is a vector of length *b*. $Q(\lambda)$ denote transition intensity matrix of a multi-state process. The transition intensity matrix (2.21) is used to define the multi-state model. The transition intensity matrix (2.21) again is also used to calculate the transition probability matrix (2.3) but definition of (2.3) is a complicated function of *Q*. So definition of (2.48 in chapter 2) can be used to calculate $P(t)$ for given *Q*. The elements in each row of the transition intensity matrix (2.21) must sum to zero and off diagonal elements must be non-negative that is

$$\sum_{j=1}^{K} \lambda_{ij} = 0$$ (2.22)

Multi-state models

and $\lambda_{ij} \geq 0$ *for* $i \neq j$ respectively. The elements in diagonal must be negative for all i is not

equal to j that is

$$\lambda_{ii} = -\sum\nolimits_{i \neq j} \lambda_{ij} \quad for \quad i = 1, \ldots, K. \tag{2.23}$$

This implies that subjects in those states remain in their state while the off diagonals are rates in which subjects move to other states. The Q matrix (2.21) is called the transition intensity or rate matrix where each element (that is λ_{ij}) represent rate at which transitions are made from state i to state j. For example, let $K = 3$, be the number of states of interest then to illustrate the conditions or constraints mentioned above for transition intensity matrix (2.21) we use the following transition intensities *3 x 3* matrix

$$Q(\lambda) = \begin{bmatrix} -(\lambda_{12} + \lambda_{13}) & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & -(\lambda_{21} + \lambda_{23}) & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & -(\lambda_{31} + \lambda_{32}) \end{bmatrix} \tag{2.24}$$

The off diagonals elements in transition intensity matrix (2.24) are rates at which subjects move into other states, while the diagonals elements are rates at which subject remain in their state that is no progress to other state.

2.4 Markov models

A Markov model is defined by a set of states as well as set of transitions with associated probabilities. A Markov model is a multi-state model where the multi-state model is defined as a model for a stochastic process ($X(t), t \in T$) with a finite space

$$S = \left\{ s_1, s_2, s_3, \ldots, s_N \right\} \tag{2.25}$$

and the multi-state process between the states is fully governed by a continuous time stochastic process (stochastic process has been discussed above in section 2.1) which is characterised through the transition probabilities between different states (Meira-Mechado, 2009)

$$P_{ij}(s, t, F_t) = P\{X(t) = j \mid X(s) = i, F_s\} \tag{2.26}$$

Multi-state models

Definition (2.4.1) can also written as follows

$$p_{ij} = \Pr\left\{ state \quad j \quad at \quad t \mid state \quad i \quad at \quad time \quad 0, F_s \right\} \tag{2.27}$$

where F_s is the history of the observation of the process over the interval $[0,t)$ that is generated and for $j, i \in S, s \leq t$. $X(t)$ in definition (2.4.1) denote the state being occupied at time t. Definition (2.26) denotes the probability of going to state j from state i in a period of time t. The transitions between the transient states occur with rates λ_{ij} defined by

$$\lambda_{ij} \quad = \quad \lim_{\Delta t \to 0} \frac{P\{X(t, t+\Delta t) = j \mid X(t) = i, F_t\}}{\Delta t} \tag{2.28}$$

Alternatively definition (2.28) can be written as follows

$$\lambda_{ij} = \lim_{dt} \frac{\Pr\left\{ transition \quad i \to j \quad in \quad [t, t+dt] \mid state \quad i \quad at \quad t, F_t \right\}}{dt} \tag{2.29}$$

Definition (2.28) means that a subject in state i at time t will have moved to state j $(j \neq i)$ by time $t + \Delta t$ with probability $\lambda_{ij}(t)\Delta t$, and a subject in state i at time t will have moved out of the system (died) by time $t + \Delta t$ with probability $\lambda_{0j}(t)\Delta t$. The intensity represents the instantaneous risk of moving from state i to state j and both (2.26 & 2.28) depends on the history. The next state to which the individual moves, and the time of change, are governed by a set of transition intensities (2.28) for each pair of states i and j. The intensities may also depend on the time of the process t or time-varying explanatory variables F_t.

The Markov assumption (this assumption is discussed in next section 2.6) is implicitly present in definition (2.28).We estimate the transition probability matrix (2.3) from transition intensity matrix (2.21) using maximum likelihood estimation method (discussed in section 2.7.3) in order to fit the multi-state model to data and in this thesis will focus on time homogenous Markov models. The transition rate matrix is recovered from the data then we can derive transition probability matrix for any t we choose from the given transition intensity matrix rate. If transitions occurs at fixed points in time (discrete-time Markov chains) and then we work

Multi-state models

with transition probabilities. If transitions occurs at any point in time (continuous-time Markov chains) and then we work with transition rates.

2.5 Sojourn time

In sojourn time the random variable (that is time spent by process X in the given subset of the state space in its *nth* visit to the subset) is considered. Therefore the sojourn time of a process X in a subset of states will be an integer-valued random variable if X is a chain or real-valued one in the case of a continuous-time process (Rubino and Sericola, 1988). Sojourn time is the length of time the process X remains in the state being occupied at the time *t*. The sojourn times of a continuous-time Markov process in a state j are independent, exponential (geometrically distributed in case of discrete Markov process) random variables with mean

$$-1/\lambda_{ii} \tag{2.30}$$

or rate given by $-\lambda_{ii}$ and it can be expressed in terms of passage times between states in continuous-time Markov and semi-Markov chains (Cinlar, 1975).

The other remaining elements of the *ith* row of transition intensity matrix (2.24) are proportional to the probabilities governing the next state after i to which the individual makes a transition. The probability that the subject's next move from state i to state j is

$$-\lambda_{ij}/\lambda_{ii}. \tag{2.31}$$

The sojourn time and the new state depend only on state i and not on the history of the system prior to time *t*. Given that the current state is *i*, the sojourn time and the new state are independent of each other. Mean sojourn times describe the average period in a single stay in a state for an example we may want to forecast the total time spent healthy or diseased before death. To illustrate sojourn time and conditional probabilities consider the following transition intensity matrix

$$Q = \begin{bmatrix} -0.2 & 0.2 & 0 \\ 0.2 & -0.7 & 0.5 \\ 0 & 0.8 & -0.8 \end{bmatrix} \tag{2.32}$$

Multi-state models

The transition intensity matrix (2.32) is a 3-state model and a subject that is currently occupying state 1 can only progress to state 2. A subject that is currently occupying state 2 can progress to state 1 or state 3. A subject that is currently in state 3 can make a move to state 2. The time the subject spends in state 1 before moving to state 2 (sojourn time) is

$$-1/\lambda_{ii} = -1/-0.2 = 5 \qquad (2.33)$$

units of time and if observation times are measured in years then this means that it would be 5 years. The time the subject spends in state 2 before moving to state 1 or state 3 is

$$-1/\lambda_{ii} = -1/-0.7 = 1.14 \qquad (2.34)$$

units of time that is a year and almost 2 months. The time a subject spends in state 3 before progressing to state 2 is

$$-1/\lambda_{ii} = -1/-0.8 = 1.25 \qquad (2.35)$$

units of time that is a year and almost 3 months.

The conditional probability that a subject currently in state 2 can move to state 1 is

$$-\lambda_{ij}/\lambda_{ii} = -0.2/-0.7 = 0.29 \qquad (2.36)$$

and the conditional probability that a subject currently in state 2 can move to state 3 is

$$-\lambda_{ij}/\lambda_{ii} = -0.5/-0.7 = 0.71. \qquad (2.37)$$

The conditional probability that a subject currently in state 3 can move to state 2 is

$$-\lambda_{ij}/\lambda_{ii} = -0.8/-0.8 = 1.0. \qquad (2.38)$$

The conditional probability that a subject currently in state 1 can move to state 2 is

$$-\lambda_{ij}/\lambda_{ii} = -0.2/-0.2 = 1.0. \qquad (2.39)$$

The mean sojourn times and conditional probabilities for the above transition intensity matrix (2.5.1) are summarised in the following matrix

Multi-state models

$$S/P = \begin{bmatrix} 5 & 1.0 & 0 \\ 0.29 & 1.14 & 0.71 \\ 0 & 1.0 & 1.25 \end{bmatrix} \qquad (2.40)$$

Above matrix (2.5.2) denote the sojourn/probability matrix (S/P) where the diagonal values represent the mean sojourn time and the off-diagonal values represents the conditional probabilities. From the above matrix (2.5.2) we can see that subjects in state 1 take longer time (5 years) to progress to state 2, from state 2 to state 3 take a year and one month.

2.6 Model assumptions

Different model assumptions can be made about the dependence of the transition rates on time (Meira-Macado, 2009). Markov property and the homogeneity assumptions are strong assumptions which may lead to biased estimates if violated, therefore it is very important to assess and further investigate a multi-state model once it has been fitted to the model (model assumptions assessment has been discussed in chapter 4). These assumptions include the following ones:

2.6.1 Markov model assumption

The Markov assumption state that the future progress only depends on the current state not on the past states and the current state should include all relevant history. This means that the transition times from each state are independent of the history of the process prior to entry to that state. To put it in simple terms Markov assumption simple means that to make the best possible prediction of what happens "tomorrow", we only need to consider what happens "today", as the "past" (yesterday) gives no additional useful information. The past history of a system plays no role in its future evolution, which is usually known as the "memoryless property of a Markov process" (Barbu & Limnios (2008)). This assumption applies to both discrete and continuous data. The Markov assumption is implicitly present in definition (2.28). The definition (2.26) and (2.28) can be simplified as

$$P_{ij}(s_i, t_i, F_j) = P_{ij}(s,t) = P\{X(t) = j \mid X(s) = i\} \qquad (2.41)$$

and

Multi-state models

$$\lambda_{ij}(t,F_t) = \lambda_{ij}(t) = \lim_{\Delta t \to 0} \frac{P\{X(t,t+\Delta t) = j \mid X(t) = i\}}{\Delta t} \tag{2.42}$$

where $\lambda_{ij}(t,F_t)$ is the transition rate of a multi-state process. In other words it is the instantaneous hazard/risk rate of progressing from state i to state j at time t, given the history F_t.

2.6.2 Semi-Markov assumption

The semi- Markov assumption state that the future progress not only depends on the current state i, but also on the entry time into the current state j (Meira-Macado, 2009). The definition (2.26) and (2.28) under this assumption can be simplified as

$$P_{ij}(s_i,t_i,F_{jt}) = P_{ij}(s,t,t_t) = P_{ij} = P\{X(t) = j \mid X(s) = i,t_i\} \tag{2.43}$$

and

$$\lambda_{ij}(t,F_t) = \lambda_{ij}(t_i) = \lim_{\Delta t \to 0} \frac{P\{X(t,t+\Delta t) = j \mid X(t) = i,t_i\}}{\Delta t} \tag{2.44}$$

2.6.3 Time homogeneous assumption

Under this assumption intensities are constant over time, that is, independent of time t. This means the mechanism that is chosen to decide which transition to take is the same at all times. This assumption can be assessed with a likelihood ratio test. The definition (2.26) and (2.28) can be simplified as

$$P_{ij}(s,t,F_t) = P_{ij}(0,t-s) = P\{X(t-s) = j \mid X(0) = i,t_i\} = p_{ij}(t-s) \tag{2.45}$$

and

$$\lambda_{ij}(t,F_t) = \lambda_{ij} = \lim_{\Delta t \to 0} \frac{P\{X(\Delta t) = j \mid X(0) = i\}}{\Delta t} \tag{2.46}$$

2.7 Time homogeneous Markov model

In time homogeneous Markov models, all transition intensities are assumed to be constant as functions of time, that is, independent of time t, see section (2.6). This assumption can be assessed with a likelihood ratio test (model assumptions assessment has been discussed in chapter 4. When intensities are treated as being time homogeneous then the dependency on

time can be removed. The transition probability matrix and transition intensity matrix discussed in section (2.2 and 2.3 respectively) form the building block of Kolmogorov equations that are used to yield unique solutions for probability matrix $P(t)$.

2.7.1 Kolmogorov equations

The Kolmogorov equations are used to derive the relationship between the transition intensity matrix Q and the transition probability matrix P. In other words the transition probabilities can be calculated from the intensities by solving the Kolmogorov differential equation. The relationship between the transition intensity and probability matrix involves canonical decomposition. The canonical decomposition was discussed by Kalbeisch and Lawless (1985). The Kolmogorov equations state that

$$\frac{\partial}{\partial t} P(t) = P(t) \ Q, \tag{2.47}$$

which yield unique/closed form solutions for $P(t)$ and conditional on $P(0) = I$,

$$P(t) = e^{Qt} = \sum_{r=0}^{\infty} \frac{(Qt)^r}{r!} \tag{2.48}$$

Definition (2.48) is only valid with time homogeneous intensities. Q is the transition intensity matrix therefore P can be found from Q using Kolmogorov equations (2.48). The solution for the transition probabilities in terms of the transition intensities can be found using (2.53) but the solutions are complicated functions of the intensities and it is only practical to calculate them for simple models with small intensities that is Q's. For example consider a progressive model (3-state 2-parameter model) where subjects can move only forward through the states. The last state is an absorbing state where subject cannot leave that state once entered it.

$$Q(\lambda) = \begin{bmatrix} -\lambda_{12} & \lambda_{12} & 0 \\ 0 & -\lambda_{23} & \lambda_{23} \\ 0 & 0 & 0 \end{bmatrix} \tag{2.49}$$

For an example, the probability that a subject currently in state 1 at time 0 will be in state 3 at time t ($P_{13}(t)$) is given by

Multi-state models

$$P_{13}(t) = \frac{1}{\lambda_{12} - \lambda_{23}} \left(\lambda_{12} - \lambda_{23} - \lambda_{12} e^{-t\lambda_{23}} + \lambda_{23} e^{-t\lambda_{12}} \right) \tag{2.50}$$

2.7.2 Eigenvalue Decomposition for *Q*

Solving (2.7.1.1) without the need to directly express the transition probabilities as functions of the transition rates can be accomplished with a canonical decomposition of *Q* (Kalbfleisch and Lawless, 1985). Let $d_i, \ldots, d_k$ be the distinct eigenvalues of *Q* and *A* be a *K x K* matrix with *jth* column the right eigenvector corresponding to d_j, then

$$Q = ADA^{-1}, \tag{2.51}$$

where

$$D = diag(d_i, \ldots, d_k), \tag{2.52}$$

$$P(t) = A \ diag\left(e^{d_1 t}, \ldots, e^{d_k t}\right) \ A^{-1}. \tag{2.53}$$

The transition matrix $P(t)$ is related to the intensity matrix $Q(\lambda)$ by $P(t) = \exp(Q(t))$. Definition (2.53) is the relationship between transition probability matrix and transition intensity matrix. To illustrate transition probability matrix (2.53), let the transition intensity matrix be defined as

$$Q = \begin{bmatrix} -\lambda_{12} & \lambda_{12} \\ \lambda_{21} & -\lambda_{21} \end{bmatrix} \tag{2.54}$$

Let $\lambda_{12} = 3$ and $\lambda_{21} = 1$ be the parameters associated with the transition intensity matrix defined in (2.54) then

$$Q = \begin{bmatrix} -3 & 3 \\ 1 & -1 \end{bmatrix} \tag{2.55}$$

The eigenvalues of (2.55) are (0,-4) and the associated eigenvalues are $[1,1]^t$ and $[-3,1]^t$.

Multi-state models

Therefore

$$A = \begin{bmatrix} 1 & -3 \\ 1 & 1 \end{bmatrix} \tag{2.56}$$

and

$$A^{-1} = \begin{bmatrix} 0.25 & 0.75 \\ -0.25 & 0.25 \end{bmatrix} \tag{2.57}$$

Then

$$P(t) = e^{Qt} = \begin{bmatrix} 0.25 + (0.75) \; e^{-4t} & 0.75 - (0.75) \; e^{-4t} \\ 0.25 - (0.25) \; e^{-4t} & 0.75 + (0.25) \; e^{-4t} \end{bmatrix} \tag{2.58}$$

where

$$e^{Qt} = \sum_{r=0}^{\infty} \frac{t^r \left(ADA^{-1} \right)^r}{r!} = A \left(\sum_{r=0}^{\infty} \frac{t^r D^r}{r!} \right) A^{-1} = A e^{Dt} A^{-1}. \tag{2.59}$$

To estimate the maximum likelihood estimates of parameters, the transition probabilities derivatives are required and are calculated in a similar way to (2.53). The matrix with entries $\partial p_{ij}(t;\lambda)/\partial \lambda_u$ is obtain as

$$\frac{\partial P(t)}{\partial \lambda_u} = A V_u A^{-1}, \quad u = 1, \ldots b, \tag{2.60}$$

with b the number of independent transition rates and V_u a $K \, x \, K$ matrix with (i,j) entry

$$\begin{aligned} & \frac{g_{ij}^{(u)} \left(e^{d_i t} - e^{d_j t} \right)}{\left(d_i - d_j \right)}, \quad i \neq j, \\ & g_{ii}^{(u)} t e^{d_i t}, \quad i = j, \end{aligned} \tag{2.61}$$

and g_{ij}^u the (i,j) entry in $G^{(u)} = A^{-1} \left(\partial Q/\partial \lambda_u \right) A$ (Kalbfleisch and Lawless, 1985).

2.7.3 Maximum likelihood estimation

The method of maximum likelihood estimation enables the unknown parameters in the model to be estimated. The maximum likelihood estimate is the number of transitions from state i to state j divided by number of overall transitions from state i to other states calculated from the transition probability matrix. Maximum likelihood estimates for a particular class of a model can be computed from transition probability matrix $P(t)$ (2.3), with (i,j) entry defined in (2.26) which depends on unknown parameters in Q (2.21) through the Kolmogorov relationship $P(t) = \exp(tQ)$ (Cox and Miller, 1965). Suppose we have the following transition intensity matrix Q

$$Q(\lambda) = \begin{bmatrix} -\lambda_{12} & \lambda_{12} & 0 \\ \lambda_{21} & -(\lambda_{21} + \lambda_{23}) & \lambda_{23} \\ 0 & \lambda_{32} & -\lambda_{32} \end{bmatrix} \tag{2.62}$$

Let $\lambda = (\lambda_{12}, \lambda_{21}, \lambda_{23}, \lambda_{32})$ denote the vector of intensities and the aim is to maximize the likelihood to obtain estimates of λ. To obtain the maximum likelihood estimates of λ, is accomplished by having the first and second derivatives of the likelihood function by considering the values of log-likelihood on grids of points. Let $t_0 < t_1 < \ldots < t_m$ be the observation times for individuals in the sample and n_{ijl} be the number of individuals in state i at t_{l-1} and in state j at t_l, then the likelihood and log-likelihood functions are defined as

$$L(\lambda) = \prod_{l=1}^{m} \left\{ \prod_{i,j=1}^{k} p_{ij}(t_{l-1}, t_l)^{n_{ijl}} \right\}, \tag{2.63}$$

$$\log(L(\lambda)) = \sum_{i=1}^{m} \sum_{i,j=1}^{k} n_{ijl} \log p_{ij}(t_l - t_{l-1} \mid \lambda), \tag{2.64}$$

(Kalbfleisch and Lawless, 1985), where λ is defined as the vector of b independent unknown transition intensities defined in (2.21) and definition (2.64) can be viewed as the general form for any multi-state model and can be modified based on the type of data under study.

Multi-state models

The general form needs to be modified under the following conditions (Jackson, 2014):

- Death state exist in the model
- Exactly observed transition times
- Censoring exist in the data

The above conditions as well as Quasi-Newton (or scoring) procedure are further discussed below but firstly we start with the full likelihood.

2.7.3.1 The full-likelihood

Suppose *i* indexes *n* individuals in the dataset. The data for individual *i* consists of a series of time points $(t_{i0},\ldots,t_{in_i})$ and corresponding states at these time points $(S(t_{i1}),\ldots,S(t_{in_i}))$. An individual's contribution to the likelihood is his or her path through the different states (Jackson et al., 2003). Consider an observed pair of states, $S(t_j)$ and $S(t_{j+1})$, at times t_j,t_{j+1}. Then the contribution to the likelihood from these two states is

$$L_{i,j} = Ps(t_i),s(t_{i+1})(t_{j+1}-t_j) \tag{2.65}$$

The *(i,j)* entry of (2.3) evaluated at $t=t_{j+1}-t_j$. The full-likelihood is the product of all such terms $L_{i,j}$ over all individuals and transitions which depend on the unknown transition matrix *Q,* which was used to determine $P(t)$.

2.7.3.2 Death state exist in the model

In studies where there is a death state, it is common to know the time of death, but the previous state before the death state is not always known. Let

$$S(t_{j+1}) = D \tag{2.66}$$

be a death state, and then the contribution to the likelihood is summed over the unknown states *m* on the day before death

$$L_{i,j} = \sum_{m\neq D} Ps(t_i),m(t_{j+1}-t_j)\ \lambda_{mD} \tag{2.67}$$

The sum is taken over all possible states *m* which can be visited between $S(t_j)$ and *D*.

2.7.3.3 Exactly observed transition times

If the times are exact transition times between the states, with no transitions between the observation times, then the contribution to the likelihood is

$$L_{i,j} = Ps(t_i)s(t_i)\big(t_{j+1} - t_j\big)\lambda s(t_i)s(t_{i+1}), \tag{2.68}$$

since the interval stays in state $S(t_j)$ in the interval t_j to t_{j+1} with a known transition at time t_{j+1}.

2.7.3.4 Censoring exist in the data

If at the end of the study, it is known that a subject is alive but not in what state that subject is in, that observation has to be treated as a censored observation. The contribution to the likelihood of a censored observation is

$$L_{i,j} = \sum_{m \in C} Ps(t_i), m\big(t_{j+1} - t_j\big), \tag{2.69}$$

with C defined as the known subset of states that the subject could have entered before being censored.

2.8 Quasi-Newton (or scoring) procedure

A quasi-Newton (or scoring) procedure is implemented to obtain the maximum likelihood estimates of λ and estimates of the asymptotic covariance matrix. This procedure was proposed by Kalbfleisch and Lawless (1985). Let $w_l = t_l - t_{l-1}$, where $l = 1, \ldots, m,$ then from (2.67) the first and second derivatives of the log likelihood is given as

$$S_u(\lambda) = \frac{\partial \log L}{\partial \lambda_u} = \sum_{l=1}^{m} \sum_{i,j=1}^{k} n_{ijl} \frac{\partial p_{ij}(w_l)/\partial \lambda_u}{p_{ij}(w_l)}, \quad u = 1, \ldots, b \tag{2.70}$$

$$\frac{\partial^2 \log L}{\partial \lambda_u \partial \lambda_v} = \sum_{l=1}^{m} \sum_{i,j=1}^{k} n_{ijl} \left\{ \frac{\partial^2 p_{ij}(w)/\partial \lambda_u \partial \lambda_v}{p_{ij}(w_l)} - \frac{\partial p_{ij}(w_l)/\partial \lambda_u \partial p_{ij}(w_l)/\partial \lambda_v}{p_{ij}^2(w_l)} \right\}. \tag{2.71}$$

Instead of directly using a Newton-Raphson algorithm and thus evaluating the first and second derivatives, a scoring device is used were the second derivatives are replaced by estimates of their expectations. This gives an algorithm that only requires the first derivatives of the log-

Multi-state models

likelihood. Let $N_i(t_{l-1}) = \sum n_{ijl}$ denote the number of individuals in state i at time $t_{i=1}$. Taking the expectation of n_{ijl} conditional on $N_i(t_{l-1})$ and noting that

$$\sum_{i,j=1}^{k} \frac{\partial^2 p_{ij}(w)}{\partial \lambda_u \partial \lambda_v} = 0, \text{ gives } E\left(-\frac{\partial^2 \log L}{\partial \lambda_u \partial \lambda_v}\right) = \sum_{l=1}^{m}\sum_{i,j=1}^{k} \frac{E\{N_i(t_{l-1})\}}{p_{ij}(w_l)} \frac{\partial p_{ij}(w_l)}{\partial \lambda_u} \frac{\partial p_{ij}(w_l)}{\partial \lambda_v}. \qquad (2.72)$$

This can be estimated by

$$M_{uv}(\lambda) = \sum_{l=1}^{m}\sum_{i,j=1}^{k} \frac{N_i(t_{l-1})}{p_{ij}(w_l)} \frac{\partial p_{ij}(w_l)}{\partial \lambda_u} \frac{\partial p_{ij}(w_l)}{\partial \lambda_v}. \qquad (2.73)$$

The $p_{ij}(w_l)$ and $\partial p_{ij}(w_l)/\partial \lambda_u$ terms in (2.70) and (2.73) are computed using (2.53) and (2.60)

To obtain an estimate of λ using (2.70) and (2.73), let λ_0 be an initial estimate of λ, $S(\lambda)$ be the $b \times 1$ vector $(S_u(\lambda))$ and $M(\lambda)$ be the $b \times b$ matrix $(M_{uv}(\lambda))$. An updated estimate λ_1 is obtained as

$$\lambda_1 = \lambda_0 + M(\lambda_0)^{-1} S(\lambda_0), \qquad (2.74)$$

Where it is assumed that $M(\lambda_0)$ is nonsingular. This process is repeated with λ_1 replacing λ_0 and with a good initial estimate, this produces λ upon convergence (Kalbfleisch and Lawless, 1985).

2.9 Semi-Markov process

The Markov assumption state that the future progress only depends on the current state not on the past states and the current state should include all relevant history. But this assumption imposes restrictions on the distribution of the sojourn time in a state, which should be exponentially distributed in case of continuous-time Markov process and geometrically distributed in case of a discrete-time Markov process. To overcome this, the Markov assumption must be relax in order to allow arbitrarily distributed sojourn times in any state and still have the Markov assumption but in a more flexible manner. The resulted process based on these two properties is called a semi-Markov process. A semi-Markov process is concerned with the random variables that describe the state of the process at some time and it is also a generalization of the Markov process. A semi-Markov process is a process that makes transitions from state to state like a Markov process, however the amount of time spent in each

Multi-state models

state before a transition to the next state occurs is an arbitrary random variable that depends on the next state the process will enter (Ibe, 2009). The semi-Markov chain can be described as follows;

- The initial state i_0 is chosen according to the initial distribution λ,

- Then next visited state i_1 is determined according to the transition probability matrix p.

- And the chain stays in state i_0 for a time t determined by the sojourn time distribution in state i_0 before going to state i_1.

2.9.1 Discrete-Time Semi-Markov processes

In a discrete-time Markov process, the assumption is made that the amount of time spent in each state before a transition to the next state occurs is a unit time (Ibe, 2009). Let the finite-state discrete-time random process be denoted by

$$\{X_n \mid n = 0,1,2,\cdots K\} \tag{2.75}$$

Here K reflects the number of states and let the state space be denoted by

$$S = \{0,1,2,\cdots,K\} \tag{2.76}$$

Let the probability of transitions between the two states be denoted by p_{ij}, where

$$\sum_{j=0}^{K} p_{ij} = 1$$
$$p_{ij} \geq 0 \quad i,j \in S \tag{2.77}$$

The above conditions were also discussed in section (2.2). Let $T_0, T_1, T_2, \cdots, T_K$ denote the transition periods on the nonnegative real line such

$$0 = T_0 \leq T_1 \leq T_2 \leq \cdots \leq T_K. \tag{2.78}$$

Multi-state models

Let the interval be defined by

$$W_i = T_{i+1} - T_i. \tag{2.79}$$

This refers to waiting time or holding time in state $i \in S$; before making the transition the process spends a waiting time W_{ij}. The W_{ij} is a positive, integer-valued random variable with the

$$p_{W_{ij}}(r) = P[W_{ij} = r], r = 1,2,\cdots \tag{2.80}$$

It is assumed that the system spends at least one unit of time before making a transition that is

$$E(W_{ij}) < \infty$$
$$p_{W_{ij}}(0) = 0 \tag{2.81}$$

for all *i* and *j*. By ignoring the times between transitions and focus only on the transitions then the resulting process will be Markov. If we include the waiting times then the process will no longer satisfy the Chapman-Kolmogorov equation. Thus

$$p_{Wi}(r) = \sum_{j=0}^{K} p_{ij} p_{W_{ij}}(r) \quad r = 1,2,\cdots \tag{2.82}$$

The mean waiting time in state *i* is given by

$$E(W_i) = \sum_{j=0}^{K} p_{ij} E(W_{ij}) \quad i = 1,2,3,\cdots,K \tag{2.83}$$

Thus the discrete-time semi-Markov process is defined as the two-dimensional stochastic process

$$\{(X_n, T_n) | n = 0,1,2,\cdots K\} \tag{2.84}$$

if the following conditions are satisfied:

- $\{X_n | n = 0,1,2,\cdots K\}$ is a Markov chain

- $P[X_{n+1} = j, T_{n+1} - T_n \le r | X_0, X_1, \cdots, X_n = i; T_0, \cdots T_n]$
  $= P[X_{n+1} = j, W_n \le r | X_n = i], r = 0,1,\cdots,$

Multi-state models

where $W_n = T_{n+1} - T_n$

2.9.2 Continuous-Time Semi-Markov process

In a continuous-time Markov process, we assume that the amount of time spent in a state before a transition to the next state occurs is exponentially distributed (Ibe, 2009). Let the finite-state continuous stochastic process denoted by

$$\{X(t), t \geq 0\} \tag{2.85}$$

Here K is the number of states and let the state space defined as

$$S = \{0,1,2,\cdots,K\} \tag{2.86}$$

Assume that the process just entered state i at time $t=0$, then it chooses the next state j with probability p_{ij}, where

$$\sum_{j=0}^{K} p_{ij} = 1$$
$$p_{ij} \geq 0 \quad i \in S, j \in S \tag{2.87}$$

The time W_{ij} that the process spends in state i until the next transition has the PDF

$$f_{W_{ij}}(t), \quad t \geq 0. \tag{2.88}$$

The W_{ij} is a random variable called the waiting time or holding time for a transition from i to j and it is assumed that

$$E(W_{ij}) < \infty \tag{2.89}$$

Multi-state models

The time W_i that the process spends in state i before making a transition is called the waiting time in state i and its PDF is given by

$$f_{W_i}(t)\sum_{j=0}^{K} p_{ij} f_{W_{ij}}(t) \quad t \geq 0 \tag{2.90}$$

The mean waiting time in state i is

$$E(W_i) = \sum_{j=0}^{K} p_{ij} E\left(W_{ij}\right) \quad i = 1,2,3,\cdots,K \tag{2.91}$$

Thus the continuous -time semi-Markov process is defined as the two-dimensional stochastic process

$$\left\{(X_n, T_n) \mid n = 0,1,2,\cdots K\right\} \tag{2.92}$$

if the following conditions are satisfied:

- $\left\{X_n \mid n = 0,1,2,\cdots K\right\}$ is a Markov chain

- $P\left[X_{n+1} = j, T_{n+1} - T_n \leq t \mid X_0, X_1, \cdots, X_n = iT_0, \cdots T_n\right]$
  $= P\left[X_{n+1} = j, W_n \leq t \mid X_n = i\right], t \geq 0,$

  where $W_n = T_{n+1} - T_n$

2.10 Discrete-Time Markov chains

Let the discrete-time stochastic process be defined by

$$\{X_k, k = 0,1,2,\cdots\} \tag{2.93}$$

Then the above process is called a Markov chain (Ibe, 2009) if for all $i, j, k, \cdots, m$, the following is true

$$
\begin{aligned}
P[X_k &= j \mid X_{k-1} = i, X_{k-2} = n, \cdots, X_0 = m] \\
&= P[X_k = j \mid X_{k-1} = i] \\
&= p_{ijk}
\end{aligned}
\tag{2.94}
$$

The state transition probability is denoted by p_{ijk}. The p_{ijk} means that the conditional probability that the process will be in state j at time k immediately after the next transition, given that it is in state i at time $k - 1$. This is called a nonhomogeneous Markov chain. For homogeneous Markov chains the $p_{ijk} = p_{ij}$ which means that the homogenous Markov chains do not depend on the time unit, which implies that

$$
\begin{aligned}
P[X_k &= j \mid X_{k-1} = i, X_{k-2} = n, \cdots, X_0 = m] \\
&= P[X_k = j \mid X_{k-1} = i] \\
&= p_{ij}
\end{aligned}
\tag{2.95}
$$

The homogenous state transition probability p_{ij} satisfies the following condition:

$$
\begin{aligned}
\sum_j p_{ij} &= 1 \quad i = 1,2,3,\cdots,n \\
0 &\le p_{ij} \le 1
\end{aligned}
\tag{2.96}
$$

Then the Markov chain rule is as follows:

$$
\begin{aligned}
P[X_k &= j, X_{k-1} = i_1, X_{k-2}, \cdots, X_0] \\
&= p_{i_1 j} p_{i_2 i_1} p_{i_3 i_2} \cdots p_{i_k i_{k-1}} P[X_0 = i_k]
\end{aligned}
\tag{2.97}
$$

Multi-state models

Thus, when we know the initial state X_0 and the state transition probabilities, we can evaluate the joint probability

$$P[X_k, X_{k-1}, \cdots, X_0]$$

(2.100)

2.11 Continuous-Time Markov chains

Let the stochastic process be defined by

$$\{X(t) \mid t \geq 0\}$$

(2.101)

Then the stochastic process defined above is a continuous-time Markov chain (Ibe, 2009) if, for all $s, t \geq 0$ and nonnegative integers *i, j, k,*

$$P\left[X(t+s) = j \mid X(s) = i, X(u) = k, \ 0 \leq u \leq s\right]$$
$$= P\left[X(t+s) = j \mid X(s) = i\right]$$

(2.102)

This means that the conditional probability of the future state at time *t + s* given the present state at *s* and all past states depends only on the present state and is independent of the past. The stochastic process defined above is said to be time homogenous or have time homogeneity property if $P\left[X(t+s) = j \mid X(s) = i\right]$ is independent of *s*. Time homogenous Markov chains have homogenous transition probabilities. Let

$$p_{ij}(t) = P[X(t+s) = j \mid X(s) = i]$$
$$p_j(t) = P[X(t) = j]$$

(2.103)

In the above probabilities, $p_j(t)$ is the probability that a Markov chain is in state *j* at time *t* and $p_{ij}(t)$ is the probability that a Markov chain is presently in state *i* will be in state *j* after an additional time *t*. Thus, the $p_{ij}(t)$, are the transition probability functions that satisfy the following conditions

Multi-state models

$$\sum_j p_j(t) = 1$$

$$\sum_j p_{ij}(t) = 1 \tag{2.104}$$

$$0 \le p_{ij} \le 1$$

The Chapman-Kolmogorov equation for continuous-time Markov chain is defined as

$$p_{ij}(t+s) = \sum_k P[X(t) = k \mid X(0) = i] P[X(t+s) = j \mid X(t) = k]$$
$$= \sum_k p_{ik}(t) p_{kj}(s) \tag{2.105}$$

The first equation is due to the Markov property.

If the transition probability matrix is defined then the Chapman-Kolmogorov equation becomes

$$P(t+s) = P(t)P(s) \tag{2.106}$$

The amount of time spends by a continuous-time Markov chain in a state is called holding time and the holding time in state i is exponentially distributed with mean

$$\frac{1}{v_i} \tag{2.107}$$

The v_i, is the rate at which the process leaves state i. After the expiration of holding time the process transit to another state with probability p_{ij}, where

$$\sum_j p_{ij} = 1 \tag{2.108}$$

The $v_i p_{ij}$ represents the rate at which the process makes a transition to state j when in state i. Since the holding times are exponentially distributed, the probability that when the process is in state i a transition to state $j \ne i$ will take place in the next small time Δt is $p_{ij} v_i \Delta t$. The probability that no transition out of state i will take place in Δt given that the process is presently in state i is

$$1 - \sum_{j \ne i} p_{ij} v_i \Delta t \tag{2.109}$$

And the probability that it leaves state i in Δt is

Multi-state models

$$\sum_{i \neq i} p_{ij} v_i \Delta t \tag{2.110}$$

The transition equations for state i for the small time interval are obtain as

$$v_i p_i \sum_{j \neq i} p_{ij} = \sum_{j \neq i} p_j p_{ji} v_j$$

$$1 = \sum_i p_i$$

The left side of the first equation is the rate of transition out of state i while the right side is the rate of transition into state i. This means that in the steady state the two rates are equal for any state in the Markov chain.

2.12 Conclusion

In this chapter we discussed in details the stochastic process which governs the multi-state process through its transition probabilities and transition rates. In stochastic process the system enters a state, spends an amount of time then moves to another state where it spends another time. The time that a system spends in a state is called sojourn or mean time which is also discussed in this chapter. We also looked at the probability transition matrix and transition intensity matrix where both matrices indicate the probability and intensity rates of transition of subject through different states respectively. Markov models which is the set of states, transition rates and probabilities or a model for a stochastic process was also discussed in details.

We discussed the model assumptions such as Markov model assumption which state that the future progress only depends on the current state not on the past states, semi-Markov assumption which state that the future progress not only depends on the current state but also on the entry time into the current state and time homogeneous assumption where the intensities are constant over time. We also looked at time homogenous Markov model particularly for Kolmogorov equations, Eigenvalue decomposition for transition intensity matrix and maximum likelihood estimation method. The Kolmogorov equations are used to derive the relationship between the transition intensity matrix and probability transition matrix, instead of solving the Kolmogorov differential equation to calculate the transition probabilities from transition intensities we use Eigenvalue decomposition and to estimate the unknown parameters in the model we use likelihood estimation method.

We also discussed the Quasi-Newton or scoring procedure which is implemented to obtain the maximum likelihood estimates of the parameters and estimates the asymptotic covariance matrix. We discussed the semi-Markov process which state that the process makes transitions from state to state but the amount of time spent in each state before a transition to the next state occurs at arbitrary random variable that depends on the next state the process will enter. We also discussed in details discrete-time Markov chains and continuous-time Markov chains.

Chapter 3

Multi-state model features

In this chapter we discuss in details the multi-state model features that can have significant influences in the model we fit. The following features will be discussed below

- Covariates
- Model structure

3.1 Covariates

Explanatory variables can be included at each level of the model through generalized regressions in order to incorporate covariates (Christopher H. Jackson, 2005). Once the covariates are incorporated in the model, the interest is not only on the movement of subjects through different states but also on how these covariates influence this movement. Variables associated with transition intensities are assumed to have a multiplicative effect. Each transition intensity can have a separate set of covariate effects. These effects are introduced as covariates in the model via the transition intensities. That is they are included in the model by assuming that the transition intensities are functions of the covariates of interest and are of the form

$$\lambda_{ij}(z) = e^{z^T \beta_{ij}}, \quad i \neq j, \tag{3.1}$$

where z is a vector of covariates and β_{ij} is the vector of regression coefficients corresponding to z. For example let $Q: K \ x \ K$ denote a transition intensity matrix as follows

$$Q = \begin{bmatrix} -\lambda_{12} & \lambda_{12} \\ 0 & 0 \end{bmatrix} \tag{3.2}$$

To incorporate the covariates into the model then the transition intensity matrix (3.2) for the model now become

Multi-state model features

$$Q(z) \;=\; \begin{bmatrix} -\lambda_{12}e^{z^T\beta_{12}} & \lambda_{12}e^{z^T\beta_{12}} \\ 0 & 0 \end{bmatrix} \tag{3.3}$$

The effect of covariate in the transition $i \to j$ is measured by the coefficient β_{ij}. If transition intensity matrix (3.3) is considered for the model and definition (2.48) is used then the transition intensity matrix will depend on z and will be denoted by $P(t;z)$, with entries $p_{ij}(t;z)$, $i,j = 1,2,\ldots,K$. A proportional hazards model can be used to relate the transition intensities to covariates (Marshall and Jones, 1995)

$$\lambda_{ij}(z) \;=\; \lambda_{ij}e^{z^T\beta_{ij}}, \tag{3.4}$$

where z is a vector of covariates, β_{ij} is the vector of regression coefficients corresponding to z and λ_{ij} is the baseline transition rate. If covariates are included in the model, the parameter estimates of the covariate effects $\beta's$ (in 3.4) can be used to calculate the hazard ratios (e^{β}) for each covariate in the model. The hazard ratios show what effect each covariate has on the different transition rates in the model. To illustrate this let the parameter $\lambda_{12} = 0.053$ and parameter $\beta_{12} = -0.3153$ then transition intensity matrix (3.4) become

$$Q(z) \;=\; \begin{bmatrix} -0.053e^{z^T-0.3153} & 0.053e^{z^T-0.3153} \\ 0 & 0 \end{bmatrix} \tag{3.5}$$

where the $z = 1 \; or \; 0$. If $z = 1$, the transition probability matrix for the transition intensity matrix defined in (3.5) become

$$P(t) = \exp\left(\begin{bmatrix} -0.053e^{-0.3153} & 0.053e^{-0.3153} \\ 0 & 0 \end{bmatrix} t \right) \tag{3.6}$$

For $t = 1$ this result to

Multi-state model features

$$P(1) = \begin{bmatrix} 0.9621 & 0.0379 \\ 0 & 0 \end{bmatrix} \qquad (3.7)$$

This shows that transition intensities are affected by covariates. Sometimes covariates are observed at different times to the main response for an example in recurrent disease events. If this time variation is deterministic, for example age, the resultant process is a time inhomogeneous Markov model, even if the baseline intensities are not dependent on time. Then the transition intensities could be written as

$$\lambda_{ij}(z(t)) \;=\; \lambda_{ij} e^{z(t)^T \beta_{ij}}. \qquad (3.8)$$

The quasi-Newton MLE algorithm can be extended to estimate the coefficients of the covariates (Kalbfleisch and Lawless, 1985). A separate canonical decomposition of $Q(z)$ is required for each of the r distinct covariate vectors z in the sample. Let these be denoted by

$$z_h \;=\; (z_{1h}, \ldots, z_{sh}) \qquad (3.9)$$

$$z_{ih} \;=\; 1 \qquad (3.10)$$

$$Q_h \;=\; Q(z_h) \;=\; (\lambda_{ij}(z_h)), \qquad h = 1, \ldots, r, \qquad (3.11)$$

and lastly let $n_{ijl}^{(h)}$ be the number of individuals with covariate values z_h that are in state i at t_{l-1} and state j at t_l.

Then log-likelihood is

$$\log L(\theta) \;=\; \sum_{h=1}^{r} \sum_{l=1}^{m} \sum_{i,j=1}^{k} n_{ijl}^{(h)} \log p_{ij}(w_l; z_h), \qquad (3.12)$$

where

$$P_h(t) \;=\; \exp(Q_n t) \;=\; (p_{ij}(t; z_h)). \qquad (3.13)$$

Multi-state model features

In definition of (2.70) we have the score vector which now involves the sum of *r* terms for each distinct covariate vector,

$$S(\theta) \; = \; \sum_{h=1}^{r} S^{(h)}(\theta), \tag{3.14}$$

The baseline transition intensities and regression parameters of the covariates that need to be estimated are denoted by $\theta \; = \; (\lambda, \beta)$. $S^{(h)}(\theta)$ is a *b x 1* vector, where *b* is the total number of parameters to be estimated in θ ,

$$S_u^{(h)}(\theta) \; = \; \frac{\partial \log L}{\partial \theta_u} \; = \; \sum_{i=1}^{m} \sum_{i,j=1}^{k} n_{ijl}^{(h)} \frac{\partial p_{ij}(w_l; z_h)/\partial \theta_u}{p_{ij}(w_l; z_h)}, \qquad u \; = \; 1,\ldots,b, \tag{2.15}$$

The Fisher scoring matrix $M(\theta)$ in the presence of covariates is given by the following

$$M(\theta) \; = \; \sum_{h=1}^{r} M^{(h)}(\theta), \tag{3.16}$$

and this calculated using definition (2.73) for each *h* and definitions (2.53) and (2.60). Now the derivatives in (2.73) are with respect to each element in θ and a separate diagonalisation is required of each Q_h .

3.2 Model structure

The Markov chain model is defined by a set of states and a set of transitions with associated probabilities. Here the states are denoted by the rectangular box (with relevant state specified) and transitions probabilities denoted by a parameter λ_{ij} where $i, j \; = \; 1,\ldots K$. With *K,* being the number of states in the model. The arrows indicate the path through the different states. Models with single arrows pointing in one direction are called unidirectional models which allow one direction transitions between some transient states. And those models with double arrows pointing in different direction are called bi-directional models which contain absorbing state (death state) but can allow transitions in either direction between some of the transient states. Depending on the specific data under investigation multi-state models are uniquely defined and the types of transitions allowed in a model have implications for inferences about the model. The features of a multi-state model structure that affect the model are considered in details below.

Multi-state model features

- **Basic survival model**

Two states are considered here the first one being Alive (A) and the second being Dead (D). This is a unidirectional model consist of one simple chain states, subjects begin in Alive (A) state can only progress through the Dead (D) state. The aim here is to study the failure that causes a subject to end up in a Dead (D) state for an example if the status is an age of person denoted by (y) then we study the failure time T(y). The model describes the probability of moving from Alive (A) state to Dead (D) state at various points in time. Figure 3.1 illustrates 2-state basic survival model.
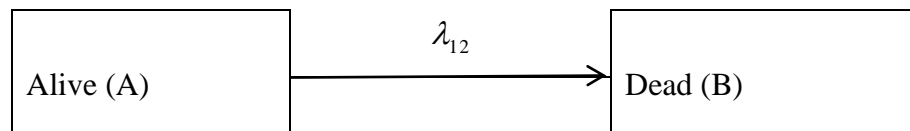
Figure 3.1 Basic survival model

The corresponding transition intensity matrix for the above basic survival model in figure 3.1 has been shown below in (3.17) where λ_{12} indicate the transition from state 1 (Alive state) to state 2 (Dead state). In section (2.3) we indicated that each row elements of a transition intensity matrix must sum zero and each row of a Q matrix in (3.17) sum to zero. The second row of Q matrix in (3.17) contains zero elements since we don't have transition from state 2 (Dead state) to state 1 (Alive state). This indicates that once the subject enters state 2 (Dead state) there's no turning back or recovery.

$$Q \;=\; \begin{bmatrix} -\lambda_{12} & \lambda_{12} \\ 0 & 0 \end{bmatrix} \qquad\qquad (3.17)$$

- **Multiple-decrement survival model**

Here we investigate the time of failure of a status denoted by (y) as in the basic survival model above, that is we study the failure of time T(y). As well as *m* causes that resulted in Withdraw (W) state and Dead (D) state. Three states are considered here the first one

Multi-state model features

being Alive (A), the second being Withdraw (W) and third Dead (D) state. This model is for life insurance where the contract between the subject and insurance can be terminated either by withdraw or die. The model describes the probability of moving from Alive (A) state to Withdrawn (W) state or from Alive (A) state to Dead (D) state at various points in time. Figure 3.2 illustrates 3-state multiple-decrement survival model.
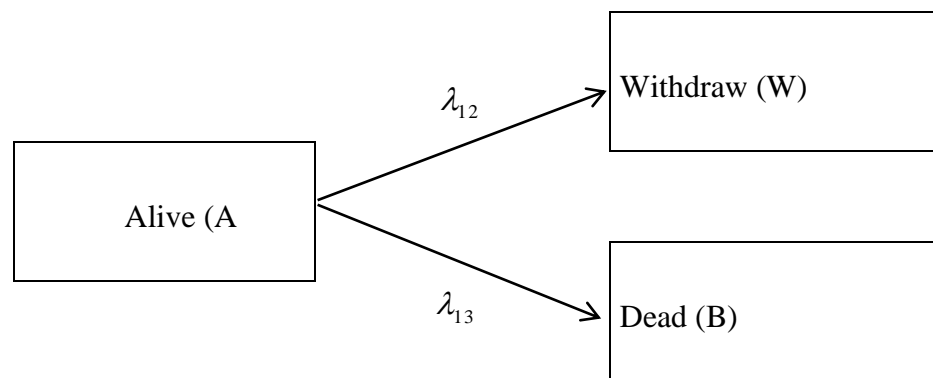
Figure 3.2 Multiple decrement survival model

The corresponding transition intensity matrix for the multiple decrement survival model in figure 3.2 has been shown below in (3.18) where λ_{12} indicate the transition from state 1 (Alive state) to state 2 (Withdraw state) and λ_{13} indicate transition from state 1 (Alive state) to state 3 (Dead state).

$$Q = \begin{bmatrix} -(\lambda_{12} + \lambda_{13}) & \lambda_{12} & \lambda_{13} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (3.18)$$

Multi-state model features

- **Progressive Model**

    In this model all states has only one possible transition into the state. This model is unidirectional where subjects can only move forward to the next state. Once the subject has left a state, it cannot return to previous state from the current state sitting in. An example of a progressive model is the three state chronic disease model where subjects begin in state 1 (healthy) from which they can only progress to state 2 (which is a diseased), from state 2 they may only progress to state 3. Here we have four states such as state one, state two, state three and state four. The last state is normally a death state once subject enters this state the subject can never leave that state. Figure 3.3 illustrates 4-state progressive model.

Figure 3.3 Progressive model

The corresponding Q transition intensity matrix for the progressive model in figure 3.3 has been shown below in (3.19) where λ_{12} indicate the transition from state 1 to state 2, λ_{23} indicate transition from state 2 to state 3 and λ_{34} indicate transition from state 3 to state 4. State 4 is the absorbing state once subject enters this state can never leave this state.

$$
Q \;=\; \begin{bmatrix} -\lambda_{12} & \lambda_{12} & 0 & 0 \\ 0 & -\lambda_{23} & \lambda_{23} & 0 \\ 0 & 0 & -\lambda_{34} & \lambda_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.19}
$$

Multi-state model features

- **Disability Model**

This model has four states such as Active (A), Disabled (d), Withdraw (W) and Dead (D). A person starts in active (healthy) state and then can die or become disable and then die. This model is use to model workers' eligibility for various employee benefits. It is possible to return to the Active (A) state from Disabled (d) state. The model describe probabilities of moving among these various states such as moving back and forth between Active (A) state and Withdraw (W) state several times. Figure 3.4 illustrates 4-state disability model.
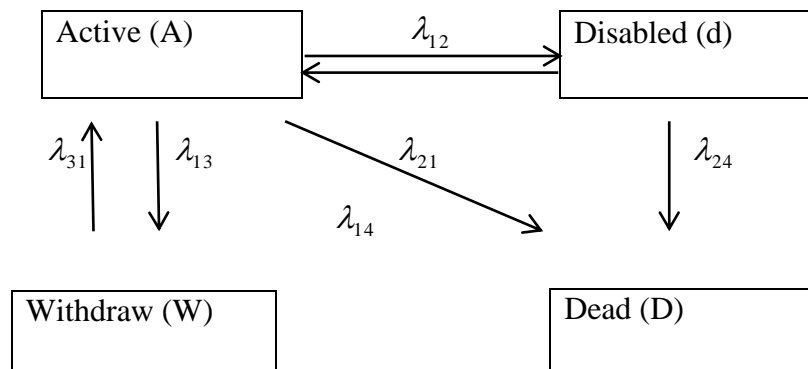
Figure 3.4 Disability model

The corresponding *Q* transition intensity matrix for the disable model in figure 3.4 has been shown below in (3.20) where λ_{12} indicate the transition from state 1 (Active state) to state 2 (Disable state), λ_{21} indicate transition from state 2 (Disable state) back to state 1 (Active state) and so forth. Here subjects can be inactive by moving directly from active state to disable state, withdraw state or die (Dead state). Subject moved to disable state can also move from disable state to dead state and subject moved to withdraw state can be active again by moving from withdraw state back to active state.

Multi-state model features

$$Q \;=\; \begin{bmatrix} -\left(\lambda_{12} + \lambda_{13} + \lambda_{14}\right) & \lambda_{12} & \lambda_{13} & \lambda_{14} \\ \lambda_{21} & -\left(\lambda_{21} + \lambda_{24}\right) & 0 & \lambda_{24} \\ \lambda_{31} & 0 & -\lambda_{31} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.20)$$

- **Recurrent Model**

In recurrent model we have three states such as state one, state two and state three. Fortunately we don't have termination state (Dead (D)). Subjects move more than once between the different states with a probability that the process will eventually return to a state is 1. The example of a recurrent model is the two state illness-recovery model where state 1 represents healthy and state 2 represents illness meaning that the subject can get ill then also can recover from illness going back to state 1 from state 2. Figure 3.5 illustrates 3-state recurring model.
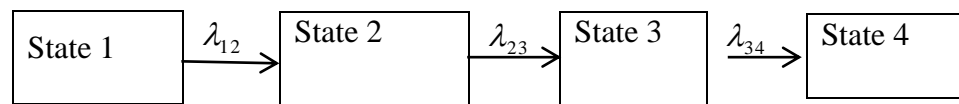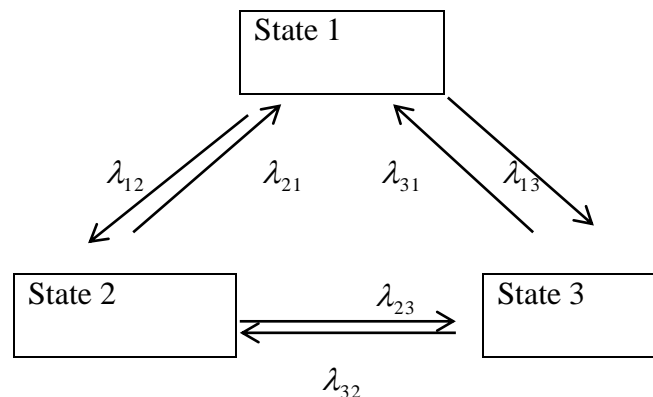

Figure 3.5 Recurring model

The corresponding Q transition intensity matrix for the recurring model in figure 3.5 has been shown below in (3.21). Here the Q transition intensity matrix (3.21) has no absorbing state, subject can move repeatedly between the different states for example subject in state 1 can move to state 2 and subject from state 2 can move back to state 1.

Multi-state model features

$$Q \;=\; \begin{bmatrix} -(\lambda_{12}+\lambda_{13}) & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & -(\lambda_{21}+\lambda_{23}) & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & -(\lambda_{31}+\lambda_{32}) \end{bmatrix} \qquad (3.21)$$

- **Competing Risk Model**

There are four states in this model such as Alive (A) state and the three Dead (D) states characterize by for an example Dead of heart disease state, Dead of cancer state and Dead of other causes. The three states are absorbing while the Alive (A) state is not. All transitions are from the Alive (A) state to the other three states. This model is concerned of different causes of death. Figure 3.6 illustrates 4-state recurring model.
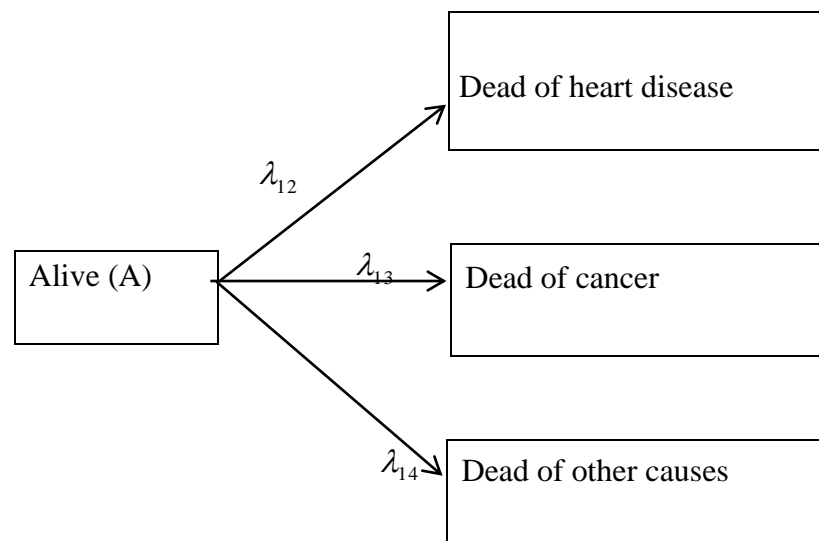
Figure 3.6 Competing model

The corresponding Q transition intensity matrix for the competing model in figure 3.6 has been shown below in (3.22) where λ_{12} indicate the transition from state 1 to state 2, λ_{13} indicate

Multi-state model features

movement from state 1 to state 3 and λ_{14} show transition from state 1 to state 4. Here the transition intensity matrix Q has several absorbing states. Here the causes of death are studied simultaneously for example heart disease, cancer as well as other causes.

$$Q = \begin{bmatrix} -(\lambda_{12} + \lambda_{13} + \lambda_{14}) & \lambda_{12} & \lambda_{13} & \lambda_{14} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (3.22)$$

- **Continuing Care Retirement Communities**

CCRC model has three states that include Healthy (H) state, Independent Living Unit (I) state, Skilled Nursing Facility (S) and Death (D). In this model subjects may move among various states mentioned above under this model. Model describes the probabilities of moving among these states at various points in time. Figure 3.7 illustrates 4-state recurring model.
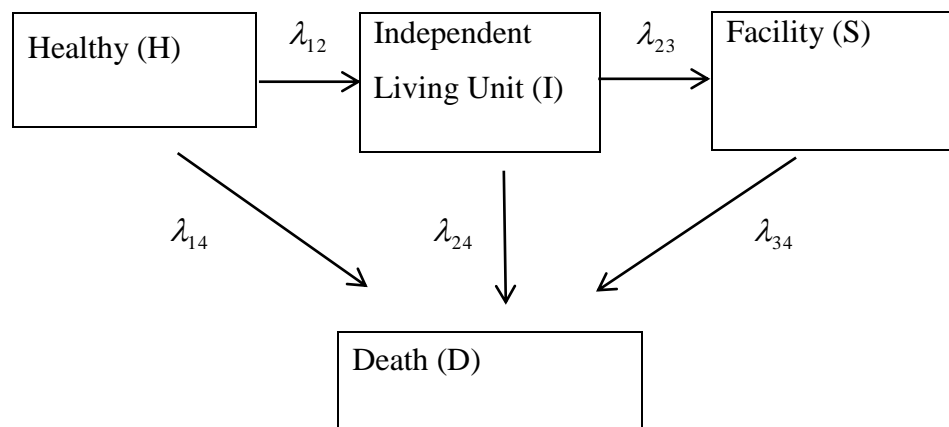
Figure 3.7 CCRC's model

Multi-state model features

The corresponding Q transition intensity matrix for the CCRC's model in figure 3.7 has been shown below in (3.23). Here transitions are unidirectional meaning subjects move from the particular state to the next state no returning to the previous state.

$$Q \;=\; \begin{bmatrix} -\left(\lambda_{12} + \lambda_{14}\right) & \lambda_{12} & 0 & \lambda_{14} \\ 0 & -\left(\lambda_{23} + \lambda_{24}\right) & \lambda_{23} & \lambda_{24} \\ 0 & 0 & -\lambda_{34} & \lambda_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.23}$$

3.3 Conclusion

In this chapter we discussed in details the covariates and model structure. Covariates are the explanatory variables that are incorporated to the model through the transition intensities. The effects of covariates are assumed to be multiplicative. Once the covariates are incorporated in the model, the interest is not only on the movement of subjects through different states but also on how these covariates influence this movement. The effect of covariate is measured by the regression coefficient.

We considered in details the basic survival model, multiple-decrement survival model, progressive model, disability model, recurrent model, competing risk model and continuing care retirement communities model. Usually in basic survival model we have two states and unidirectional transition between the two states. The model describes the probability of transition from state 1 to state 2. In multiple-decrement survival model we have three states and this model is for life insurance where the contract between the subject and insurance can be terminated either by withdraw (state 2) or die (state 3) where state 1 is being alive.

In progressive model the subjects can only move forward to the next state no returning to the previous state. An example of progressive model is the three state chronic disease model where subjects start at state 1 (healthy) from which they can only progress to state 2 (disease) then to state 3 (death). In disability model a person starts in active state (healthy) then can die or become disabling and then die. This model is use to model workers' eligibility for various employee benefits. In recurrent model we don't have death state. The example of a recurrent model is the two state illness-recovery model where state 1 represents healthy and state 2 represents illness meaning that the subject can get ill then also can recover from illness going back to state 1 from state 2.

The competing risk model is concerned of different causes of death. In this model all states are absorbing (death states) except starting state. For example there are four states in this model such as Alive (A) state and the three Dead (D) states characterize by for an example Dead of heart disease state, Dead of cancer state and Dead of other causes. And lastly the continuing care retirement communities' (CCRC) model has three states that include Healthy (H) state,

Multi-state model features

Independent Living Unit (I) state, Skilled Nursing Facility (S) and Death (D). In this model subjects may move among various states.

Chapter 4

Multi-state model assessment

The Markov property and the homogeneity assumptions are strong assumptions that may lead to biased estimates if violated; therefore, it is very important to assess and further investigate a multistate model once it has been fitted to the model. Since the exact times of transition between the states are unknown it makes difficult to assess Markov assumption but the assumption of homogeneity of transition rates through time and across subjects can be assessed. For discrete time and continuous time Markov models these two assumptions can be assessed by modelling transition rates on observed covariates. The fit of the model can be checked by testing the specific assumptions of the model individually and by general goodness-of-fit tests, but continuous covariates, exact death times and irregular sampling times present additional challenges. The Markov models have the following assumptions that need to assessed or validated:

1) Homogeneity of the transition rates through time
2) Homogeneity of the transition rates across the subject population
3) The Markov property or assumption

The above assumptions are investigated in details below together with assessment of covariates effect in the model and the model fit assessment.

4.1 Model assumptions

4.1.1 Homogeneity of the transition rates through time

The key characteristic of time homogenous Markov models is that the transition intensities remain constant through time. This assumption can be tested using piecewise constant transition intensities, this was originally proposed by Faddy (1976) and Kay (1986). For piecewise constant intensities, the number and location of change points must be determined. The likelihood ratio test is used to test the assumption of constant rates through time and again the likelihood ratio test can be used to compare the piecewise constant model with homogenous model. As the alternative Kalbfleisch and Lawless (1985) suggest the fitting of parametric time-dependent model

$$\lambda_{ij} \quad = \quad \lambda_{ij} e^{-\beta t}. \tag{4.1}$$

The likelihood ratio test is performed on the hypothesis that

$$H_0: \quad \beta_0 \quad = \quad 0 \tag{4.2}$$

to assess the homogeneity of the transition rates through time. If the null hypothesis is rejected based on p-value then we will conclude that intensities are constant across time, implying that the assumption of homogeneity of transition intensities across time is valid.

4.1.2   Homogeneity of the transition rates across the subject population

This assumption can be checked by including covariates and treatment indicators

$$x \quad = \quad (x_1, \ldots, x_p)^T \tag{4.3}$$

on individuals in the modelling process. In this model the parameters can be re-parameterized as

$$\lambda_{ij}^x \quad = \quad \lambda_{ij} e^{-\beta_{ij} x} \tag{4.4}$$

Multi-state model assessment

where x is a binary variable with 0 and 1 values that is equivalent of dividing the population into two groups according to its value and $i, j = 1,2,…,K$ denote transitions rates in the model. As in homogeneity of the transition rates through time discussed above, the likelihood ratio test can be used to test

$$H_0 : \quad \beta_{ij} \quad = \quad 0.$$

(4.5)

This can be used to test if the transition rates differ with regard to the two population groups. An overall likelihood ratio test of homogeneity can be obtained by comparing the overall log-likelihood with the sum of the log-likelihoods obtained from the two subpopulation groups. If there is a significant difference between the two population groups then assumption of homogeneity of the transition intensities across population groups has been violated. This will imply that the model is not perfectly fit for the data. Conversely, if no significant difference is found between the two groups in terms of transition intensities, assumption of homogeneity is valid.

4.1.3 Markov property or assumption

This assumption state that the future evolution only depends on the current state at time t. This implies that the history of the process is summarised by the state occupied at time t. The Markov assumption may be assessed by including covariates depending on the history. Markov assumption or property is the key to many analyses even when not appropriate it can provide a base case analysis against which to assess other model. It seems impossible to assess this assumption because it is difficult to test the assumption explicitly for panel observed data in the absence of data on exact transitions. It is necessary therefore to undertake some interpolation before in order to estimate exact transition times. A method suggested by Kay (1986) involves creating data for the exact transition times between states using interpolation. A test can then be performed on this completed dataset to test the Markov assumption. For instance, consider a disease model where death is the absorbing state and which includes state 1 and state 2 and transitions between them is *bi*-directional (*bi*-directional means transition is in both directions). Let x denote the time spent in state 2 during last sojourn (sojourn has been discussed in section (2.5)) from state 1.

Multi-state model assessment

To fit the model where intensity or rate is denoted by λ_{12} is given by

$$\lambda_{12} = \lambda_0 e^{(\beta\, x)} \tag{4.6}$$

Then we set the hypothesis that

$$H_0 : \quad \beta = 0 \tag{4.7}$$

which would assess the Markov assumption that the transition intensity to death from state 1 is unaffected by the previous sojourn time. This method is applicable to test other Markov assumption but accuracy of any conclusions depends on the accuracy with which the exact transition times can be determined through interpolation.

4.2 Covariates effect in the model assessment

Explanatory variables can be included at each level of the model through generalized regressions in order to incorporate covariates (Christopher H. Jackson, 2005). Once the covariates are incorporated in the model, the interest is not only on the movement of subjects through different states but also on how these covariates influence this movement. Variables associated with transition intensities are assumed to have a multiplicative effect. Each transition intensity can have a separate set of covariate effects. These effects are introduced as covariates in the model via the transition intensities. That is they are included in the model by assuming that the transition intensities are functions of the covariates of interest and are of the form

$$\lambda_{ij}(z) = e^{z^T \beta_{ij}}, \quad i \neq j, \tag{4.8}$$

where z is a vector of covariates and β_{ij} is the vector of regression coefficients corresponding to z. For example let $Q : K \ x \ K$ denote a transition intensity matrix as follows

$$Q = \begin{bmatrix} -\lambda_{12} & \lambda_{12} \\ 0 & 0 \end{bmatrix} \tag{4.9}$$

To incorporate the covariates into the model then the transition intensity matrix for the model now (4.9) become

Multi-state model assessment

$$Q(z) \quad = \quad \begin{bmatrix} -\lambda_{12}e^{z^T\beta_{12}} & \lambda_{12}e^{z^T\beta_{12}} \\ 0 & 0 \end{bmatrix} \tag{4.10}$$

So it is very important to assess the significance of these covariates in the model which is done by using the likelihood ratio and Wald test.

4.3 Model fit assessment

It is very important to assess model suitability once the model is fitted and the constraints has been complied, for instance the validation of underlying assumptions. To assess the fit of the multi-state model Person goodness-of-fit and informal model diagnostic tool can be used which are further investigated below.

4.3.1 Informal diagnostic tool

4.3.1.1 Prevalence counts

Prevalence counts provide a more informal empirical measure of state occupancy which involves comparing the observed state occupancies at a fixed set of times with those expected under the fitted model. This method is applicable for exact death times and tries to eliminate problems of irregular observation times. Here of a table of observed and expected state occupancies at a sequence of times is constructed. In prevalence counts the intermediate states will be underestimated when the observed transitions imply the passing through of a series of states. The expected counts are calculated by summing the probability a subject is in the specified state given their initial state over all subjects who are under observation at the time of interest. A subject is under observation until their absorbing state. An indication of where the data deviate from the model is achieved by comparing the observed count O_{uv} with the expected count E_{uv} for particular state u and time t_i using

$$M_{uv} \quad = \quad \frac{(O_{uv} - E_{uv})^2}{E_{uv}}, \tag{4.11}$$

where

$$E_{uv} \quad = \quad \sum_l Pg_{l,u}(t_i, z_l), \tag{4.12}$$

where g_u is the initial state (assumed known) and z_l the covariate vector, subject l. A large value of M_{uv} would indicate a poor fit (Gentleman *at al.* (1994)). Formal tests to determine whether the deviates observed are statistically significant are not possible because due the *ad hoc* interpolation of observed states and also the dependence between the rows of the tables. This shows that the prevalence counts can only be used as an informal measure of fit.

4.3.2 Pearson goodness-of-fit test

The Pearson-type- χ^2 test is a procedure for testing parametric continuous-time panel observed multi-state Markov models. For balanced observations and when there are no continuous covariates, a test with an asymptotic χ^2 null distribution can be found. Contingency table methods provide an assessment of overall fit of the assumed model. The model fit can be assessed by considering observed and expected transition frequencies either through a likelihood ratio test or the asymptotically equivalent Pearson λ^2 statistic. Assessing goodness of fit by prevalence counts involves estimating the observed prevalence at a series of points by some form of interpolation. This is applicable if observation times are close together. An alternative method of assessing goodness of fit is to construct tables of observed and expected numbers of transitions which leads to a formal test of goodness-of-fit which is similar to the Pearson λ^2 test for contingency tables. This was proposed by Aguirre-Hernández, R. & Farewell (2002). The observed and expected numbers of transitions in each group are defined by

$$O_{hl_h rscg} \quad = \quad \sum I(S(t_{i,j+1}) \quad = \quad s, S(t_{ij}) \quad = \quad r) \tag{4.13}$$

$$E_{hl_h rscg} \quad = \quad \sum P(S(t_{i,j+1}) \quad = \quad s \mid S(t_{ij}) \quad = \quad r) \tag{4.14}$$

where $I(A)$ is the indicator function for an event A and summation is over the set of transitions in the category defined by $h, l_h, c, g,$ over all individuals i. The Pearson-type test statistic is then given by

$$T \quad = \quad \sum_{hl_n rscg} \frac{(O_{hl_h rscg} - E_{hl_h rscg})^2}{E_{hl_h rscg}} \tag{4.15}$$

Multi-state model assessment

The Pearson test statistic has a λ^2_{n-p} distribution, where n-p is the number of independent cells in the table (n) minus the number of estimated parameters p. The observed transitions are realizations from a set independent but non-identical multinomial distribution, since the distribution of T is not exactly λ^2. Aguirre-Hernández, R. & Farewell (2002) showed that λ^2_{n-p} is a good approximation if there are no covariates in the model. For models with covariates, the null mean of T is higher than n-p, but lower than n, which implies that the lower and upper bounds for p-value of the statistic can be obtained from the λ^2_{n-p} and λ^2_n distributions. To obtain accurate p-value, bootstrap procedure is required as described by Aguirre-Hernández, R. & Farewell (2002). Pearson goodness of fit requires an arbitrary grouping of the observations and it is implemented within R package ***msm*** including the modified test for exact death times.

4.4 Conclusion

The focus in this chapter is to assess the multi-state model through model assumptions, covariates effect in the model and the fit of the model. Specifically we looked at the assumptions such as homogeneity of the transition rates through time, homogeneity of the transition rates across the subject population and the Markov property or assumption. The first assumption indicate that the intensity rates are constant through time, the second assumption also indicate that the intensity rates are homogeneous across subject population and the last one state that the future progress only depends on the current state not on the past states. The likelihood ratio test can be used to test the first assumption, the second assumption can also be tested using likelihood ratio test but including covariates and the last assumption can be tested by including the covariates depending on the history.

Covariates are the explanatory variables that are incorporated to the model through the transition intensities. The effects of covariates are assumed to be multiplicative. Once the covariates are incorporated in the model, the interest is not only on the movement of subjects through different states but also on how these covariates influence this movement. The effect of covariate is measured by the regression coefficient. To assess the significance of these covariates we employ likelihood ratio and Wald test.

To assess the fit of the multi-state model we use Pearson goodness-of-fit and informal model diagnostic tool. The model fit can be assessed by considering observed and expected transition frequencies either through a likelihood ratio test or the asymptotically equivalent Pearson chi-square statistics. Informal diagnostic tool involves prevalence counts where the observed state occupancies are compared with the expected occupancies under the fitted model.

Chapter 5

Data simulation and application

The main purpose of this study was to assess the fit of model particular to assess or validate the Markov assumptions. In order to be able to assess those assumptions we firstly need a dataset that can be used to fit the model. Therefore with regard to this we will need to simulate a panel or longitudinal dataset that is suitable for Markov models. Longitudinal dataset consists of repeated measurements of the state of a subject and the time between observations. The period of time with observations in longitudinal dataset is being made on subject at regular or irregular time intervals until the subject dies then the study ends. So in this chapter we discuss the data simulation and application. There are three Markov models assumptions that need to be validated or assessed and they are:

1) Homogeneity of the transition rates through time
2) Homogeneity of the transition rates across the subject population
3) The Markov property or assumption

In this thesis I wrote an R code that simulate panel dataset where specifically these assumptions are violated. We will implement the first two assumptions that is Homogeneity of the transition rates through time and Homogeneity of the transition rates across the subject population by passing some parameters in R function written. The Markov property or assumption is already accommodated in the function whether you pass the parameters or not. For more information with regard to this assumption please refer to section (4.1) of this thesis.

5.1 Models considered for simulation and application

For the purpose of the simulation of data set the following transition intensity matrices corresponding to each multi-state model will be considered in details in this thesis and the corresponding transition intensity matrices will be used in data simulation and application.

5.1.1 Four-state model

This model is unidirectional where subjects can only move forward to the next state. Once the subject has left a state, it cannot return to previous state from the current state

Data simulation and application

sitting in. Let $\lambda_{12}, \lambda_{23}$ and λ_{34} denote transitions intensity between the states, then multi-state model is

$$
\boxed{\text{State 1}} \xrightarrow{\lambda_{12}} \boxed{\text{State 2}} \xrightarrow{\lambda_{23}} \boxed{\text{State 3}} \xrightarrow{\lambda_{34}} \boxed{\text{State 4}}
$$

Figure 5.1 Four-state model

The corresponding transition intensity matrix for the above model in figure 5.1 is

$$
Q = \begin{bmatrix}
-\lambda_{12} & \lambda_{12} & 0 & 0 \\
0 & -\lambda_{23} & \lambda_{23} & 0 \\
0 & 0 & -\lambda_{34} & \lambda_{34} \\
0 & 0 & 0 & 1
\end{bmatrix} \tag{5.1}
$$

To illustrate the transition intensity matrix specified in (5.1) let $\lambda_{12} = 0.4$, $\lambda_{23} = 0.8$

and $\lambda_{34} = 0.6$ then (5.1) become

$$
Q = \begin{bmatrix}
-0.4 & 0.4 & 0 & 0 \\
0 & -0.8 & 0.8 & 0 \\
0 & 0. & -0.6 & 0.6 \\
0 & 0 & 0 & 1
\end{bmatrix} \tag{5.2}
$$

This is a four-state model where transitions are allowed between the 2 states from state 1 to state 2, from state 2 to state 3 and from state 3 to state 4. The mean time spent in state 1 is

$$
-1/\lambda_{11} = -1/-0.4 = 2.5 \tag{5.3}
$$

the mean time spent in state 2 is

Data simulation and application

$$-1/\lambda_{22} = -1/-0.8 = 1.25 \qquad (5.4)$$

and the mean time spent in state 3 is

$$-1/\lambda_{33} = -1/-0.6 = 1.66. \qquad (5.5)$$

The probability that the subject's next move from state i to state j is

$$-\lambda_{ij}/\lambda_{ii}. \qquad (5.6)$$

So the probability of a subject move from state 1 to state 2 is

$$-\lambda_{12}/\lambda_{11} = -0.4/-0.4 = 1 \qquad (5.7)$$

the probability of a subject moving from state 2 to state 3 is

$$-\lambda_{23}/\lambda_{22} = -0.8/-0.8 = 1. \qquad (5.8)$$

and the probability of a subject moving from state 3 to state 4 is

$$-\lambda_{34}/\lambda_{33} = -0.6/-0.6 = 1 \qquad (5.9)$$

This implies that the probability of moving from a lower state to higher state is always possible.

5.1.2 Three-state model

Let λ_{12}, λ_{21}, λ_{23} and λ_{32} denote transitions intensity between the states, then multi-state model is
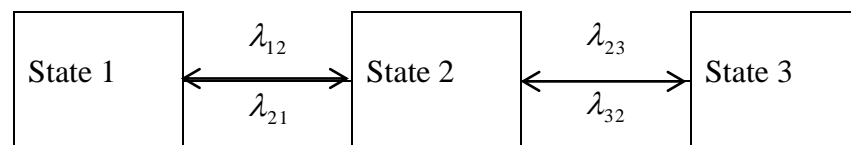
Figure 5.2 Three-state model

The corresponding transition intensity matrix for the above model in figure 5.2 is

Data simulation and application

$$Q = \begin{bmatrix} -\lambda_{12} & \lambda_{12} & 0 \\ \lambda_{21} & -(\lambda_{21} + \lambda_{23}) & \lambda_{23} \\ 0 & \lambda_{32} & -\lambda_{32} \end{bmatrix} \tag{5.10}$$

To illustrate the transition intensity matrix specified in (5.10) let $\lambda_{12} = 0.4, \ \lambda_{21} = 0.4, \ \lambda_{23} = 0.4$ and $\lambda_{21} = 0.4$ then (5.10) become

$$Q = \begin{bmatrix} -0.4 & 0.4 & 0 \\ 0.4 & -0.8 & 0.4 \\ 0 & 0.4 & -0.4 \end{bmatrix} \tag{5.11}$$

The probability intensity matrix (5.11) denote a three-state model where transitions are allowed from state 1 to state 2, from state 2 to state 1, from state 2 to state 3 and from state 3 to state 1. The mean time spent in state 1 and state 3 is

$$-1/\lambda_{ii} = -1/-0.4 = 2.5 \tag{5.12}$$

and the mean time spent in state 2 is

$$-1/\lambda_{ii} = -1/-0.8 = 1.25. \tag{5.13}$$

The probability of the subject that it will move from state 1 to state 2 is

$$-\lambda_{ij}/\lambda_{ii} = -0.4/-0.4 = 1 \tag{5.14}$$

and the probability of the subject that it will move from state 3 to state 2 is

$$-\lambda_{ij}/\lambda_{ii} = -0.4/-0.4 = 1. \tag{5.15}$$

The probability of transitions from state 2 to state 1 or from 2 to state 3 is

$$-\lambda_{ij}/\lambda_{ii} = -0.4/-0.8 = 0.5. \tag{5.16}$$

This means that the probability of moving from a lower state to higher state is the same with the probability of moving from higher state to a lower state.

Data simulation and application

5.1.3 Four-state model

Let λ_{12}, λ_{21}, λ_{23}, λ_{32}, λ_{34} and λ_{43} denote transitions intensity between the states, then multi-state model is
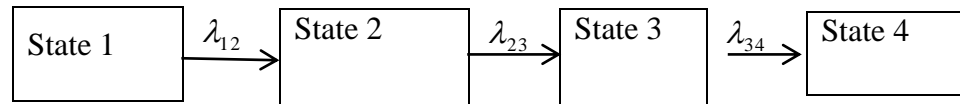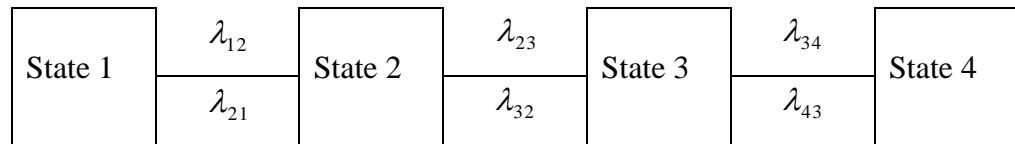

Figure 5.3 Four-state model

The corresponding transition intensity matrix for the above model in figure 5.3 is

$$Q = \begin{bmatrix} -\lambda_{12} & \lambda_{12} & 0 & 0 \\ \lambda_{21} & -(\lambda_{21}+\lambda_{23}) & \lambda_{23} & 0 \\ 0 & \lambda_{32} & -(\lambda_{32}+\lambda_{34}) & \lambda_{34} \\ 0 & 0 & \lambda_{43} & -\lambda_{43} \end{bmatrix} \tag{5.17}$$

To illustrate the transition intensity matrix specified in (5.17) let $\lambda_{12}=\lambda_{21}=\lambda_{23}=\lambda_{32}=\lambda_{34}=\lambda_{43}=0.4$ then (5.17) become

$$Q = \begin{bmatrix} -0.4 & 0.4 & 0 & 0 \\ 0.4 & -0.8 & 0.4 & 0 \\ 0 & 0.4 & -0.8 & 0.4 \\ 0 & 0 & 0.4 & -0.4 \end{bmatrix} \tag{5.18}$$

The probability intensity matrix (5.18) denote a four-state model where transitions are allowed from state 1 to state 2, from state 2 to state 1, from state 2 to state 3, from state 3 to state 2, from state 3 to state 4 and from state 4 to state 3. The mean time spent in state 1 and state 4 is

$$-1/\lambda_{ii} = -1/-0.4 = 2.5 \tag{5.19}$$

Data simulation and application

and the mean time spent in state 2 and state 3 is

$$-1/\lambda_{ii} = -1/-0.8 = 1.25.$$ (5.20)

The probability of the subject that it will move from state 1 to state 2 is

$$-\lambda_{ij}/\lambda_{ii} = -0.4/-0.4 = 1$$ (5.21)

and the probability of the subject that it will move from state 4 to state 3 is

$$-\lambda_{ij}/\lambda_{ii} = -0.4/-0.4 = 1.$$ (5.22)

The probability of transitions from state 2 to state 1 or from 2 to state 3 is

$$-\lambda_{ij}/\lambda_{ii} = -0.4/-0.8 = 0.5.$$ (5.23)

The probability of transitions from state 3 to state 2 or from 3 to state 4 is

$$-\lambda_{ij}/\lambda_{ii} = -0.4/-0.8 = 0.5.$$ (5.24)

This means that the probability of moving from a lower state to higher state is the same with the probability of moving from higher state to a lower state.

5.2 Data simulation

The R code for data simulation and application that was written is provided in the appendix; the code used the msm package developed by Jackson (2005). The R code simulated two datasets, one in which the above assumptions were specifically violated and one that used the simmulti.msm function in the msm package. The dataset using the simmulti.msm function was assumed to simulate data without violating the abovementioned assumptions. The dataset simulated by using the *simmulti.msm* function in the *msm package* had two parameters in this regard. The parameters were data and *qmatrix*. Data represented the data frame with an optional column named 'subject', which corresponded to subject identification numbers, and a mandatory column named 'time', representing observation time. The observation times were sorted according to individuals. The parameter *qmatrix* was the transition intensity matrix of the Markov process without covariate effect. For more information about the transition intensity matrix, please consult Section 2.3. For dataset simulation, the researcher used the following transition intensity matrices:

$$Q_1 = \begin{bmatrix} -0.92 & 0.51 & 0.41 \\ 0.49 & -0.98 & 0.49 \\ 0.32 & 0.93 & -0.61 \end{bmatrix} \tag{5.25}$$

$$Q_2 = \begin{bmatrix} -0.50 & 0.25 & 0.25 \\ 0.65 & -0.90 & 0.25 \\ 0.21 & 0.75 & -0.96 \end{bmatrix} \tag{5.26}$$

Data simulation and application

$$Q_3 \;=\; \begin{bmatrix} -1.00 & 0.75 & 0.25 \\ 0.25 & -1.00 & 0.75 \\ 0.45 & 0.25 & -0.70 \end{bmatrix} \qquad (5.27)$$

For each Q we simulate two datasets, one using our defined R code where assumptions are violated and the other using **msn package** where assumptions are assumed to be not violated. To assess the importance of assumptions we will compare the transition intensity matrices, probability transition matrices, sojourn time and other statistic estimates between these two datasets. If for example transition rates do not differ very much then we can conclude that assumptions are not important otherwise assumptions are very important. Once the simulated datasets has been simulated using above three different Q's then following transition intensity matrices are fitted in each case

$$Q_{11} \;=\; \begin{bmatrix} -1.00 & 0.70 & 0.30 \\ 0.60 & -0.90 & 0.30 \\ 0.52 & 0.33 & -0.85 \end{bmatrix} \qquad (5.28)$$

$$Q_{22} \;=\; \begin{bmatrix} -1.00 & 0.85 & 0.15 \\ 0.45 & -0.80 & 0.35 \\ 0.71 & 0.25 & -0.96 \end{bmatrix} \qquad (5.29)$$

The above mentioned two transition matrices (Q_{11} and Q_{22}) were arbitrary chosen or defined for illustration purposes. To assess the assumption of homogeneity of the transition rates across the subject population Q_1, sample size of 100 ($n = 100$), missed observations of (10 %) and

Data simulation and application

time period of $(0-12)$ were used to simulate the first dataset, then Q_2, sample size of 200 ($n = 200$), missed observations of (10 %) and time period of $(0-12)$ were used to simulate the second dataset and Q_3, sample size of 300 ($n = 300$), missed observations of (10 %) and time period of $(0-12)$ were used to simulated the third dataset. The sample size refers to the number of patients or subject investigated, missed observations refers to the fact that when a patient do a check-up on a monthly basis for example there will be some cases where he or she missed one of any month check-up and time period refers to the fact that subject will be observed monthly for one year. So under each of these cases we simulate two datasets, one using our defined R code where assumptions are violated and the other using msn package where assumptions are assumed to be not violated. Then to assess the importance of the assumption we fit Q_{11} and Q_{22} in each case and compare transition rates across different sample size (n) and also across different dataset.

To assess the assumption of homogeneity of the transition rates through time we used the same describe above method but now the sample size (n) is held constant at $n = 100$ and the time period vary in each case, that is for Q_1 we use time period of $(0-12)$, for Q_2 we use time period of $(0-24)$ and for Q_3 we use $(0-36)$.

The Markov assumption state that the transition rates only depend on the history of the process through the current state. To assess this assumption the R code was written in such a way that it relax this assumption and transition rates were compared. The results has been shown from table 5.3 to table 5.6 and graphical representation has also shown in figure 5.4, 5.5, 5.6 and 5.7. The extracted dataset from the whole data is shown in table 5.1 below for illustration.

Data simulation and application

Table 5.1: Illustrate the simulated longitudinal dataset

Subject	Time (years)	State	Status/State
1	0	1	Healthy
1	1	3	Severe
2	0	1	Healthy
2	1	3	Severe
3	0	1	Healthy
3	1	3	Severe
3	2	2	Mild
3	3	1	Healthy
3	4	2	Mild
3	5	3	Severe
3	6	3	Severe
3	7	1	Healthy
4	0	1	Healthy
4	1	2	Mild
4	2	1	Healthy
4	3	3	Severe
5	0	1	Healthy
5	1	3	Severe
5	2	1	Healthy
5	3	2	Mild
5	4	3	Severe
5	5	1	Healthy
5	6	2	Mild
6	0	1	Healthy
6	1	2	Mild
6	2	3	Severe
8	0	1	Healthy
8	1	3	Severe
8	2	1	Healthy
9	0	1	Healthy
9	1	2	Mild
9	2	3	Severe
9	3	1	Healthy

Data simulation and application

The above Table 5.1 can be also summarised in the following format:

Table 5.2: Illustrate subject and observation times

Subject	t_0	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}
1	1	3											
2	1	3											
3	1	3	2	1	2	3	3	1					
4	1	2	1	3									
5	1	3	1	2	3	1	2						
6	1	2	3										
8	1	3	1										
9	1	2	3	1									

The model is based on progression of a disease from individual subject. The model to be fitted has three state: State 1 corresponds to those subject that are free from disease and is denoted by 1 or label Healthy, State 2 corresponds to those subject that have a disease but the disease is a minor and it is denoted by 2 or label Mild and State 3 corresponds to those subject that have a disease that is severe and it is denoted by 3 or label Severe. It is clear from the transition intensity matrices defined in (5.25, 5.26, 5.27, 5.28 and 5.29) and from table 5.1 that a subject or patient whose healthy (not ill) can get ill by moving from state 1 (Healthy) to state 2 (Mild) or to state 3 (Severe) this mean that the person can get very sick on the onset eg. person ate poison. The person can recover from state 3 (Severe) to state 1 (Healthy) or to state 2 (Mild). The patient can also move from state 2 (Mild) to state 3 (Severe). This transitions refers to recurrent model where subject can move to any of these states and move backward. This chapter is based on recurrent model and other models that are specified in this chapter their estimates has been monitored and analysed but their results are not shown here. The aim of Markov models is to estimate transition probabilities and transition intensities rates of the subjects between the two states and in this thesis we used R package called ***msm*** to estimate those probabilities.

5.3 Data application

The models described in Section 5.1 were analysed by using the corresponding Q matrices defined in that section, and their results were monitored for comparison; unfortunately, it is impossible to show those results in this section as they are too numerous. Instead, the concern tried with the transition intensity matrices in definitions 5.25, 5.26, 5.27, 5.28 and 5.29. Here I try to assess the model fit using transition intensity matrices (5.28 and 5.29), meaning that I fitted and assessed the model using (5.28 and 5.29). Our aim here is to validate above mentioned assumptions defined in this chapter (5). Firstly the model was fitted using the *msm* function defined in *msm R package* and transition intensity matrices defined in (5.28 and 5.29) using the simulated data depicted above (Table 5.1). The model was fitted without the effect of any covariate. As mentioned in section 5.2 of this chapter two transition intensity matrices were used to fit the model in each case and the results were recorded. The aim here was to compare the transition rates for different scenarios as describe in the above section. For easy comparison the results were summarised in the following table (Table 5.3 to 5.6).

Data simulation and application

Table 5.3: Illustration of the results

		Data simulation with Q_1		Data simulation with Q_2		Data simulation with Q_3	
		Missed observation (10 %) Time ($0-12$)		Missed observation (10 %) Time ($0-12$)		Missed observation (10 %) Time ($0-12$)	
		$n = 100$		$n = 200$		$n = 300$	
Q	λ	Assumptions violated	Assumptions not violated	Assumptions violated	Assumptions not violated	Assumptions violated	Assumptions not violated
		Transition matrix estimates for Q_{11}					
Q_{11}	λ_{11}	-1.288	-1.040	-0.478	-0.501	-0.984	-0.984
	λ_{12}	0.684	0.455	0.267	0.233	0.655	0.655
	λ_{13}	0.604	0.586	0.211	0.268	0.329	0.329
	λ_{21}	0.637	0.616	0.636	0.680	0.237	0.237
	λ_{22}	-1.205	-1.082	-0.936	-1.014	-0.937	-0.937
	λ_{23}	0.568	0.466	0.300	0.333	0.700	0.700
	λ_{31}	0.216	0.330	0.227	0.315	0.428	0.428
	λ_{32}	0.757	1.073	0.740	0.689	0.222	0.222
	λ_{33}	-0.973	-1.404	-0.967	-1.004	-0.650	-0.650
		Transition matrix estimates for Q_{22}					
Q_{22}	λ_{11}	-1.268	-1.052	-0.477	-0.502	-0.984	-0.984
	λ_{12}	0.648	0.456	0.266	0.231	0.655	0.655
	λ_{13}	0.620	0.596	0.211	0.271	0.329	0.329
	λ_{21}	0.619	0.589	0.635	0.696	0.236	0.236
	λ_{22}	-1.164	-1.086	-0.936	-1.037	-0.937	-0.937
	λ_{23}	0.545	0.497	0.301	0.341	0.701	0.701
	λ_{31}	0.214	0.367	0.227	0.298	0.428	0.428
	λ_{32}	0.752	1.074	0.740	0.731	0.222	0.222
	λ_{33}	-0.966	-1.441	-0.967	-1.030	-0.650	-0.650
		Sojourn time estimates for Q_{11}					
S_{11}	λ_{11}	0.776	0.961	2.094	1.998	1.017	1.017
	λ_{22}	0.830	0.924	1.068	0.986	1.067	1.067
	λ_{33}	1.028	0.712	1.034	0.996	1.539	1.539
		Sojourn time estimates for Q_{22}					
S_{22}	λ_{11}	0.788	0.950	2.095	1.994	1.016	1.016
	λ_{22}	0.859	0.921	1.069	0.964	1.067	1.067
	λ_{33}	1.035	0.694	1.034	0.971	1.538	1.538

Fitting Q_{11} and Q_{22} transition matrices (three-state model) in different observations

Data simulation and application

Table 5.4: Illustration of the results – continued

		Data simulation with Q_1		Data simulation with Q_2		Data simulation with Q_3	
		Missed observation (10 %) Time ($0-12$)		Missed observation (10 %) Time ($0-12$)		Missed observation (10 %) Time ($0-12$)	
		$n = 100$		$n = 200$		$n = 300$	
Q	λ	Assumptions violated	Assumptions not violated	Assumptions violated	Assumptions not violated	Assumptions violated	Assumptions not violated
		Probability matrix estimates for Q_{11}					
P_{11}	P_{11}	0.249	0.323	0.501	0.509	0.264	0.264
	P_{12}	0.376	0.404	0.299	0.267	0.292	0.292
	P_{13}	0.374	0.274	0.201	0.224	0.444	0.444
	P_{21}	0.249	0.321	0.495	0.506	0.263	0.263
	P_{22}	0.377	0.404	0.303	0.270	0.290	0.290
	P_{23}	0.374	0.274	0.202	0.225	0.447	0.447
	P_{31}	0.249	0.323	0.489	0.502	0.264	0.264
	P_{32}	0.377	0.404	0.306	0.272	0.289	0.289
	P_{33}	0.375	0.274	0.205	0.226	0.447	0.447
		Probability matrix estimates for Q_{22}					
P_{22}	P_{11}	0.248	0.322	0.501	0.510	0.264	0.264
	P_{12}	0.379	0.405	0.298	0.268	0.292	0.292
	P_{13}	0.373	0.273	0.201	0.222	0.445	0.445
	P_{21}	0.248	0.322	0.495	0.506	0.263	0.263
	P_{22}	0.379	0.405	0.302	0.271	0.290	0.290
	P_{23}	0.373	0.273	0.202	0.223	0.447	0.447
	P_{31}	0.248	0.322	0.489	0.503	0.264	0.264
	P_{32}	0.379	0.405	0.306	0.273	0.289	0.289
	P_{33}	0.373	0.273	0.205	0.224	0.447	0.447
		-2 log likelihood estimates for Q_{11} and Q_{22}					
Q_{11}		2 394.661	2 407.596	4 329.189	4 312.165	6 846.045	7 019.708
Q_{22}		2 394.62	2 407.786	4 329.187	4 311.91	6 846.046	7 020.155
		Pearson statistic estimates for Q_{11} and Q_{22}					
Q_{11}		16.96 (p = 0.423)	15.81 (p = 0.488)	12.32 (p = 0.778)	17.19 (p = 0.452)	21.12 (p = 0.232)	18.33 (p = 0.37)
Q_{22}		16.97 (p = 0.424)	16.06 (p = 0.471)	12.32 (p = 0.778)	16.92 (p = 0.468)	21.12 (p = 0.232)	18.81 (p = 0.341)

Fitting Q_{11} and Q_{22} transition matrices (three-state model) in different observations – continued

Data simulation and application

Table 5.5: Illustration of the results

Fitting Q_{11} and Q_{22} transition matrices (three-state model) in different time periods							
		Data simulation with Q_1		Data simulation with Q_2		Data simulation with Q_3	
		Missed observation (10 %) Time ($0-12$)		Missed observation (10 %) Time ($0-24$)		Missed observation (10 %) Time ($0-36$)	
		$n=100$		$n=100$		$n=100$	
Q	λ	Assumptions violated	Assumptions not violated	Assumptions violated	Assumptions not violated	Assumptions violated	Assumptions not violated
Transition matrix estimates for Q_{11}							
Q_{11}	λ_{11}	-1.111	-0.956	-0.518	-0.530	-1.101	-0.886
	λ_{12}	0.501	0.323	0.278	0.284	0.838	0.625
	λ_{13}	0.610	0.633	0.241	0.247	0.262	0.261
	λ_{21}	0.504	0.341	0.714	0.689	0.260	0.226
	λ_{22}	-1.049	-0.836	-0.943	-0.954	-1.139	-0.982
	λ_{23}	0.545	0.494	0.229	0.265	0.880	0.756
	λ_{31}	0.267	0.469	0.275	0.302	0.494	0.430
	λ_{32}	0.872	1.034	0.668	0.723	0.294	0.301
	λ_{33}	-1.139	-1.503	-0.943	-1.025	-0.789	-0.731
Transition matrix estimates for Q_{22}							
Q_{22}	λ_{11}	-1.111	-0.898	-0.519	-0.527	-1.108	-0.901
	λ_{12}	0.501	0.368	0.268	0.283	0.851	0.646
	λ_{13}	0.610	0.531	0.251	0.245	0.257	0.254
	λ_{21}	0.505	0.330	0.724	0.687	0.254	0.228
	λ_{22}	-1.050	-0.924	-0.933	-0.954	-1.130	-0.977
	λ_{23}	0.545	0.594	0.209	0.267	0.876	0.749
	λ_{31}	0.265	0.386	0.247	0.298	0.499	0.432
	λ_{32}	0.873	1.109	0.654	0.726	0.286	0.291
	λ_{33}	-1.138	-1.494	-0.901	-1.024	-0.784	-0.723
Sojourn time estimates for Q_{11}							
S_{11}	λ_{11}	0.900	1.046	1.929	1.886	0.908	1.129
	λ_{22}	0.953	1.197	1.060	1.049	0.878	1.018
	λ_{33}	0.878	0.666	1.060	0.975	1.268	1.368
Sojourn time estimates for Q_{22}							
S_{22}	λ_{11}	0.900	1.113	1.926	1.896	0.903	1.110
	λ_{22}	0.952	1.082	1.072	1.049	0.885	1.023
	λ_{33}	0.878	0.669	1.109	0.976	1.275	1.383

Data simulation and application

Table 5.6: Illustration of the results – continued

Fitting Q_{11} and Q_{22} transition matrices (three-state model) in different time periods – continued							
		Data simulation with Q_1		**Data simulation with Q_2**		**Data simulation with Q_3**	
		Missed observation (10 %) Time $(0-12)$		**Missed observation (10 %) Time $(0-24)$**		**Missed observation (10 %) Time $(0-36)$**	
		$n=100$		$n=100$		$n=100$	
Q	λ	**Assumptions violated**	**Assumptions not violated**	**Assumptions violated**	**Assumptions not violated**	**Assumptions violated**	**Assumptions not violated**
Transition matrix estimates for Q_{11}							
P_{11}	P_{11}	0.263	0.290	0.510	0.504	0.265	0.280
	P_{12}	0.403	0.442	0.290	0.298	0.306	0.306
	P_{13}	0.334	0.268	0.200	0.198	0.429	0.414
	P_{21}	0.263	0.289	0.508	0.502	0.265	0.278
	P_{22}	0.403	0.444	0.292	0.300	0.306	0.305
	P_{23}	0.334	0.268	0.200	0.198	0.430	0.416
	P_{31}	0.263	0.289	0.503	0.498	0.265	0.280
	P_{32}	0.403	0.443	0.295	0.302	0.306	0.305
	P_{33}	0.334	0.268	0.203	0.200	0.429	0.416
Probability matrix estimates for Q_{22}							
P_{22}	P_{11}	0.263	0.283	0.505	0.504	0.264	0.278
	P_{12}	0.403	0.442	0.288	0.298	0.308	0.308
	P_{13}	0.334	0.276	0.207	0.198	0.429	0.414
	P_{21}	0.263	0.280	0.503	0.502	0.264	0.276
	P_{22}	0.403	0.444	0.290	0.301	0.307	0.307
	P_{23}	0.334	0.276	0.206	0.198	0.430	0.417
	P_{31}	0.263	0.281	0.497	0.498	0.264	0.277
	P_{32}	0.403	0.443	0.293	0.303	0.307	0.306
	P_{33}	0.334	0.276	0.209	0.200	0.429	0.416
-2 log likelihood estimates for Q_{11} and Q_{22}							
Q_{11}		2 418.014	2 384.42	4 302.814	4 298.525	7 051.181	6 952.542
Q_{22}		2 418.013	2 385.16	4 301.945	4 298.537	7 051.088	6 952.793
Pearson statistic estimates for Q_{11} and Q_{22}							
Q_{11}		24.38 (p = 0.1)	24.38 (p = 0.200)	19.05 (p = 0.319)	13.52 (p = 0.577)	20.79 (p = 0.136)	15.17 (p = 0.377)
Q_{22}		24.38 (p = 0.1)	24.37 (p = 0.200)	18.14 (p = 0.372)	13.53 (p = 0.577)	20.71 (p = 0.139)	15.45 (p = 0.363)

Data simulation and application

From table 5.3, Q_{11} and Q_{22} were both fitted across Q_1, Q_2 and Q_3 which each Q (that is Q_1, Q_2 and Q_3 were used to simulate two different datasets. One using our defined R code where assumptions are violated (Assumptions violated from table 5.3) and the other using ***msn package*** where assumptions are assumed to be not violated (Assumptions not violated from table 5.3.1). Comparing transition intensity rates between Q_{11} and Q_{22} within the same dataset (for Q_1, $n = 100$ and time period of $0 - 12$) we see that the transition intensity rates don't differ too much (almost the same). This chapter is based on recurrent model where the subject or patient can move from state 1 (Healthy) to any state and move backward. Looking at transition rates under assumptions violated (for Q_1, $n = 100$ and time period of $0 - 12$) we see that the estimated transition rate for moving from state 1 (Healthy) to state 2 (Mild) or to state 3 (Severe) is above 60 %. The estimated transition rate for moving backward (recovery) from state 2 (Mild) to state 1 (Healthy) is also above 60 % while the estimated transition rate for moving from state 2 (Mild) to state 3 (Severe) is just above 50 %.

Comparing probability intensity rates between P_{11} and P_{22} for Q_{11} and Q_{22} within the same dataset we see that the transition probabilities don't differ too much at all. In five years' time the probability of moving from state 1 (Healthy) to state 2 (Mild) 0.376 (38 %) and the probability of moving state 1 (Healthy) to state 3 (Severe) is 0.374 (37 %). Comparing the sojourn time between S_{11} and S_{22} for Q_{11} and Q_{22} within the same dataset we see no differences. The mean time spent in state 1 (Healthy) before moving to state 2 (Mild) is 0.900, the mean time spent in state 2 (Mild) before moving to state 3 (Severe) is 0.953 and the mean time spent in state 3 (Severe) before moving backward is (0.878). Looking at the -2Log-likelihood Estimate for Q_{11} and Q_{22} portions from table 5.4 we see that minus twice the maximised log-likelihood for Q_{11} and Q_{22} within the same dataset is the same. Looking at the Pearson Statistic Estimates for Q_{11} and Q_{22} portion within the same dataset we see that Pearson-type goodness-of-fit test for multi-state model for Q_{11} and Q_{22} are similar. Since we have small Pearson statistics for both case and p-value is greater than 0.05 both models fit very well.

Looking across Q_1, Q_2 and Q_3 from table 5.3 under (Assumptions violated) for Q_{11} and Q_{22} we see that the estimated transition intensity rates, probability transition rates, sojourn time and statistics differs significantly. Also under (Assumptions not violated) the above estimates

Data simulation and application

differs significantly but looking between (Assumptions violated) and (Assumptions not violated) for each Q we see differences but not clear enough but this has been shown by Prevalence vs Time plot (check figures below). Table 5.3 shows the results of assessing the assumption of homogeneity of the transition rates across the subject population. Based on this results we can conclude that if you ignore multi-state model assumptions you might overestimate or underestimate the transition rates or any estimates. This indicate that these assumptions are important and need to be considered when fitting a multi-state model.

Table 5.5 show the results of assessing the assumption of homogeneity of the transition rates through time and the same interpretation describe above can be applied in this table. So this assumption is also important. The third assumption is already incorporated into the other two assumptions through the R code that was written, so the same interpretation is equivalent.

Figure 5.4 and 5.5 shows the Prevalence vs Time plot for the model fitted using (Assumptions violated) and (Assumptions not violated) at time period of ($0-12$), sample size of $n = 200$ and 10 % of missed observations to assess the importance of the assumption of homogeneity of the transition rates across the subject population. From figure 5.5 we have 100 % of patients who were healthy (state 1) at the beginning of the study but as we moving away from month zero to month two the number decrease to around 50 %. From state 2 (Mild) at the beginning we had 0 % of patients but as we approach month two the number increase up to around 28 % and in state 3 (Severe) at the beginning we had 0 % of patients but as we approach month two the number increase up to around 20 %. In figure 5.4 where the assumptions were violated we can see that the plots differs significantly. So assumptions are very important.
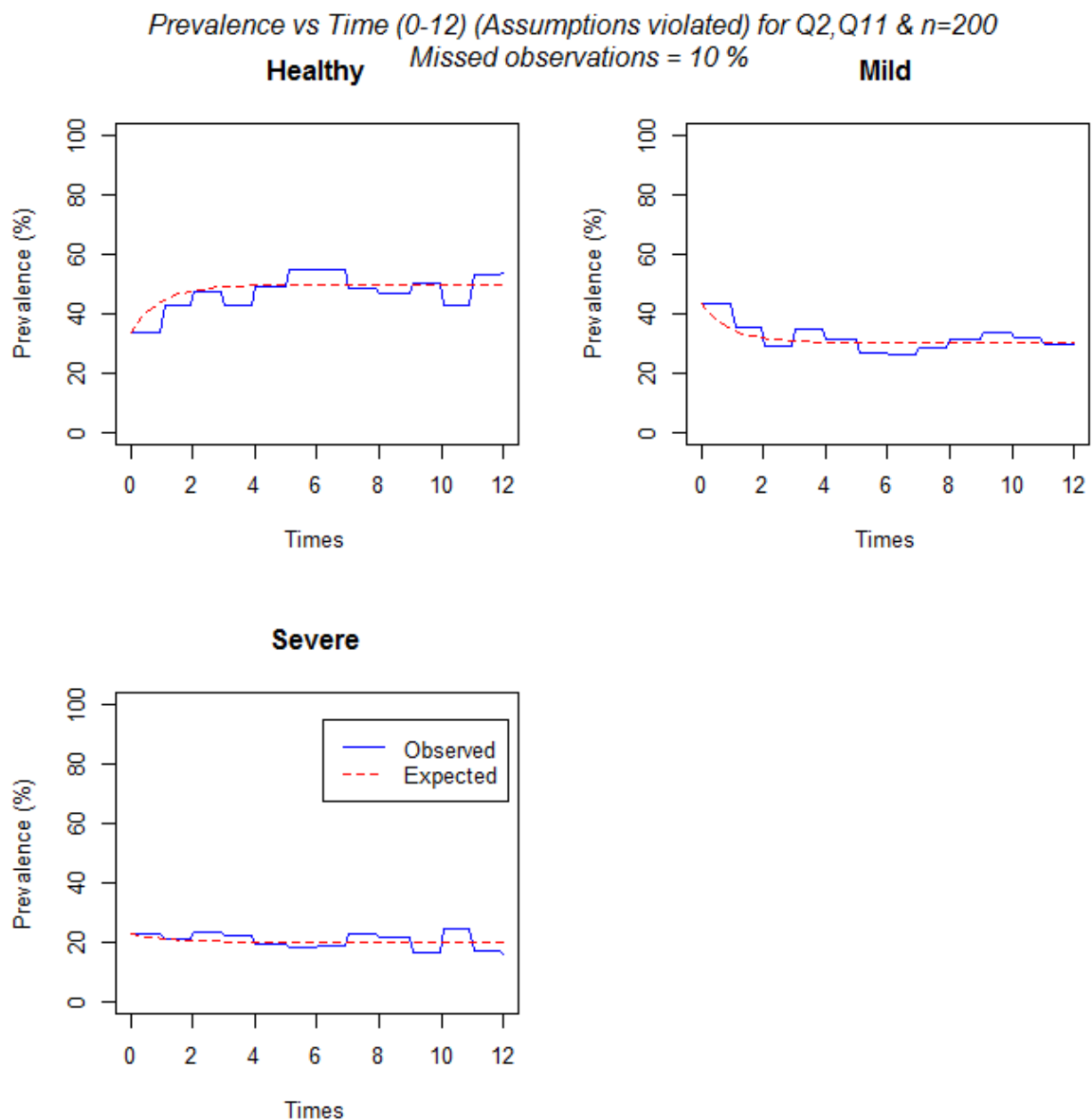
Data simulation and application

Figure 5.4: Prevalence vs Time plot when assumptions violated (testing assumption 1)
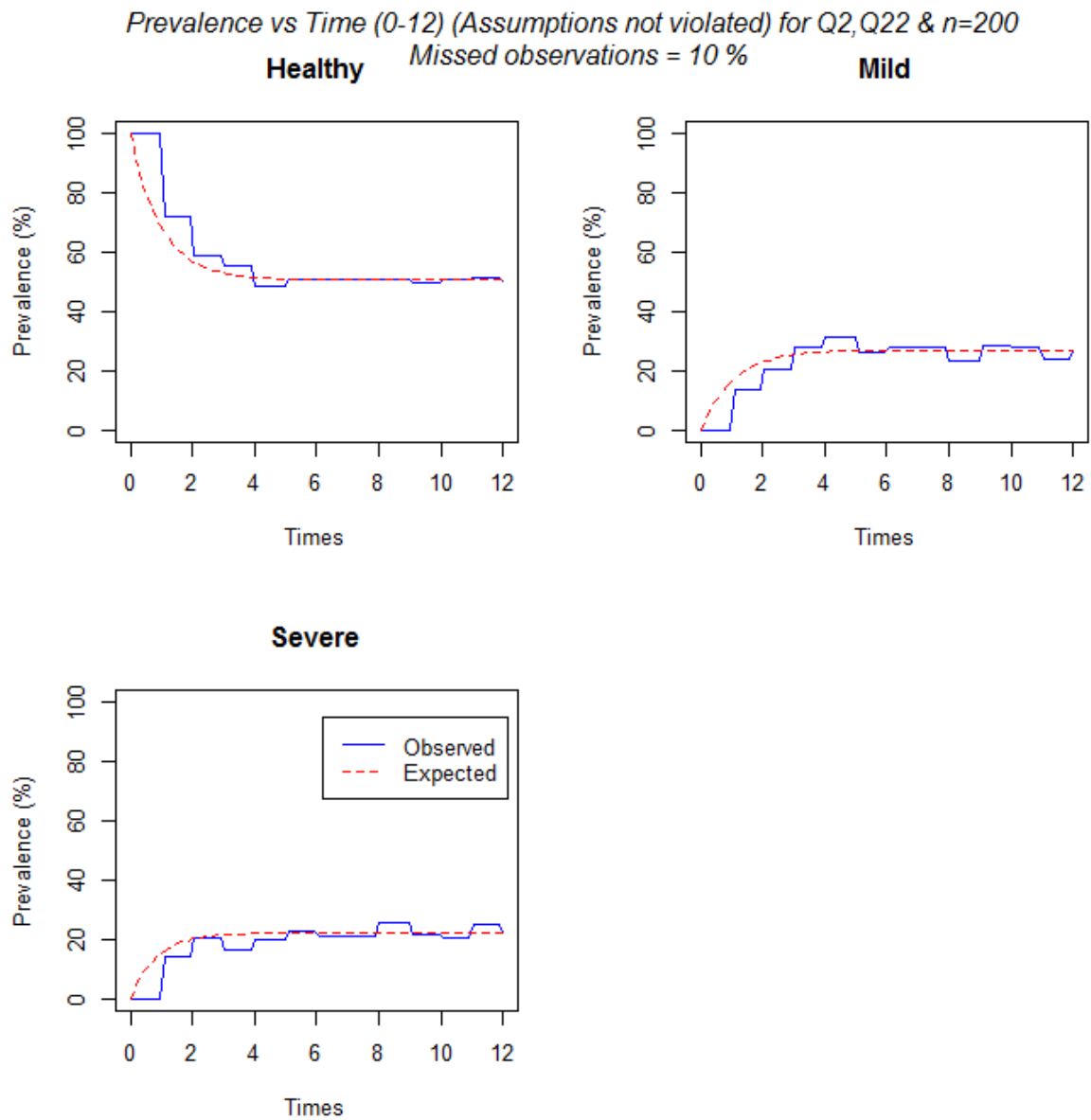
Data simulation and application

Figure 5.5: Prevalence vs Time plot when assumptions not violated (testing assumption 1)

Data simulation and application

Figure 5.6 and 5.7 shows the Prevalence vs Time plot for the model fitted using (Assumptions violated) and (Assumptions not violated) at time period of $(0-36)$, sample size of $n = 100$ and 10 % of missed observations to assess the importance of the assumption of homogeneity of the transition rates through time. Again the same interpretation as above can be applied here.
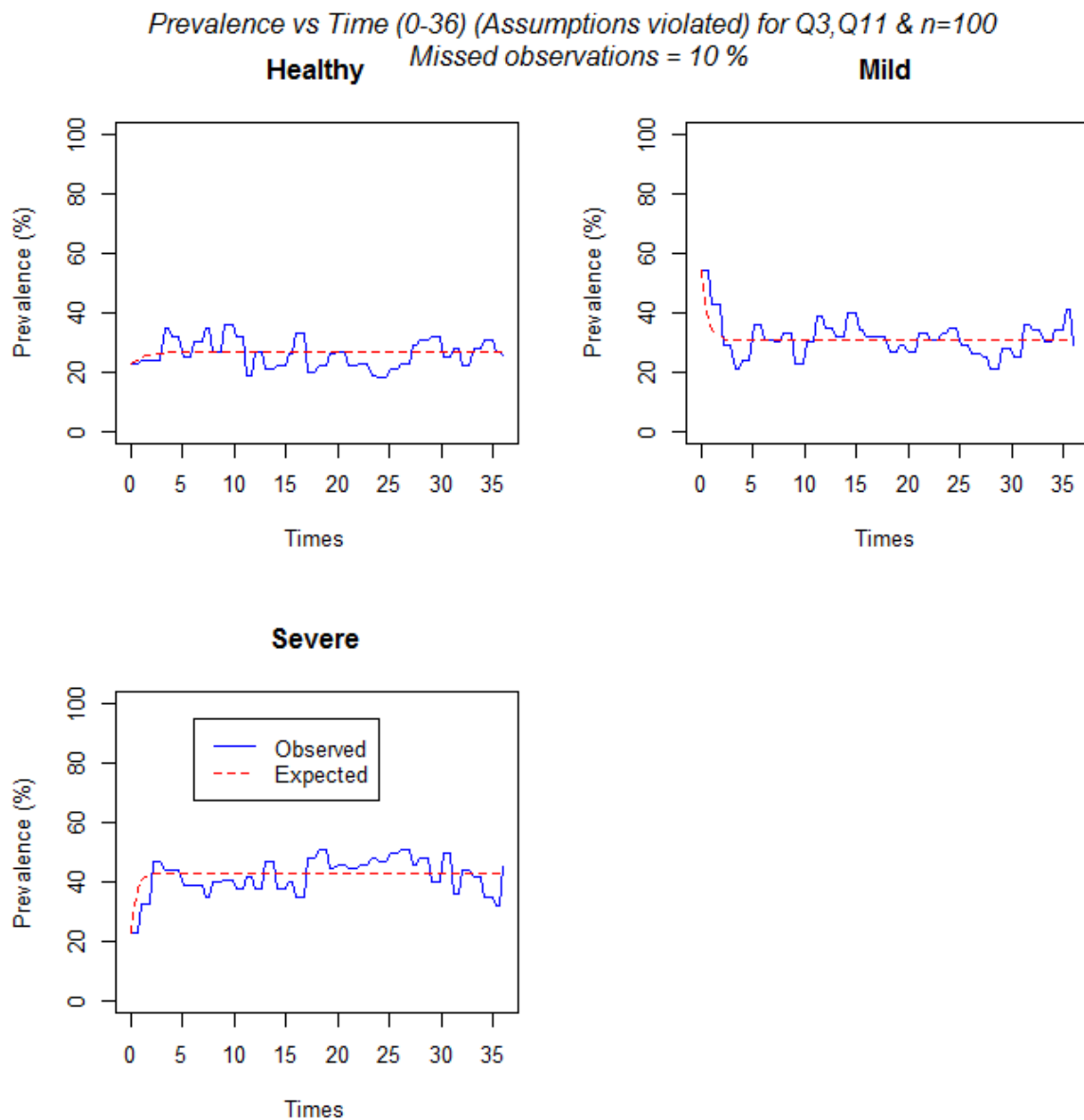

Figure 5.6: Prevalence vs Time plot when assumptions violated (testing assumption 2)
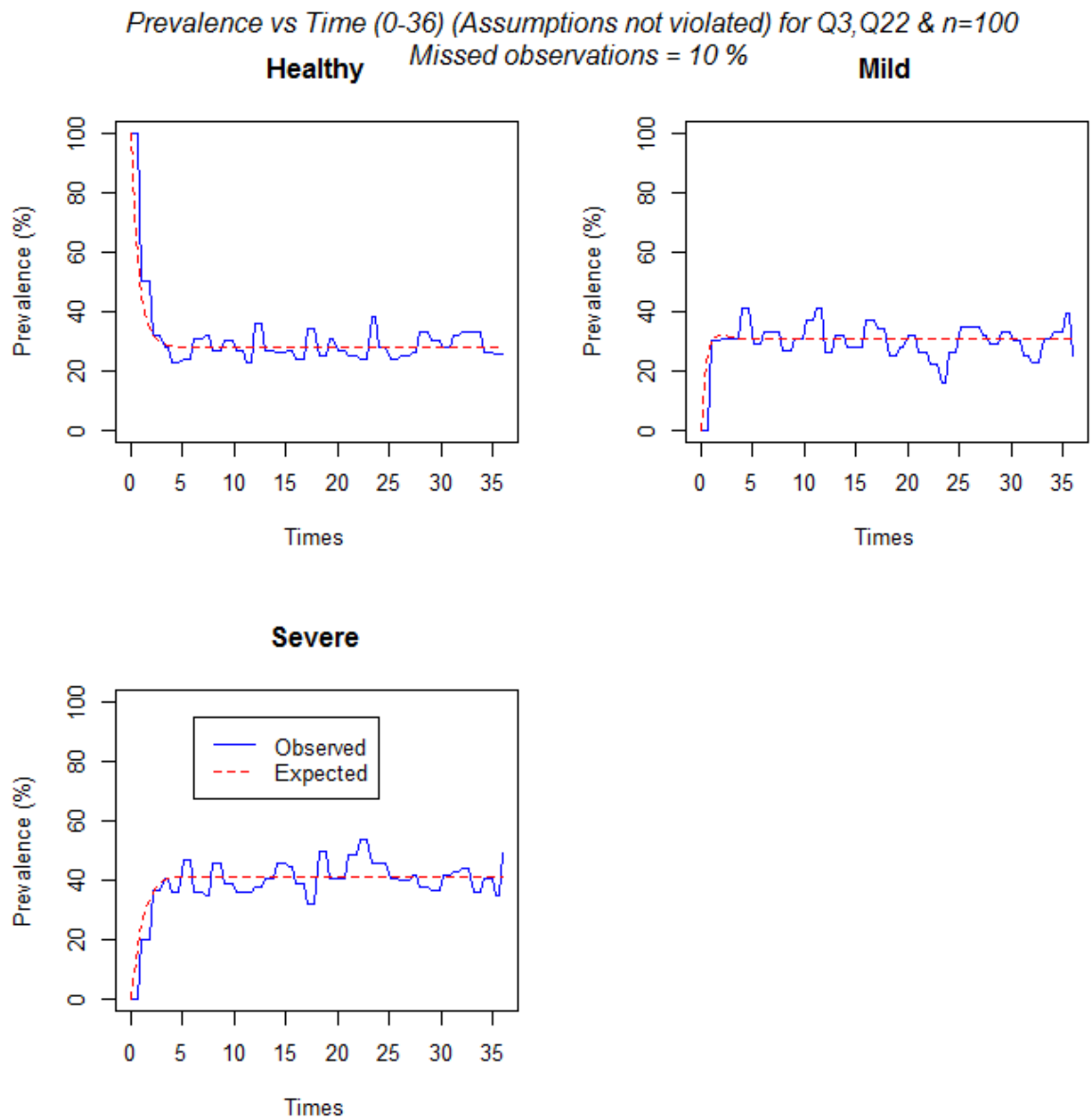
Data simulation and application

Figure 5.7: Prevalence vs Time plot when assumptions not violated (testing assumption 2)

Data simulation and application

5.4 Conclusion

In this chapter we coded an R code for data simulation. The R code simulate two datasets, one where the above assumptions are specifically violated and the other one using *simmulti.msm* function in *msm package*. The dataset using *simmulti.msm* function is assumed to simulate data without violating the above mentioned assumptions. For each Q we simulate two datasets, one using our defined R code where assumptions are violated and the other using *msn package* where assumptions are assumed to be not violated. To assess the importance of assumptions we compared the transition intensity matrices, probability transition matrices, sojourn time and other statistic estimates between these two datasets.

If for example transition rates do not differ very much then we can conclude that assumptions are not important otherwise assumptions are very important. Once the simulated datasets has been simulated using above three different Q's (that is Q_1, Q_2 and Q_3) then Q_{11} and Q_{22} transition intensity matrices were fitted in each case. To assess the assumption of homogeneity of the transition rates across the subject population Q_1, sample size of 100 ($n = 100$), missed observations of (10 %) and time period of ($0-12$) were used to simulate the first dataset, then Q_2, sample size of 200 ($n = 200$), missed observations of (10 %) and time period of ($0-12$) were used to simulate the second dataset and Q_3, sample size of 300 ($n = 300$), missed observations of (10 %) and time period of ($0-12$) were used to simulated the third dataset.

Then to assess the importance of the assumption we fitted Q_{11} and Q_{22} in each case and compared transition rates across different sample size (n) and also across different dataset. To assess the assumption of homogeneity of the transition rates through time we used the same describe above method but now the sample size (n) was held constant at $n = 100$ and the time period vary in each case, that is for Q_1 we use time period of ($0-12$), for Q_2 we use time period of ($0-24$) and for Q_3 we use ($0-36$). To assess Markov assumption the R code was written in such a way that it relax this assumption and transition rates were compared

From table 5.3, Q_{11} and Q_{22} were both fitted across Q_1, Q_2 and Q_3 which each Q (that is Q_1, Q_2 and Q_3 were used to simulate two different datasets. Comparing transition intensity rates between Q_{11} and Q_{22} within the same dataset (for Q_1, $n = 100$ and time period of $0-12$) we see that the transition intensity rates don't differ too much (almost the same).Comparing

Data simulation and application

probability intensity rates between P_{11} and P_{22} for Q_{11} and Q_{22} within the same dataset we see that the transition probabilities don't differ too much at all. Comparing the sojourn time between S_{11} and S_{22} for Q_{11} and Q_{22} within the same dataset we see no differences.

Looking at the -2Log-likelihood Estimate for Q_{11} and Q_{22} portion from table 5.3 we see that minus twice the maximised log-likelihood for Q_{11} and Q_{22} within the same dataset is the same. Looking at the Pearson statistic estimates for the Q_{11} and Q_{22} portions within the same dataset, one sees that the results of the Pearson goodness-of-fit test for multistate models are similar for Q_{11} and Q_{22}. Since we have small Pearson statistics for both case and p-value is greater than 0.05 both models fit very well. Looking across Q_1, Q_2 and Q_3 from table 5.3 under (Assumptions violated) for Q_{11} and Q_{22} we see that the estimated transition intensity rates, probability transition rates, sojourn time and statistics differs significantly. Also under (Assumptions not violated) the above estimates differs significantly but looking between (Assumptions violated) and (Assumptions not violated) for each Q we see differences but not clear enough but this has been shown by Prevalence vs Time plot (check figures in section 5.3). Based on this results we can conclude that if you ignore multi-state model assumptions you might overestimate or underestimate the transition rates or any estimates. This indicate that these assumptions are important and need to be considered when fitting a multi-state model.

Chapter 6

Summary

In chapter 1 we discussed the overview, aim and structure of the thesis, particular the aim of the thesis is to analyse the importance of the model assumptions when modelling the multi-state panel data. The multi-state model assumptions we analysed were homogeneity of the transition rates through time, homogeneity of the transition rates across the subject population and Markov assumption which state that the transition rates only depend on the history of the process through the current state. The assumptions were assessed by generating simulated datasets where the assumptions were specifically violated and assumptions were not violated then the model were fitted in each of the dataset.

In Chapter 2, the stochastic process, the transition probability matrix, the transition intensity matrix, Markov models, sojourn time, model assumptions and time homogeneous Markov models were discussed. The stochastic process that governs the multistate process through its transition probabilities and transition rates was discussed in detail. In stochastic process the system enters a state, spends an amount of time then moves to another state where it spends another time. The time that a system spends in a state is called sojourn or mean time which is also discussed in chapter 2. The probability transition matrix and transition intensity matrix both indicate the probabilities and intensity rates of transition of subject through different states respectively. The set of states, transition rates and probabilities are known as Markov model was also discussed in chapter 2. These components can be considered as building blocks for multi-state model.

In chapter 3 we discussed in details the covariates and model structure which also can be considered as building blocks for multi-state model. The model structure refers to the graphical representation of a model in terms of states and transitions between these states. The covariates on the other hand refers to the variable that is included to the model and assess its influence to the transition between the states. We considered in details in chapter 3 the basic survival model, multiple-decrement survival model, progressive model, disability model, recurrent model, competing risk model and continuing care retirement communities model. The results for these models were monitored and compared but the results that are shown in this thesis were the

results for recurrent model. The example of a recurrent model is the two state illness-recovery model where state 1 represents healthy and state 2 represents illness meaning that the subject can get ill then also can recover from illness going back to state 1 from state 2.

The focus in chapter 4 was to assess the multi-state model through model assumptions, covariates effect in the model and the fit of the model. The likelihood ratio test can be used to test the first assumption, the second assumption can also be tested using likelihood ratio test but including covariates and the last assumption can be tested by including the covariates depending on the history. To assess the significance of covariates we employ likelihood ratio and Wald test. To assess the fit of the multi-state model we use Pearson goodness-of-fit and informal model diagnostic tool. The model fit can be assessed by considering observed and expected transition frequencies either through a likelihood ratio test or the asymptotically equivalent Pearson chi-square statistics. Informal diagnostic tool involves prevalence counts where the observed state occupancies are compared with the expected occupancies under the fitted model.

After we looked in details of each components of multi-state model now it was time to fit the model and assess the importance of model assumptions and this was done chapter 5. In chapter 5 an R code was written for data simulation. The R code simulate two datasets, one where the above assumptions are specifically violated and the other one using *simmulti.msm* function in *msm package* in which we assumed that the above mentioned assumptions were not violated. After data simulation the models were fitted to each dataset and compared. Based on the results indicated in section 5.3, the multi-state model assumptions are very important and if you ignore them then you might underestimate or overestimate the estimates.

Bibliography

Agresti, A. 2002. *Categorical data analysis* (2th ed.). New Jersey: Wiley-Interscience.

Cinlar, E. 1975. Introduction to Stochastic Processes (1st ed.). New York: Dover Publications.

Aguirre-Hernández, R. & Farewell, V.T. 2002. A Pearson-type goodness-of-fit test for stationary and time-continuous Markov regression models. *Statistics in Medicine*, 21:1899–1911.

Andersen, P.K., Abildstrom, S.Z. & Rosthoj, S. 2002. Competing risks as a multi-state model. *Statistical Methods in Medical Research*, 11:203–215.

Andersen, P.K., Borgan, O., Gill, R.D. & Keiding. 1993. *Statistical models based on counting processes* (2nd ed.). Springer series in statistics. New York: Springer.

Ching, W., Micheal, K. & Ng. 2006. *Markov chains: Models, algorithms and applications*. Springer Science and Business Media.

Christopher, J. 2014. *Multi-state modelling with R: The msm package* [Online]. Available: http://cran.r-project.org/web/packages/msm/vignettes/msm-manual.pdf [2014, March 7].

Fosler-Lussier, E. 1998. *Markov models and hidden Markov models: A brief tutorial* [Online]. Available: http://www.di.ubi.pt/~jpaulo/competence/tutorials/hmm-tutorial-1.pdf [2014, February 22].

Hollander, M. & Proschan, F. 1976. Testing to determine the underlying distribution using randomly censored data. *Biometrics*, 35:393–401.

Hougaard, P. 1999. *Multi-state models: A review* [Online]. Available: http://link.springer.com/article/10.1023%2FA%3A1009672031531#page-1 [2014, February 22].

Hsiao, C. 2003. *Analysis of panel data*. Cambridge: Cambridge University Press.

Ibe, O.C. 2009. *Markov processes for stochastic modelling*. Elsevier Academic Press.

Jackson, C.H. 2005. *Multi-state modelling with R: The msm package.* R msm package software manual [Online]. Available: http://cran.r-project.org/ [2014, February 28].

Bibliography

Jackson, C.H. 2011. Multi-state models for panel data: The msm package for R. *Journal of Statistical Software*, 38:8.

Jackson, C.H., Sharples, L.D., Thompson, S.G., Duffy, S.W. & Couto, E. 2003. Multistate Markov models for disease progression with classification error. *The Statistician*, 52:193–209.

Jon. 2009. *Stochastic processes and Markov chains* [Online]. Available: http://research.sourcebyte.com/blog/stochastic-processes-and-markov-chains/ [2014, March 5].

Kalbfleisch, J.D. & Lawless, J.F. 1985. The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association*, 80:863–871.

Kay, R. 1986. A Markov model for analysing cancer markers and disease states in survival studies. *Biometrics*, 42:855–865.

Klein, J.P. & Moeschberger, M.L. 2003. *Survival analysis techniques for censored and truncated data* (2nd ed.). New York: Springer.

Lebovic, G. 2011. *Estimating non-homogeneous intensity matrices in continuous time multi-state Markov models* [Online]. Available:https://tspace.library.utoronto.ca/bitstream/1807/29786/6/Lebovic_Gerald_B_2011 06_PhD_thesis.pdf [2014, March 11].

Longini Jr., I.M., Clark, W.S., Byers, R.H., Ward, J.W., Darrow, W.W., Lemp, G.F. & Hethcote, H.W. 1989. Statistical analysis of the stages of HIV infection using a Markov model. *Statistics in Medicine*, 8:831–843.

Marshall, G. & Jones, R.H. 1995. Multi-state models and diabetic retinopathy. *Statistics in Medicine*, 14:1975–1983.

Meira-Machado, L., Cadarso-Suárez, C. & De Uña-Álvarez, J. 2008. Inference in the progressive three-state model. *International Journal of Mathematical Models and Methods in Applied Sciences*, 3:447–454.

Rubino, G. & Sericola, B. 1988. *Sojourn times in finite Markov process* [Online]. Available: http://www.irisa.fr/dionysos/pages_perso/sericola/PAPIERS/J2.pdf [2014, March 7].

Bibliography

Meira-Machado, L., De Uña-Álvarez, J., Cadarso-Suárez, C. & Anderson, P.K. 2009. Multistate models for the analysis of time-to-event data. *Statistical Methods in Medical Research*, 18:195–222.

Pérez-Ocón, R., Ruiz-Castro, J.E. & Gámiz-Pérez, M.L. 1998. A multivariate model to measure the effect of treatments in survival of breast cancer. *Biometrical Journal*, 40:703–715.

Pérez-Ocón, R., Ruiz-Castro, J.E. & Gámiz-Pérez, M.L. 2001. Non-homogeneous Markov models in the analysis of survival after breast cancer. *Journal of the Royal Statistical Society: Series C*, 50:111–124.

Rabiner, L. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–286.

Ross, M.S. 2003. *Introduction to probability models* (8th ed.). San Diego: Academic Press.

Sharples, L.D., Jackson, C.H., Parameshwar, J., Wallwork, J. & Large, S.R. 2003. Diagnostic accuracy of coronary angiography and risk factors for post-heart-transplant cardiac allograft vasculopathy. *Transplantation*, 76(4):679–682.

Titman, A.C. 2007. *Model diagnostics in multi-state models of biological systems* [Online]. Available: http://www.maths.lancs.ac.uk/~titman/thesis.pdf [2014, March 6].

Titman, A.C. & Sharples, L.D. 2007. A general goodness-of-fit test for Markov and hidden Markov models. *Statistics in Medicine*, 27:2177–2195.

Titman, A.C. & Sharples, L.D. 2010. Model diagnostics for multi-state models. *Statistical Methods in Medical Research*, 19:621–651.

Welton, N.J. & Ades, A.E. 2005. *Estimation of Markov chain transition probabilities and rates from fully and partially observed data* [Online]. Available: http://mdm.sagepub.com/content/25/6/633.full.pdf+html [2014, March 6].

Rubino, G and Sericola B. 1988. *Sojourn Times in Finite Markov Processes* [Online]. Available:http://www.irisa.fr/armor/lesmembres/Rubino/Para%20ev/Markov/sojourns.pdf [2014, March 6].

Cox, D. R. and Miller, H. D. (1965) The Theory of Stochastic Processes. London:Chapman and Hall

Appendix

Appendix

7.1 R code

```
> #### The following function simulate data ####

> #### based on violating Markov models assumptions ####

> #### and based on using msm package (simmulti.msm function) ####

> #### Then it fit the models using both dataset ####

> #### to compare and assess the assumptions ####

> simdata <- function(Q1,mis_obs,n,m,T,tg){

+       #### Set empty data ####

+       nuldata <- NULL

+       nuldata2 <- NULL

+       #### Define transition matrix ####

+       #### Q matrix is used to simulate the data #####

+       #### Q11 and Q22 are used to fit the model under selected Q ####

+       if(Q1 ==1){

+         Q <- rbind(c(-0.92,0.51,0.41),c(0.49,-0.98,0.49),c(0.32,0.93,-0.61))

+         Q11 <- rbind(c(-1.0,0.70,0.30),c(0.60,-0.90,0.30),c(0.52,0.33,-0.85))

+         Q22 <- rbind(c(-1.0,0.85,0.15),c(0.45,-0.80,0.35),c(0.71,0.25,-0.96))

+         rownames(Q11) <- colnames(Q11) <- c("Healthy", "Mild","Severe")

+         rownames(Q22) <- colnames(Q22) <- c("Healthy", "Mild","Severe")

+       }else if(Q1 == 2){

+         Q <- rbind(c(-0.5,0.25,0.25),c(0.65,-0.90,0.25),c(0.21,0.75,-0.96))
```

Appendix

```
+        Q11 <- rbind(c(-1.0,0.70,0.30),c(0.60,-0.90,0.30),c(0.52,0.33,-0.85))

+        Q22 <- rbind(c(-1.0,0.85,0.15),c(0.45,-0.80,0.35),c(0.71,0.25,-0.96))

+        rownames(Q11) <- colnames(Q11) <- c("Healthy", "Mild","Severe")

+        rownames(Q22) <- colnames(Q22) <- c("Healthy", "Mild","Severe")

+      }else if(Q1 == 3){

+        Q <- rbind(c(-1.0,0.75,0.25),c(0.25,-1.0,0.75),c(0.45,0.25,-0.70))

+        Q11 <- rbind(c(-1.0,0.70,0.30),c(0.60,-0.90,0.30),c(0.52,0.33,-0.85))

+        Q22 <- rbind(c(-1.0,0.85,0.15),c(0.45,-0.80,0.35),c(0.71,0.25,-0.96))

+        rownames(Q11) <- colnames(Q11) <- c("Healthy", "Mild","Severe")

+        rownames(Q22) <- colnames(Q22) <- c("Healthy", "Mild","Severe")

+      }

+      #### Simulate data based on violating Markov model assumptions ####

+      for(k in 1:n){

+          num_of_obs <- round(runif(1,0,mis_obs),3)

+          num_of_obs <- round((1-num_of_obs) * m)

+          obs_times <- seq(1,m,1)

+          times_for_subct <- sort(sample(obs_times,size=num_of_obs))

+          times_for_subct <- c(0,times_for_subct)

+          nxt_state <- round(runif(1,1,dim(Q)[[1]]))

+          nxt_state2 <- nxt_state

+          j <- 1

+          state <- c()

+          for(i in 2: length(times_for_subct)){
```

Appendix

```
+              time_diff <- times_for_subct[i] - times_for_subct[j]

+              P_matrx <- round(MatrixExp(Q,t=time_diff),3)

+              nxt_state <- c(rmultinom(1,1,prob=c(P_matrx[nxt_state,])))

+              j <- j + 1

+              index <- which(nxt_state !=0, arr.ind = T)

+              nxt_state <- index

+              state[i] <- nxt_state

+        }

+        state <- c(nxt_state2,state)

+        state <- state[!is.na(state)]

+        data <- data.frame(subject=rep(k,num_of_obs+1),

+        time=times_for_subct,state=state)

+        datasim <- rbind(nuldata,data)

+        nuldata <- datasim

+        #### create data frame for simmulti.msm function ####

+        data2 <- data.frame(subject=rep(k,num_of_obs+1),time=times_for_subct)

+        simdata <- rbind(nuldata2,data2)

+        nuldata2 <- simdata

+    }

+    #### Simulate data using simmulti.msm function ####

+    simdata <- simmulti.msm(simdata,Q)

+    #### Fit the model using data based on violaed assumptions ####

+    model_viol_assump <-msm(state ~ time,subject=subject,
```

Appendix

```
+     data=datasim,qmatrix=Q11)

+     Q_matrix_viol_assump <- round(qmatrix.msm(model_viol_assump)$estimates,3)

+     P_matrx_viol_assump <- round(pmatrix.msm(model_viol_assump,t=T),3)

+     Mean_Time_viol_assump <- round(sojourn.msm(model_viol_assump),3)

+     Pearson_viol_assump <- round(pearson.msm(model_viol_assump,

+     timegroups=tg)$test,3)

+     model_viol_assump2 <-msm(state ~ time,subject=subject,

+     data=datasim,qmatrix=Q22)

+     Q_matrix_viol_assump2 <- round(qmatrix.msm(model_viol_assump2)$estimates,3)

+     P_matrx_viol_assump2 <- round(pmatrix.msm(model_viol_assump2,t=T),3)

+     Mean_Time_viol_assump2 <- round(sojourn.msm(model_viol_assump2),3)

+     Pearson_viol_assump2 <- round(pearson.msm(model_viol_assump2,

+     timegroups=tg)$test,3)

+     #### Fit the model using data based on simmulti.msm function ####

+     model_with_assump <-msm(state ~ time,subject=subject,

+     data=simdata,qmatrix=Q11)

+     Q_matrix_with_assump <- round(qmatrix.msm(model_with_assump)$estimates,3)

+     P_matrx_with_assump <- round(pmatrix.msm(model_with_assump,t=T),3)

+     Mean_Time_with_assump <- round(sojourn.msm(model_with_assump),3)

+     Pearson_with_assump <- round(pearson.msm(model_with_assump,

+     timegroups=tg)$test,3)

+     model_with_assump2 <-msm(state ~ time,subject=subject,

+     data=simdata,qmatrix=Q22)
```

Appendix

```
+    Q_matrix_with_assump2 <- round(qmatrix.msm(model_with_assump2)$estimates,3)

+    P_matrx_with_assump2 <- round(pmatrix.msm(model_with_assump2,t=T),3)

+    Mean_Time_with_assump2 <- round(sojourn.msm(model_with_assump2),3)

+    Pearson_with_assump2 <- round(pearson.msm(model_with_assump2,

+    timegroups=tg)$test,3)

+    #### Compare the models ####

+    compare_models <- round(lrtest.msm(model_viol_assump,model_with_assump),9)

+    #### Plot the Prevalence vs Time for model based on assumption ####

+    windows()

+    plot.prevalence.msm(model_viol_assump,mintime=0,

+    maxtime=m,legend.pos=c(6.2, 95))

+    par(oma=c(0,0,2,0))

+    title(main="Prevalence vs Time (0-36) (Assumptions violated)

+    for Q3,Q11 & n=100 \n Missed observations = 10 %",font.main= 3,outer=T)

+    #### Plot the Prevalence vs Time for model without assumptions ####

+    windows()

+    plot.prevalence.msm(model_with_assump,mintime=0,

+    maxtime=m,legend.pos=c(6.2, 95))

+    par(oma=c(0,0,2,0))

+    title(main="Prevalence vs Time (0-36) (Assumptions not violated)

+    for Q3,Q22 & n=100 \n Missed observations = 10 %",font.main= 3,outer=T)

+    #### List and print the results ####

+    list("Transition Matrix violated assumptions"=Q_matrix_viol_assump,
```

Appendix

```
+    "Probability Matrix violated assumptions"=P_matrx_viol_assump,

+    "Sojourn Time violated assumptions"=Mean_Time_viol_assump,

+    "Pearson Statistic violated assumptions"=Pearson_viol_assump,

+    "Transition Matrix with assumptions"=Q_matrix_with_assump,

+    "Probability Matrix with assumptions"=P_matrx_with_assump,

+    "Sojourn Time with assumptions"=Mean_Time_with_assump,

+    "Pearson Statistic with assumptions"=Pearson_with_assump,

+    "Compare Models"= compare_models)

+ }

> outputdata <- simdata(Q1=3,mis_obs=0.10,n=100,m=36,T=5,tg=2)

> outputdata
```